INVESTIGATION OF FACTORS RELATED TO IMMIGRANT AND NATIVE STUDENTS' MATHEMATICS PERFORMANCE IN PISA 2018

by

Ayşe Akkır

B.S., Undergraduate Program in Primary Mathematics Education, Boğaziçi University, 2020

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Secondary School Science and Mathematics
Boğaziçi University
2023

ACKNOWLEDGEMENTS

I would first like to thank my advisor Assoc. Prof. Serkan Arıkan for his constant feedback and support. His advice and guidance have been invaluable in helping me become a better researcher. He has given me advice and encouragement that has enabled me to stay motivated and on task. I am deeply grateful to him for his guidance and mentorship.

In addition, I would like to express my sincere gratitude to my thesis committee members, Assist. Prof. Gaye Defne Ceyhan and Assoc. Prof. Seher Yalçın, for reviewing my thesis and providing critical comments in order to make it better. Their feedback and support have been very helpful in developing a strong thesis.

I am also grateful for the support of my friends who have been with me throughout this journey. In particular, I would like to thank Esila, my best friend, who has always been there to listen to me and offer advice. I'm so grateful for her friendship and the support she has given me. Moreover, I would like to thank Sare, who has always understood me and encouraged me to complete this thesis. As well, I would like to thank my friends and my study buddies, Aybüke and Selen, for their support in my academic journey and their encouragement.

Lastly, I would like to thank my parents, Hacer and Sebahattin, who have been the biggest source of motivation and support throughout my life. They have always believed in me and given me the strength to keep going even when things get tough. I am truly blessed to have them in my life. As well, I would also like to express my sincere gratitude to my sister Betül who has always been by my side during all of this. I am incredibly lucky to have her in my life and thankful for everything she has done for me.

ABSTRACT

INVESTIGATION OF FACTORS RELATED TO IMMIGRANT AND NATIVE STUDENTS' MATHEMATICS PERFORMANCE IN PISA 2018

This study examined the mathematics performance gap between native and immigrant students, and between first- and second-generation immigrant students in European countries in the 2018 Program for International Student Assessment (PISA). Another objective was to understand how the performance gap changes after controlling for economic and social status (ESCS) and gender. Factors that can predict the mathematics performance of immigrant and native students were examined. Student-level variables were chosen based on Walberg's academic achievement theory. Country-level variables, such as the Migrant Integration Policy Index (MIPEX), the Human Development Index (HDI), and government expenditure on education were included. An independent sample t-test was used to investigate the performance gap. Multilevel regression analyses were conducted to investigate mathematics performance at the student- and country-levels. Propensity score matching was used to investigate the achievement gap after controlling gender and ESCS. Immigrants' and natives' mathematics performance is significantly predicted by government expenditure on education. In addition, the MIPEX was a significant predictor of first-generation immigrants' mathematics performance. Students' attitudes toward immigrants, discriminating school climate, and exposure to bullying were significant predictors of the mathematics performance of immigrants and natives. Resilience significantly predicted immigrants' mathematics performance. There was a statistically significant performance gap between natives and immigrants in 21 European countries. The performance gap narrowed after controlling for (ESCS) and gender.

ÖZET

GÖÇMEN VE YERLİ ÖĞRENCİLERİN PISA 2018 MATEMATİK PERFORMANSLARINA İLİŞKİN FAKTÖRLERİN İNCELENMESİ

Bu çalışma, Uluslararası Öğrenci Değerlendirme Programının (PISA) 2018'de Avrupa ülkelerindeki yerli ve göçmen öğrenciler ile birinci ve ikinci nesil göçmen öğrenciler arasındaki matematik performansı farkını incelemiştir. Bu çalışmanın bir diğer amacı, ekonomik ve sosyal statü (ESCS) ve cinsiyeti kontrol ettikten sonra performans farkının nasıl değiştiğini anlamaktır. Ayrıca göçmen öğrencilerin ve yerli öğrencilerin matematik performansını yordayan faktörler incelenmiştir. Öğrenci düzeyindeki değişkenler Walberg'in akademik başarı teorisine dayalı olarak seçilmiştir. Ülke düzeyindeki değiş- kenler göçmen entegrasyon politikası endeksi (MIPEX), insani gelişmişlik endeksi (HDI) ve eğitime yapılan devlet harcamasıdır. Yerli ve göçmen öğrenciler arasındaki matematik performans farkını araştırmak için bağımsız örneklem t-testi kullanılmıştır. Öğrenci ve ülke düzeylerinde matematik performansını araştırmak için çok düzeyli regresyon analizleri yapılmıştır. Cinsiyet ve ESCS kontrol edildikten sonra performans farkını araştırmak için eğilim skoru eşleştirmesi kullanılmıştır. Devletin eğitime yaptığı harcamaların, göçmenlerin ve yerlilerin matematik performansını yordadığı bulunmuştur. Ek olarak, MIPEX, birinci nesil göçmen öğrencilerin matematik performansının yordayan bir değişkendir. Öğrencilerin göçmenlere karşı tutumları, ayrımcı okul iklimi ve zorbalığa maruz kalma, göçmenlerin ve yerlilerin matematik performansı ile ilişkilidir. Dayanıklılığın, göçmenlerin matematik performansını yordadığı bulunmuştur. 21 Avrupa ülkesinde yerliler ve göçmenler arasında istatistiksel olarak önemli bir performans farkı vardır. ESCS ve cinsiyet kontrol edildikten sonra performans açığının daraldığı görülmektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

ρ	Intraclass Correlation
$\sigma^2_{u_e}$	The variance of the lowest-level errors
$\sigma^2_{u_0}$	The variance of the highest-level errors

LIST OF ACRONYMS/ABBREVIATIONS

ESCS	The Index of Economic, Social and Cultural Status
HDI	Human Development Index
ICT	Information and Communication Technologies
MIPEX	Migrant Integration Policy Index
OECD	Organization for Economic Cooperation and Development
PISA	Program for International Student Assessment
TIMSS	Trends in International Mathematics and Science Study
ZPRED	Standardized Predicted Value
ZRESID	Standardized Residuals

1. INTRODUCTION

1.1. Migration and Immigrants

Migration has a prolonged history dating back many centuries. People had chosen to migrate in the early times of human lives for some reasons. In today's world, people are immigrating, hoping to find a new job or studying at a university. In addition, there are times when people have to immigrate because of tragic incidents such as disasters or wars. Immigration may occur both domestically and internationally.

According to the International Organization for Migration (2020), there were 272 million foreign migrants out of 7.7 billion people in 2019. As a percentage, 3.5% of the world population were immigrants. It means that there is one immigrant for every 30 individuals. In addition, 82 million international migrants were hosted in Europe in 2019 (International Organization for Migration, 2020). Europe has long been a central place for international immigrants to settle. Even though the rate of migration to European countries has slowed slightly in recent years, the share of immigrants in Europe (82 million) and Asia (84 million) in 2019 is 61% of total immigrants (International Organization for Migration, 2020).

As previously mentioned, a large number of people migrate from one country to another for various reasons. Immigrant students may face challenges in the societies in which they migrate. These challenges may affect them either psychologically or academically. Guerra *et al.* (2019) investigated the difference in perceived discrimination, well-being, and school achievement between immigrant students and natives. Their result revealed that immigrant students have significantly lower levels of peer acceptance and well-being among their peers. Guerra *et al.* (2019) claimed that immigrant children are at risk of social exclusion. Moreover, the school achievement of immigrant students was found to be lower than native students (Guerra *et al.*, 2019). To ensure a successful integration of immigrant children and their families, effective policies are required (Cerna *et al.*, 2021). A wide range of policy dimensions, including those related to education, employment, welfare, health, housing, urban planning, and economic development, is involved in developing and implementing integration measures (OECD, 2018a). Education plays a significant role because it may assist immigrants' socio-emotional well-being, help them develop the skills they need to participate in the economy and encourage their involvement in social and civic life of their communities (Cerna *et al.*, 2021).

1.2. Large-Scale Assessments

Large-scale assessments give insights about educational attainment for a specific curriculum, such as the Trends in International Mathematics and Science Study (TIMSS), or a specific age level, such as Program for International Student Assessment (PISA) (Clarke and Luna-Bazaldua, 2021). The results of large-scale assessments not only share information about nations' achievements but also provide some information to policymakers (Clarke and Luna-Bazaldua, 2021). Large-scale assessments can be either national or international. International large-scale assessments, such as PISA, not only measure students' performance but also gather data about students' backgrounds, teachers, principals, and schools (Clarke and Luna-Bazaldua, 2021). The purpose of gathering this data is to investigate how students' backgrounds relate to students' performance (Clarke and Luna-Bazaldua, 2021).

Large-scale international assessment facilitates comparing student achievement to peers from various countries (Clarke and Luna-Bazaldua, 2021). Greaney and Kellaghan (2008) stated that national and international assessments have common purposes, such as assessing how effectively students are learning within the educational system, accurately predicting specific areas of strength and weakness in the knowledge and skills that students have attained, comparing the achievements of subgroups within the population such as boys and girls, ascertaining the relationship between achievement and various aspects of the school settings as well as of homes and communities. On the other hand, international assessment has the force over a national assessment in that it aims to inform policymakers, educators, and the community on how a country's educational system differs from those of other systems (Beaton *et al.*, 1999).

1.3. The Program for International Student Assessment (PISA)

The PISA is an international large-scale assessment applied every three years (OECD, 2019a). PISA's objectives are to assess students' knowledge at the end of their compulsory education and investigate the application of knowledge within and without the school (OECD, 2019a). What makes PISA important is that it has unique features such as policy orientation, innovative concept of literacy, the relevance of lifelong learning, regularity, and breadth of coverage (OECD, 2019a). PISA is policy-oriented, allowing it to integrate data on student learning outcomes with student backgrounds to highlight differences in performance and the characteristics of students, schools, and educational systems that function well. PISA has its own innovative concept of literacy in essential areas such as reading, mathematics, and science. Literacy is explained as the ability to employ skills and knowledge in the fields of reading, math, and science, as well as the capacity to conduct analysis, reason, make communications, and solve problems in various contexts (OECD, 2019a). PISA is relevant to lifelong learning so that students' learning motivations and students' beliefs about themselves, and students' strategies for learning are asked students to report. Since PISA is regularly conducted over a three-year period, it allows countries to keep track of their progress on important learning objectives over the years (OECD, 2019a). PISA has broad coverage of countries because 37 OECD countries and 42 partner countries participated in 2018. Many regions of the world employ PISA as an assessment instrument. Forty-three nations and economies used it during the initial assessment. This number rose to 79 countries and economies in 2018 (OECD, 2019a). It can be said that PISA is a massively important assessment that helps nations compare their performance and formulate policy.

Reading was the primary emphasis of the PISA 2018 survey, with the assessments of mathematics, science, and global competence coming in second, and financial literacy being optional. The PISA 2018 assessment was completed by nearly 600000 students as a representation of 32 million students (OECD, 2019a). Most of the countries applied computer-based assessments except countries that were not able to apply them. Those countries used previously developed paper-based assessments (OECD, 2019a). The computer-based test takes two hours to complete. Moreover, student background questionnaires take 35 minutes to complete. Student background questionnaires are important because they collect information about their attitudes, dispositions, beliefs, and their homes, schools, and learning experiences (OECD, 2019a).

1.4. Large-Scale Assessments of Immigrant Students

The large-scale assessment provides an opportunity to address some issues about immigrant students. Since education plays an important role in integrating immigrants, the assessment may show evidence of immigrant students' achievement so that policy-makers can take action about immigrant education based on the analysis of immigrant students' achievement. One significant advantage of PISA is that it collects data on the background of students and their parents. In PISA 2018, parents were asked about their immigrant or non-immigrant background in the parents' questionnaire (OECD, 2019a). The parents' questionnaire was applied in 17 countries (OECD, 2019a).

1.5. Walberg's Theory of Academic Achievement

Walberg's theory of academic achievement claims that students' achievement is influenced by their traits and their surroundings as well. The theory is comprised of three dimensions such as student aptitude, instruction, and psychological dimension. Student aptitude dimension includes ability, development, motivation, or a student's propensity for intense persistence in educational tasks. Instruction dimension is defined as the quantity and the quality of the instructional time. Psychological environments encompass the morale or students' views of their peers in the classroom and the home environment (Walberg, 2004).

The choice of this theory is based on a number of factors. It is first important to note that the theory was developed based on meta-analyses of earlier research, including large-scale assessments (Walberg, 2004). According to a preliminary synthesis of 2575 study comparisons, three categories of psychological factors were identified as the primary drivers of academic success. Following the initial synthesis, a successive synthesis produced congruent results with the earlier synthesis (Walberg, 2004). Therefore, one of the reasons for choosing this theory is that it was developed by an in-depth examination of the literature and validated by the findings of a meta-analysis. As the theory focuses on a wide range of research, it is comprehensive in that it includes both inner variables such as students' aptitudes, as well as extrinsic factors such as classroom instruction students' home environment. Consequently, the theory provides a framework for understanding the relationship between these variables and a students' achievement.

In the current study, for the student aptitude dimension of Walberg's Theory, motivation to master tasks, resilience, and cognitive flexibility/adaptivity were included. For the psychological environment dimension of Walberg's theory, exposure to bullying, sense of belonging, discriminating school climate, and students' attitudes toward immigrants were included.

1.6. Purpose of the Study

The purpose of the study was to investigate which student level (motivation to master tasks, resilience, cognitive flexibility/adaptivity, exposure to bullying, sense of belonging, discriminating school climate, students' attitudes toward immigrants) and country level (migrant integration policy index, human-development index, governmental expenditure on education) variables could predict the mathematics performance of immigrant students in Europe in PISA 2018, and native students as well. In order to choose the student-level variables, Walberg's theory of academic achievement as well as existing literature were considered. The literature also provided a basis for choosing country-level variables. In addition, the performance gap between immigrant and native students and the performance gap between first- and second-generation students was examined. To gain a better understanding of the differences, the performance gap was investigated after controlling for economic, cultural, and social status and gender.

1.7. Significance of the Study

This study investigated the performance gap between immigrant and native students because of its potential implications for policy on education. Education of immigrant students is considerably important because education may promote immigrants' involvement in society (Cerna *et al.*, 2021). In addition, the labor force measured by mathematics and science skills in TIMSS and National Assessment of Educational Progress (NAEP) has a positive correlation with economic growth (Hanushek and Kimko, 2000). Therefore, this study examined the variables that contribute to the mathematics performance of immigrants in Europe based on the PISA 2018 data and native students to provide insight into what accounts for achievements. This study is unique because it includes numerous country-level variables such as the Migrant Integration Policy Index (MIPEX), the Human Development Index (HDI), and government expenditure on education. These country-level variables are critical for comprehending how social, economic, and immigration policies may influence educational outcomes. As a result, the study is an important resource for educational stakeholders seeking to understand the impact of policies on educational results. Moreover, student-level variables such as exposure to bullying, sense of belonging, resilience, and cognitive flexibility were included to investigate how much variance in performance is explained by these variables. These variables are important because they may provide information about how students perform academically. Understanding which variables predict performance might aid in developing strategies or interventions to enhance student outcomes. Overall, the findings of this study are intended to understand the factors related to the mathematics performance of immigrant and native students in Europe, as well as provide significant insights for policymakers.

1.8. Research Questions

RQ1: Which student-level (motivation to master tasks, resilience, and cognitive flexibility/adaptivity, exposure to bullying, sense of belonging, discriminating school climate, students' attitudes toward immigrants) and country-level (migrant integration policy index and human development index, governmental expenditure on education) variables could predict mathematics performance of immigrant students and native students across European countries in PISA 2018?

RQ2: Is there a statistically significant difference between the mathematics performance of immigrant students and native students across European countries in PISA 2018?

RQ3: Is there a statistically significant difference between the mathematics performance of first-generation and second-generation immigrant students across European countries in PISA 2018?

RQ4: Is there a statistically significant difference between the mathematics performance of immigrant students and native students after controlling economic, social, and cultural status and gender?

RQ5: Is there a statistically significant difference between the mathematics performance of first-generation and second-generation immigrant students after controlling economic, social, and cultural status and gender?

2. LITERATURE REVIEW

2.1. Performance Gap Between Immigrant and Native Students

Pivovarova and Powers (2019) investigated the performance differences between immigrants and third-plus generation peers in PISA 2012 for mathematical literacy. Their study investigated the students and school background to see how these variables relate to performance. Their study participants were first, second-generation immigrants and third-plus-generation peers in the United States. Multiple regression models were used to assess variables. Student-level variables were chosen as students' generation status (first- or second-generation and third plus), gender, race, parental education, wealth index, a language other than English spoken at home, and grade. Since school factors may affect student achievement, school and class size, student-teacher ratio, school level (public, urban, suburban, rural), teacher shortage, shares of mathematics teachers with degree (bachelor or master), dropouts' rate larger than 10%, and student climate were included as school factors. The descriptive statistics showed that first- and second-generation students resembled the characteristics of schools attended. If the students' backgrounds were accounted for, the achievement difference between the first and third plus generations faded away. Moreover, it was found that if the students' backgrounds were controlled, the second generation performed slightly better than the third plus generation. Pivovarova and Powers (2019) have found that student background variables of wealth and parental education have a role in the mathematics achievement of first-generation immigrant students rather than their third-plus generation. Parental education and wealth variables positively correlated with mathematics achievement, indicating that students from advantageous families have had higher mathematics achievement (Pivovarova and Powers, 2019).

Martin *et al.* (2012) conducted a study to investigate the relationship between problem-solving skills, mathematics, and science performance of immigrant students in PISA with multiple countries, including Australia, Austria, Belgium, Canada, Denmark, France, Germany, Luxembourg, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United States, and the partner countries Hong Kong– China, Macao–China and the Russian Federation. Socio-demographic and settlement factors were included to see the role of these factors on performance. Settlement factors were chosen as follows: language at home, language proficiency, and age at immigration. For socio-demographic factors, gender, school, country, and first- and second-generation were included. As a result, Martin *et al.* (2012) found that the relationship between first-generation status and problem-solving skills varies significantly as a function of country and school, while the relation between second-generation status and problem-solving skills varies by school.

Martin *et al.* (2012) found that problem-solving skills are linked with achievement after controlling socio-demographic factors (gender, school, country, first- and second-generation immigrants) and settlement factors (language at home and language proficiency, age at immigration). Also, it was found that first-generation students' problem-solving skills strongly mediate the relation to achievement (Martin *et al.*, 2012). Furthermore, it was found that males are better at achievement and problem-solving skills, and older students and students with high socioeconomic status have higher problem-solving skills and achievement (Martin *et al.*, 2012).

In conclusion, Pivovarova and Powers (2019) examined the association between student and school background variables and mathematics performance in PISA 2012 and concluded that wealth and parental education positively correlate with mathematical proficiency for first-generation immigrant students. On the other hand, Martin *et al.*, (2012) looked into the relationship between immigrant students' performance in PISA across several countries and their problem-solving, math, and science abilities. They discovered that the link varied greatly by school and country. After adjusting for socio-demographic and settlement characteristics, problem-solving abilities were associated with achievement, according to Martin *et al.* (2012). Both studies have revealed a correlation between socioeconomic status and achievement, showing positively.

2.2. Factors Related to Achievement

There has been a lot of research has been done on the factors related to the achievement of immigrant students. Some studies emphasize immigrant's resilience (Rodriguez *et al.*, 2020), while others focus on exposure to bullying (Karakus *et al.*, 2022; Ponzo, 2013).

Rodriguez *et al.* (2020) investigated the differences in well-being indicators (life satisfaction, positive affect, self-efficacy-resilience, and a sense of belonging to the school) and academic performance in mathematics and science measured in PISA 2018. They have found that positive affect, self-efficacy resilience, and sense of belonging to school are statistically different between natives and first- and second-generation immigrants (Rodriguez *et al.*, 2020). Post-hoc analysis showed that native students had higher scores than immigrant students on three indicators of well-being such as positive affect, self-efficacy-resilience, and a sense of belonging to the school (Rodriguez *et al.*, 2020). In addition, first-generation students had lower scores than second-generation immigrant students in positive affect and sense of school belonging (Rodriguez *et al.*, 2020).

Some research indicates that experiencing bullying at school has a negative effect on students' academic performance (Karakus *et al.*, 2022; Ponzo, 2013). Ponzo (2013) claims that those who experience bullying at school perform significantly worse on reading comprehension tests made by Progress in International Reading Literacy Study (PIRLS) in 2006. Moreover, they argue that students bullied at school do much worse than their non-victim classmates in the fourth and eighth grades in mathematics and science in TIMSS 2007 (Ponzo, 2013).

Ponzo (2013) argues that home items linked to family wealth and book ownership have a positive correlation with student performance. A recent study found that immigrant students perform better academically across mathematics, science, and reading in PISA 2018 when they come from a higher ESCS background (Karakus *et al.*, 2022).

In this study, countries' gross domestic product (GDP) per capita was utilized as a potential indicator of performance differences between immigrant students and their native peers. Therefore, as a country-level variable, only GDP per capita was included. Notable variances in the immigrant students' performance were found at the country level; however, these differences were not accounted for by the GDP per capita (Karakus *et al.*, 2022). On the other hand, He, Vijver and Kulikova (2017) have found that GDP per capita and educational expenditure positively correlated with students' achievement in TIMSS and PISA. In addition to this, MIPEX is positively associated with PISA performance (Arıkan *et al.*, 2017; Arikan *et al.*, 2020; He *et al.*, 2017). However, Dronkers and de Heus (2016) argue that the results of the science performance of students from immigrant backgrounds were unaffected by the immigration-related policies of the destination countries as measured by MIPEX. Since there is a mixed result, MIPEX is included as a country-level variable in this study. On the other hand, some research claims that the human-development index is associated with achievement (Arikan *et al.*, 2020), while others did not find an association between achievement and the human development index (Arıkan *et al.*, 2017). Therefore, the human development index is included to investigate its association with achievement.

2.3. Mathematics Performance Explained by ESCS

Arikan *et al.* (2017) examined a single immigrant group in a variety of nations to investigate the relationship between the nation's integration policies and the achievement gap between mainstream and immigrant group. As student-level factors gender, immigrant status, and an index of ESCS were chosen.

As country-level factors, multicultural policies were chosen to have knowledge of difference of performances in European countries. To compare different nations' migrant policies, the Migrant Integration Policy Index (MIPEX) was chosen for this study. Moreover, the human development index developed by United Nations Development Program (UNDP), the education domain score of MIPEX, the general integration score, anti-discrimination score and the school stratification index were chosen. The school stratification index indicates the age of selection to higher education and the opportunity to different options to higher education.

Arikan *et al.* (2017) investigated the effects of these factors on performance in reading and mathematics in PISA. They claimed that immigrant Turkish students had lower scores in reading and mathematics than mainstream European students. However, Cohen's d had a median value of 0.65 for reading and 0.58 for mathematics after controlling for ESCS, indicating that differences in economic, social, and cultural status had an important impact on the gap in reading and mathematics performance between mainstream European students and immigrant Turkish students. In addition to this, it is claimed that immigrant Turkish students living in countries whose national policies are more inclusive tend to show higher scores. They had chosen immigrant participants according to the categories such as taking the test out of Turkey and either born in Turkey or their parents were born in Turkey or Turkish language used at home. As a result, it was found that immigrant Turkish students had lower scores in reading and mathematics than European mainstream students. At the student-level, gender, immigrant status and ESCS index were found significant in the prediction of performance. Girls scored higher in reading, while boys scored higher in math. When the index of economic, social, and cultural status controlled, the achievement gap decreased but still remained. At the country level, there were no country factors related significantly to mathematics performance. On the other hand, the education domain score was found statistically significant. It was implied that the reading score of students in countries whose educational integration is better was higher (Arikan *et al.*, 2017).

Karakus *et al.* (2022) investigated the immigrant students reading, math, and science performance in PISA 2018. Student characteristics and experiences, as well as educational resources and policies were investigated. As a country level variable, GDP per capita was chosen, while school level variables were chosen as follows: school type, creative extracurricular activities' number, educational material's shortage, educational staff's shortage (Karakus *et al.*, 2022). In addition to these, immigrant school-policy related variables were included. On the other hand, student level variables such as gender, immigrant status, ESCS, parental emotional support, teacher interest, teacher adaptivity of instruction based on students' needs, experiencing bullying were included (Karakus *et al.*, 2022). Students' experiences with bullying appeared to have a substantively negative impact, while being male and having a higher ESCS appeared to have a substantively positive impact on differences in mathematics ability within schools. Although less significant, the adaptation of classroom instruction, teachers' interest in students, and immigrant parents' emotional support all seem to have positive and statistically significant effects on mathematics performance of immigrant students. Moreover, extracurricular activities' number statistically significant impact on mathematics performance of immigrant students. No policies of school did not have an effect on mathematics performance. Karakus *et al.* (2022) claims that parental emotional support may be especially beneficial for 1G immigrant students' mathematics performance. Similar results were found for reading and science performance of immigrants (Karakus *et al.*, 2022). On the other hand, Karakus *et al.* (2022) argue that the effect of school type is not significant indicating that regardless of the type of school a school's average socioeconomic status is a more significant indication of performance.

The studies of Arikan *et al.* (2017) and Karakus *et al.* (2022) differs regarding the variables since Arikan *et al.* (2017) included country-level variables while Karakus *et al.* (2022) included both school level variables and one country level variable different than Arikan's study. Arikan *et al.* (2017) have found that students' performance was predicted by gender, immigrant status, economic, social, and cultural index, and education domain score of MIPEX. In contrast, Karakus *et al.* (2022) reported that bullying, being male, and having a higher ESCS had significant effects on mathematics performance. Both studies used PISA data, but in different years, and they found similar results, indicating that ESCS and gender are related to mathematics performance.

2.4. Association of MIPEX and HDI with Mathematics Performance

Arikan *et al.*, (2020) investigated the difference in mathematics achievement between immigrant and native students in fourth grade in a variety of European countries included in TIMSS. Arikan *et al.* (2020) defined immigrants in two categories such as first-generation and second-generation. The achievement difference between the first and second-generation was investigated. While most research focuses on the high school immigrant students' comparison, in this research the participants were 4th-grade students. Student liking of mathematics, home resources for learning, experience bullying, confidence in math, and sense of belonging were chosen as student-level background variables. The Human Development Index (HDI), the Migrant Integration Policy Index (MIPEX), and the percentage of total expenditure for education were chosen as country-level variables. In this study, home resources for learning are controlled to examine mathematics achievement differences. The achievement difference was measured with and without controlling home resources for learning variable. As a result, they found that the difference between immigrants and mainstream students became non-significant when home resources for learning is controlled. In addition to this, experienced bullying, liking of learning math, and confidence in mathematics were found to be significant predictors of mathematics achievement of immigrant students. At the country-level, significant relations were found for only the HDI. Other country-level variables were found non-significant in mathematics achievement. On the other hand, for Turkish immigrants, higher MIPEX and HDI were associated with higher mathematics achievement. On the other hand, the mathematics achievement difference between first- and second-generation did not show a systematic difference (Arikan *et al.*, 2020). In general, immigrant students lag behind mainstream students in the fourth grade.

Campbell, McIntyre and Kucirkova (2021) examined the long-term relationship between the HDI and PISA performance between 2006 and 2018. A positive correlation between the HDI and PISA scores have found. All countries with a score less than 420 had a lower human development index. Campbell *et al.* (2021) claim that the HDI is dependent on PISA scores. Despite having a strong association with PISA scores at every time point and exhibiting persistent positive change during the 12 years of this research, change in the HDI was not found to be a statistically significant predictor of change in PISA scores (Campbell *et al.*, 2021).

Mazurek, Garcia and Rico (2021) used data from 70 countries to investigate the effect of various types of inequality on PISA reading, mathematics, and science performance. The effects of gender, wealth, and educational inequity were investigated using indices such as GDP per capita, government expenditure on education, gender inequality index, Gini coefficient HDI, and inequality in education index. While the index of income inequality is not statistically significant, the index of gender inequality and the index of educational inequality are. Furthermore, Mazurek *et al.* (2021) discovered that GDP per capita, the HDI, and government spending on education are statistically significant for OECD countries although insignificant for non-OECD countries. On the other hand, GDP per capita and expenditure on education are in a negative direction, indicating that higher expenditure on education is associated with lower performance. According to Mazurek *et al.* (2021), money is squandered because economic resources are employed inefficiently.

2.5. Relationship between Motivation and Achievement

Motivation is one of the topics researchers have focused on regarding its relation to achievement. McClelland's early work defines academic achievement goal theory, and academic achievement can be explained as the urge to work toward rewards, including approval from others, fulfillment, and feelings of mastery of oneself (McClelland, 1985). On the other hand, Elliot and McGregor (2001) highlights a framework of achievement goal consisting of two categories: mastery goals and performance goals. Moore *et al.*, (2010) have found a strong correlation between achievement goal, including mastery and performance, and academic achievement in English and mathematics. On the other hand, Huang (2012) has found an association between academic achievement and mastery motivation but not with performance motivation. Furthermore, Korpershoek (2016) investigated the relationship between motivation, school commitment, and achievement. Mastery motivation was not related to achievement, while mastery motivation was correlated with school commitment (Korpershoek, 2016). As a whole, research on motivation focuses on different dimensions of motivation and their relations to achievement. Some studies have found that mastery and performance goals correlated with achievement (Elliot and McGregor, 2001), while others have found a relation with motivation only for one dimension, changing mastery or performance.

3. METHODOLOGY

3.1. Participants

The target population of PISA is students at the age of 15 when most young people are still enrolled in formal education, so PISA evaluates the overall effects of education and learning (OECD, 2019b). One of the reasons for defining the target population based on age is that it is independent of the institutional frameworks of national educational systems (OECD, 2019b). Students who were aged between 15 years and three months and 16 years and two months at the start of the assessment period, plus or minus a permitted 1-month variation, and enrolled in a school at grade 7 or higher were included in PISA 2018 (OECD, 2019b).

3.1.1. Immigrant Status

According to OECD (2019b), students' country of birth and their parents' country of birth were gathered in PISA 2018. The immigrant background index was constructed by using the information of birth country gathered from students and their parents. This index includes three categories: first-generation students, second-generation students, and native students. According to OECD (2019b), students born in another country and whose parents were born in another country are considered first-generation. Second-generation students are those born in the country of assessment but whose parents were born elsewhere. Native students are those whose parents (at least one) were born in the assessment country (OECD, 2019b). Missing responses indicate that either student's or both parents' information is missing (OECD, 2019b).

Based on this definition, the numbers, and percentages of first- and second-generation students and native students across European countries were shown in Table 3.1. According to Hox (2013), the greater number of groups is suggested for the multilevel regression analysis. However, in the current study, the number of groups is less so that groups with 100 immigrants is included in the sample for multilevel regression analysis. Croatia, Estonia, Germany, Greece, Iceland, Italy, Ireland, Latvia, Lithuania, Malta, Portugal, Serbia, Slovenia, Spain, and Switzerland were included in the multilevel regression analysis.

Table 3.1. Percentage of immigrant students in European Countries in PISA 2018.

	First-generation		Second-generation		Native	
Country	N	%	N	%	N	%
Albania	17	0.3	20	0.3	6170	97.0
Austria	474	7.0	933	13.7	5303	78.0
Belgium	661	7.8	812	9.6	6782	80.0
Bosna and Herzegovina	90	1.4	87	1.3	6068	93.6
Bulgaria	28	0.5	29	0.5	4911	92.8
Czech Republic	123	1.8	128	1.8	6674	95.1
Croatia	87	1.3	511	7.7	5882	89.0
Denmark	269	3.5	1282	16.7	5858	76.5
Estonia	41	0.8	498	9.4	4635	87.2
France	336	5.3	611	9.7	5220	82.8
Finland	179	3.2	134	2.4	5211	92.2
Germany	299	5.5	751	13.8	3677	67.5
Greece	173	2.7	529	8.3	5570	87.0
Iceland	99	3.0	80	2.4	3025	91.8
Italy	499	4.2	572	4.9	10283	87.3
Ireland	529	9.5	429	7.7	4454	79.9
Malta	212	6.3	69	2.1	2955	87.9
Montenegro	147	2.2	236	3.5	6121	91.8
Latvia	47	0.9	197	3.7	4936	93.1
Lithuania	34	0.5	114	1.7	6558	95.3
Luxembourg	1249	23.9	1549	29.6	2318	44.3
Netherlands	143	3.0	518	10.9	3951	82.9
Norway	347	6.0	348	6.0	4882	84.0
Poland	11	0.2	16	0.3	5523	98.2
Portugal	140	2.4	202	3.4	5254	88.6
Serbia	69	1.0	529	8.0	5810	87.9
Slovenia	323	5.0	250	3.9	5730	89.5
Spain	2507	7.0	1610	4.5	30727	85.5
Sweden	499	9.1	556	10.1	4283	77.8
Switzerland	692	11.9	1234	21.2	3731	64.1
Turkey	17	0.2	37	0.5	6663	96.7
United Kingdom	877	6.3	853	6.2	11249	81.4

3.1.2. Sampling Methods in PISA

Two-stage stratified sampling was used in PISA (OECD, 2018b). First of all, schools whose students at the age of 15 were sampled. The school sampling frame, a complete nationwide list of all schools appropriate for PISA, was used to systematically sample schools so that the probability should be proportionated to the measure of size. The measure of size is defined as the result of the number estimation of 15-year-old students registered in schools. This specific kind of sampling is known as systematic probability proportional to size (PPS) sampling (OECD, 2018b).

Schools were stratified before the sampling process. One of the benefits of stratification is that it increases the effectiveness of the sampling design so that the reliability of the estimates of the survey increases (OECD, 2018b). In addition, schools from various strata ensure that the sample represents the population. Moreover, stratification ensures that the sample sufficiently represents each relevant category in the target population. Stratification takes place either implicitly or explicitly. Explicit stratification involves classifying schools handled separately, such as schools in a region or a state. Implicit stratification classifies schools within an explicit stratification utilizing some implicit variables such as type of school or school size (OECD, 2018b).

After the schools were sampled, the students were sampled in the second stage. Forty-two students are required for each school if the country takes the assessment as computer-based and takes the global competency assessment. If a country administers the assessment as paper-based or computer-based but does not take the global competence assessment, 35 students are required for each school. Students were sampled with equal probability if a school contains more than 42 students or 35 for some countries.

A major analytical goal of PISA requires that the number of students chosen for each school be at least 25 to guarantee appropriate accuracy in evaluating variance within and across schools. At least 150 schools should be sampled for each country. If a country has less than 150 schools, all of them are sampled. At least 6300 students are sampled in countries that use computer-based assessments. In countries that employ paper-based or computer-based tests but do not use the global competence assessment, 5250 students are sampled as a minimum (OECD, 2018b).

3.2. Instruments

3.2.1. Student-level Variables

Based on Walberg's theory of academic achievement, motivation to master tasks, resilience, cognitive flexibility/adaptivity, exposure to bullying, sense of belonging, discriminating school climate, and students' attitudes toward immigrants were included as student-level variables. The following section explains how these variables are measured.

3.2.1.1. Motivation to Master Task (WORKMAST). The index of motivation to master task (WORKMAST) was constructed with the three statements asked students in PISA 2018. The statements are: "Once I start a task, I persist until it is finished"; "I find satisfaction in working as hard as I can"; "Part of the enjoyment I get from doing things is when I improve on my past performance" (OECD, 2019d, p. 216). Students rate these statements, whether they agree or not, using a 4-point Likert scale (OECD, 2019d).

3.2.1.2. Resilience (RESILIENCE). The resilience index was constructed with the four statements in which students rated if they agreed or disagreed using a 4-point Likert scale (OECD, 2018b). The statements are: "I usually manage one way or another"; "I feel proud that I have accomplished things,"; "I feel that I can handle many things at a time,"; "My belief in myself gets me through hard times"; "When I'm in a difficult situation, I can usually find my way out of it" (OECD, 2019b, p. 189).

<u>3.2.1.3. Cognitive Flexibility/Adaptivity (COGFLEX).</u> The items of this scale were chosen from Martin and Rubin (1995), and Dennis and Vander Wal (2010), and OECD's experts made minor word changes (OECD, 2018b). The elements of flexibility/adaptability are intercultural adaptation and adaptation to unfamiliar situations. The first four items are written for the element of adaptation to unfamiliar situations, such as “I can deal with unusual situations”; “I can change my behavior to meet the needs of new situations”; “I can adapt to different situations even when under stress or pressure”; “When encountering difficult situations with other people, I can think of a way to resolve the situation”. The last item developed for “intercultural adaptation” is “I am capable of overcoming my difficulties in interacting with people from other cultures”. Five-point Likert scale was used. Options are “Very much like me”, “Mostly like me”, “Somewhat like me”, “Not much like me” and “Not at all like me”. Prior to scaling, all items were reverse-coded (OECD, 2020, p. 101).

<u>3.2.1.4. Exposure to Bullying (BEINGBULLIED).</u> Students were asked how frequently they had experienced school-related experiences in the 12 months prior to the PISA assessment, involving experiences that occur in social media: ”Other students left me out of things on purpose,”; ”Other students made fun of me,”;” I was threatened by other students,” (OECD, 2019b, p.214). Four-point Likert scale was used including ratings that are “once a week or more”, “a few times a month”, “a few times a year”, “never or almost never” (OECD, 2019b, p.214).

In order to create the index of exposure to bullying, these statements were merged so that a positive value means that the student experienced bullying at school to a greater extent than the average (OECD, 2019b).

<u>3.2.1.5. Sense of Belonging (BELONG).</u> The index of sense of belonging was developed based on students' answers to questions about the sense of belonging to school (OECD, 2019c). Students were asked to assess how much they agreed with statements such as “I feel like an outsider (or left out of things) at school; I make friends easily

at school; I feel like I belong at school; I feel awkward and out of place in my school; Other students seem to like me; and I feel lonely at school" (OECD, 2019c, p. 217). Three of these items were reversely coded, indicating that positive results on this scale show that students felt more sense of belonging at school (OECD, 2019c).

3.2.1.6. Discriminating School Climate (DISCRIM). This scale questions the environment of diversity in schools. It queries whether teachers show respect for students from various cultural backgrounds. Thus, it examines the absence of discrimination, bias, and stereotyping. The items are "They have misconceptions about the history of some cultural groups"; "They say negative things about people of some cultural groups"; "They blame people of some cultural groups for problems faced by ¡country of test¿", "They have lower academic expectations for students of some cultural groups." How most of the statements in the question apply to their teachers are asked to students (OECD, 2020, p. 215). The options are "To none or virtually none of them," "To some of them," "To most of them," and "To all or almost all of them" on a four-point Likert scale (OECD, 2020).

3.2.1.7. Students' Attitudes Toward Immigrants (ATTIMM). This scale comprises four questions based on the International Civic and Citizenship Education Study (Schulz *et al.*, 2011). It asks students about their attitudes toward immigrants' equal rights. Students rate the following statements if they agree or not. Four point-Likert scale was used. The statements are "Immigrant children should have the same opportunities for education that other children in the country have"; "Immigrants who live in a country for several years should have the opportunity to vote in elections"; "Immigrants should have the opportunity to continue their own customs and lifestyle"; "Immigrants should have all the same rights that everyone else in the country has" (OECD, 2020, p. 105). As can be seen in Table 3.2, Cronbach's alpha values for variables of each country were reported (OECD, 2018b). A value less than 0.70 is considered unacceptable, while a value between 0.70 and 0.79 is considered fair. It is considered good if it is between 0.80 and 0.90 and excellent if it is greater than 0.90 (Cicchetti, 1994).

As indicated in Table 3.2, motivation to master tasks had a reliability coefficient ranging from 0.66 to 0.80. Resilience reliability coefficients varied from 0.71 to 0.87. Cognitive flexibility/adaptability's reliability coefficient ranged from 0.70 to 0.85. The range of the reliability coefficient for exposure to bullying was 0.77 to 0.87. The reliability coefficient for the sense of belonging ranged from 0.82 to 0.90. Moreover, the Cronbach alpha values for discriminating school climate ranged from 0.81 to 0.91. The reliability coefficient for students' attitudes towards immigrants varied from 0.81 to 0.92. These high-reliability coefficients indicates that the variables are reliable so that it can be included in the study.

Table 3.2. Cronbach alpha values of student-level variables.

Country	WORK-MAST	RESIL-IENCE	COG-FLEX	BEING BULL-IED	BEL-ONG	DIS-CRIM	ATT-IMM
Croatia	0.73	0.77	0.81	0.83	0.88	0.88	0.88
Estonia	0.67	0.78	0.71	0.81	0.86	0.81	0.85
Germany	0.66	0.73	0.71	0.80	0.84	0.85	0.82
Greece	0.71	0.76	0.71	0.80	0.80	0.84	0.80
Iceland	0.80	0.87	0.73	0.87	0.90	0.91	0.92
Italy	0.75	0.71	0.80	0.79	0.84	0.82	0.80
Ireland	0.71	0.73	0.75	0.83	0.82	0.89	0.84
Latvia	0.70	0.74	0.70	0.82	0.83	0.81	0.84
Lithuania	0.73	0.78	0.85	0.77	0.88	0.86	0.88
Malta	0.77	0.78	0.80	0.78	0.85	0.84	0.85
Portugal	0.72	0.74	0.78	0.81	0.89	0.84	0.83
Serbia	0.76	0.83	0.84	0.79	0.88	0.87	0.85
Slovenia	0.75	0.80	0.79	0.80	0.86	0.86	0.83
Spain	0.71	0.76	0.78	0.85	0.87	0.88	0.81

3.2.2. Country-Level Variables

For country-level variables, MIPEX, HDI, and governmental expenditure on education were chosen. The construction of indices and the definition of each variable were explained in the following section.

3.2.2.1. Migrant Integration Policy Index (MIPEX). The MIPEX is a specialized tool that evaluates and integrates policies for migrants in nations on five continents, including 52 countries. This index can be used to evaluate, compare, and develop integration policy. It portrays a range of integration policies across a wide variety of diverse situations. MIPEX includes eight areas: Education, Health, Anti-discrimination, Labor Market Mobility, Family Reunion, Political Participation, Long-term Residence, and Access to Nationality. Standards are taken from Council of Europe Conventions, European Union Directives, and International Conventions for these areas. These eight areas are evaluated by a policy indicator which is a question on a particular policy component of one of the eight areas. The policy scores ranged from 0 to 100. The scores of eight areas are averaged for each country. The latest MIPEX scores of countries were calculated in 2019, and scores were listed in Table 3.3 (Solano and Huddleston, 2020). Data were only available in 2019, thus the data included in this study are the only available data.

According to Solano and Huddleston (2020), education score on MIPEX is measured by eight policy indicators such as access to both mandatory and non-mandatory education, higher education accessibility, support for learning the language of education, education counseling at all levels, teacher education to meet the learning needs of migrants, diversity-focused teacher education, measures to enhance migrant populations' educational condition, perspective of diversity in the curriculum, and efforts to increase the number of immigrant teachers (Solano and Huddleston, 2020).

3.2.2.2. Human Development Index (HDI). Human Development Index has three dimensions: knowledge, healthy life, and a decent standard of living (UNDP, 2014). The indicator of long and healthy life is the life expectancy at birth. Life expectancy is defined as the number of years a newborn child should anticipate to live if the current trends in the rate of age-related mortality at birth stay the same throughout the infant's life (UNDP, 2014). This dimension is assessed by using the life expectancy index. The indicators of knowledge are expected years of schooling and mean years of schooling. Expected years of schooling are defined as the number of years of education that a child entering school can anticipate receiving if current trends in age-specific enrollment rates continue throughout the child's life. Mean years of schooling are the average years of schooling obtained by students aged 25 and older. The education index was used to measure the knowledge dimension. A decent standard of living is measured by Gross National Income (GNI). These three indices construct the HDI (UNDP, 2014). The score of the HDI ranges from 0 to 1. The scores are divided into four categories such as low human development (below 0.55), medium human development (0.55-.70) and high human development (0.7-0.79), and very high human development (0.8-1.0). The most recent data on this index was published in 2019. The latest HDI scores of each country were shown in Table 3 (UNDP, 2014).

3.2.2.3. Government Expenditure on Education. Government expenditure on education is shown as a share of all general government spending across all sectors, such as education, social services, and health. It is computed by dividing total government spending on education by total government spending on all sectors, then multiplying the result by 100. The data source is UNESCO Institute for Statistics (World Bank, 2022b). The latest available data was used in 2018. The reason for using 2018 data is that the data was not available in some countries in 2019. Data was available for all countries included in the study in 2018. The data from 2018 and 2019 did not differ greatly when compared, only the decimals differed.

According to Table 3.3, Estonia and Portugal have the highest education scores on MIPEX, and Latvia has the lowest. Regarding HDI, the highest score is in Iceland, Lithuania, and Switzerland, while the lowest is in Serbia. In terms of government expenditure on education, Iceland has the highest percentage, while Greece has the lowest.

Table 3.3. Government Expenditure on Education, MIPEX, and HDI scores.

Country	Education Score on MIPEX	HDI Score	Government Expenditure on Education (%)
Croatia	33	0.86	8.6
Estonia	69	0.89	13.4
Germany	55	0.94	11.2
Greece	36	0.89	7.4
Iceland	45	0.96	17.3
Ireland	45	0.95	13.3
Italy	43	0.90	8.8
Latvia	26	0.86	11.2
Lithuania	48	0.96	15.5
Malta	40	0.92	14.0
Portugal	69	0.87	10.8
Serbia	43	0.80	8.8
Slovenia	33	0.92	11.3
Spain	43	0.91	10.0
Switzerland	48	0.96	15.5

3.2.3. Control Variable: Economic, Social and Cultural Status

The index of economic, cultural, and social status was used as a control variable to investigate the difference between the mathematics performance of first-generation, second-generation, and native students if students have similar socio-economic status.

According to PISA 2018 Assessment and Analytical Framework (OECD, 2019d), most of the measures outlined in the PISA 2009 Initial Report were present in the

2018 instruments, allowing for a comparison of results between 2009 and 2018. To track trends in social, cultural, and economic factors, PISA 2018 retains questions about socio-economic status and other background variables mostly unaltered (OECD, 2019c). However, a few minor adjustments are required. For instance, significant advancements in Information and Communications Technology (ICT) have made products once only sometimes seen in students' households more prevalent and less discriminating as a measure of socio-economic position (OECD, 2019c). As a result, the measurements of household possessions will be adjusted to better account for within- and between-country heterogeneity (OECD, 2019c).

According to OECD (2010), the international socio-economic index of occupation status of the mother or father (whose status is higher), the level of education of the mother or father (whichever is greater) transferred into years of schooling, and the index of home possessions were the variables used to create the index of economic and social status. Home possessions were defined and asked participants as follows:

> "desk at which they studied at home", a room of their own, a quiet place to study, educational software, a link to the Internet, their own calculator, classic literature, books of poetry, works of art (e.g., paintings), books to help them with their school work, a dictionary, a dishwasher, a DVD player or VCR, three other country-specific items and the number of cellular phones, televisions, computers, cars and books at home" (OECD, 2010, p. 29).

3.2.4. Dependent Variable: Performance in Mathematics

Mathematical literacy is defined based on three major features such as processes, content, and context (OECD, 2013a). Contents are quantity, change and relationships, space and shape, uncertainty, and data. Contexts are categorized as scientific, occupational, social, and personal. Contexts are required to include items with different levels of difficulty appropriate for 15-year-old students (OECD, 2013a). There are three processes: formulate, employ, and interpret/evaluate. Formulating means acknowledging and finding possibilities for the usage of mathematics. It means recognizing real-world problems and making an explanation mathematically (OECD, 2013a).

Employing means using mathematical facts and procedures, reasoning, and concepts to make a conclusion. Interpreting means making sense of conclusions and solutions in a real-life context. This feature includes both interpreting and evaluating (OECD, 2013a). Item formats are multiple-choice, open-ended, and closed ended. Most items are scored dichotomously, and some open-ended items are scored partially (OECD, 2013a). On the other hand, to reduce the testing burden on individual students while ensuring precise population estimates of a complicated item administration, the rotated booklet design is utilized in PISA (Rutkowski *et al.*, 2010). The rotated booklet design helps to ensure that items are introduced to enough students in the sample. Although this administration technique lessens students' testing time, it makes it difficult to determine each student's proficiency. Traditional techniques for determining individual proficiency may not produce an accurate or consistent variance. As an effective alternative, a plausible value approach is used to produce population-level proficiency estimates from test designs in which a small portion of the overall items are presented to each student (Rutkowski *et al.*, 2010). Responses to the limited number of cognitive items given to the population of interest are used to produce a distribution of student ability. Each plausible value is chosen at random from a set of draws from the ability distribution for each student (Rutkowski *et al.*, 2010). There were ten plausible values in PISA 2018. Ten plausible values of mathematics were used in the present study.

3.3. Data Analysis

In order to answer the first research question, multilevel regression analysis was used to investigate which student-level and country-level variables could predict the mathematics performance of immigrant and native students. Croatia, Estonia, Germany, Greece, Iceland, Italy, Ireland, Latvia, Lithuania, Malta, Portugal, Serbia, Slovenia, Spain, and Switzerland were included in the multilevel regression analysis because some of the countries did not gather data on some student level variables such as cognitive flexibility/adaptivity in PISA 2018. Therefore, countries that collect data on student level variables were included in the multilevel regression.

For the second research question, an independent samples t-test was used to compare the performance of immigrant and native students. The sample weights and plausible values were included in the analyses to ensure unbiased results (IEA, 2022). In order to answer the third research question, propensity score matching was used first, and then related comparisons were performed to examine the performance gap between immigrant and native students after controlling for socio-economic status and gender.

3.3.1. Regression

Regression analysis is used to predict dependent variables using one or more independent variables by fitting a model to the data (Field *et al.*, 2012). Simple linear regression is used when a prediction is made based on just one independent variable (Field *et al.*, 2012). At the same time, multiple linear regression is used when the prediction is based on more than one independent variable (Field *et al.*, 2012). The regression analysis equation can be expressed as outcome = (model) + error (Field *et al.*, 2012). This equation tells that the results of a person can be predicted using a model and some error.

In linear regression, the model is a linear line that best describes the data. The least squares method is used to fit the data (Field *et al.*, 2012). Mathematically, a linear line has two parameters such as slope and intercept. The slope (b1) and intercept (b0) are referred to as regression coefficients (Field *et al.*, 2012). The slope indicates the nature of the relationship in the data while intercept shows the location of the line. The main aim of regression is to find the best line to describe data and make an estimation of the regression coefficient of the line (Field *et al.*, 2012). In order to assess the goodness of fit, R2 is used, and it indicates how much variance is explained by the model (Field *et al.*, 2012).

As previously mentioned, a regression model's predictor contains a coefficient (b1) that symbolizes the gradient of the regression line representing a unit change in the predictor that causes an outcome to change (Field *et al.*, 2012). A predictor variable should have a b-value that is considerably different from zero if it significantly predicts an outcome. This hypothesis is tested using a t-test (Field *et al.*, 2012).

Multiple regression is the extension of simple regression by using a variety of predictor variables (Field *et al.*, 2012). For every predictor variable added, the coefficient is added to the regression equation. Since multiple regression includes more than one predictor variable, it is important to decide which predictors to use (Field *et al.*, 2012).

3.3.2. Multilevel Regression Analysis

In real life, data are typically hierarchical, and hierarchical data shows that some variables are nested inside other variables (Field *et al.*, 2012). For instance, data can be gathered from students, which is the bottom hierarchically referred to as the level 1 variable (Field *et al.*, 2012). The students belong to classrooms indicating that classes are at a higher level of the hierarchy, and it is referred to as the level 2 variable or contextual variable (Field *et al.*, 2012). A new level should be introduced to the hierarchy if data are gathered from several schools and classrooms from the schools (Field *et al.*, 2012). Hierarchical data is essential because contextual variables create dependence on the data (Field *et al.*, 2012). This implies the correlation between the residuals because of the dependency of the observations (Field *et al.*, 2012). Intraclass correlation is used to evaluate the degree of dependency among observations (Field *et al.*, 2012). The intraclass correlation estimates the share of the outcome's total variability that can be due to level two (Field *et al.*, 2012).

The intraclass correlation shows the percentage of the overall variation that can be accounted for by the second level of the hierarchy (Hox *et al.*, 2017). The intraclass correlation is defined as the ratio of the group second-level variance to the total variance. The formula of ICC is:

$$\rho = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_e^2}. \tag{3.1}$$

High intraclass correlation values indicate that the grouping level matters and observations are similar within groups (Tabachnick and Fidell, 2007).

In this study, multilevel analysis was used. The dependent variable in the present study is mathematics performance measured by PISA 2018. As student-level variables, motivation to master tasks, resilience, and cognitive flexibility/adaptivity, exposure to bullying, sense of belonging, discriminating school climate, students' attitudes toward immigrants were used. The MIPEX, HDI, and government expenditure on education were included as country-level variables. The intraclass correlation was used to indicate the variation in immigrant students' mathematics performance explained by country-level and student-level differences. In addition to this, R^2 was used to show the percentage of variances in mathematics performance of immigrant students explained by student level variables and country level variables. For the multilevel regression analyses, MPLUS 7.4 was used.

3.3.2.1. Assumptions of Multilevel Regression Analysis. In multilevel regression analysis, linearity, homoscedasticity, and independence of residual error in different levels are assumed (Hox, 2013). Random coefficients are assumed to have normal distributions (Field *et al.*, 2012). The normally distributed error shows that differences bigger than zero occasionally occur and that most of the time, there are no differences between the model and the observed data (Field *et al.*, 2012). The linearity of the relationships indicates that outcome and predictor variables make a straight line indicating that the model is linear. Nonlinear models are obstacles to the results' generalizability (Field *et al.*, 2012). Homoscedasticity means that the variance of the residuals in every level of the predictor(s) is the same (Field *et al.*, 2012).

In addition to this, the sample size is significant, especially in the second or higher level. An unequal sample size is not problematic because an equal sample is not an assumption (Hox, 2013).

3.3.3. Independent Samples T-Test

An independent-sample t-test is used when there are two experimental conditions and different individuals are placed in each condition (Field *et al.*, 2012). T-test is also used for comparison between means of samples (Field *et al.*, 2012). In addition to this, standard error is used to measure variance between sample means (Field *et al.*, 2012). T value is calculated by dividing the difference of sample means into standard error.

There are some assumptions about the t-test. Scores of the samples are expected to be normally distributed. Homogeneity of the variances is one of the assumptions indicating that samples' variances are equal. In addition to this, when participants are chosen randomly for a sample, two samples are considered independent (Heiman, 2011). The scores are considered independent implying that the likelihood of a specific score occurring in one sample is unaffected by the scores occurring in the other (Heiman, 2011).

In this study, multiple independent samples t-test was used to test if a statistically significant difference exists between the mathematics performance of immigrant students and native students. The mathematics performance of immigrant and native students was compared. Then, second-generation and first-generation immigrant students' mathematics performance was compared. After that, t-scores and effect sizes were presented. IDB Analyzer was used. On the other hand, applying multiple t-tests may increase the chance of type 1 error. The Bonferroni adjustment is used to lower the likelihood of receiving false-positive findings (Napierala, 2012). The adjustment is made by dividing the p-value by the number of t-tests (Napierala, 2012). Therefore, the correction was made by dividing the p-value (0.5) by the number of t-tests (2) in the current study.

3.3.4. Propensity Score Matching

Matching is a way of sampling from a large pool of samples to establish a control group of reasonable size with a distribution of covariates comparable to the distribution in the treatment group (Rosenbaum and Rubin, 1983). Propensity score matching is the process of pairing control and treatment units with matching propensity score values, along with other covariates and rejecting any unmatched units (Rubin, 2001). In propensity score matching, various methods are used, such as exact matching, subclassification, nearest neighbor, and optimal. According to Randolph and Falbe (2014), treated units are matched exactly with control units that have identical covariate values in exact matching. Exact matching might not be feasible when several variables or covariates can take a wide range of values (Randolph and Falbe, 2014). On the other hand, the goal of the optimal matching approach is to reduce the average absolute distance between all matched pairs (Randolph and Falbe, 2014). Through subclassification, the data set is divided into subclasses such that each subclass is similar in its covariates (Randolph and Falbe, 2014). Another method, the nearest neighbor method, assigns each treatment unit the closest control unit according to their distance measured with a logarithmic function. The nearest neighbor is preferred in the current study because it allows the control units to be as close to the treatment unit as possible, reducing selection bias. This method also reduces the bias of the estimation of the treatment effect by eliminating the differences in the covariates between the treatment and control group (Randolph and Falbe, 2014).

Nearest neighbor propensity score matching was used in the current study to investigate the performance difference between immigrant and native students after ESCS and gender had been controlled. The scores of ESCS and gender were matched for immigrant and native groups, so the groups are similar regarding ESCS and gender. The MatchIt R package (Ho *et al.*, 2011) was used for nearest neighbor matching propensity score matching. Then, the performance of immigrant and native students was compared by applying the t-test. IDB Analyzer was used for the t-test. The effect size of the performance difference before and after matching was calculated. Effect size shows the magnitude of the difference between groups (Cohen, 1988). Cohen's d, an effect size index, is found by dividing the mean difference between two samples into their standard deviation (Cohen, 1988). Cohen's d was calculated before and after controlling economic and social status. If the d value is 0.8, it signifies a big effect size. A value of 0.5 indicates a medium effect size, while 0.2 suggests a small effect size (Cohen, 1988).

4. RESULTS

4.1. The Evaluation of Assumptions

4.1.1. Linearity, Independence of Errors, and Homoscedasticity

Among the methods employed to check linearity, homoscedasticity, and independence of errors are the scatterplots of the standardized predicted values (ZPRED) and the standardized residuals (ZRESID) (Field, 2009). As there were four multilevel regression analyses, the scatterplots for each analysis were checked. The scatterplots were provided below. Since the histograms did not indicate any violation of linearity, independence of errors, or homoscedasticity, all assumptions were met.

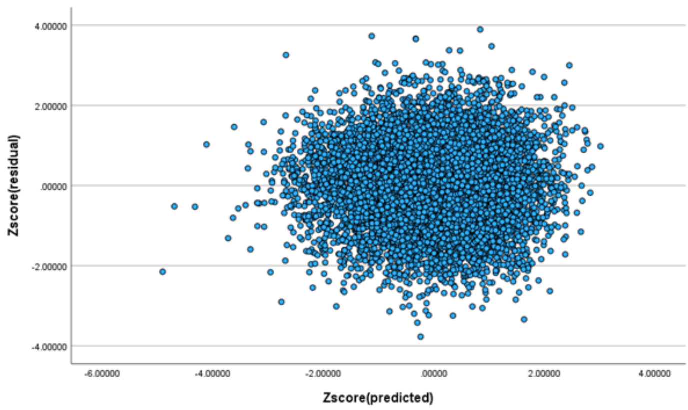

Figure 4.1. Scatterplot of ZRESID and ZPRED for immigrants.

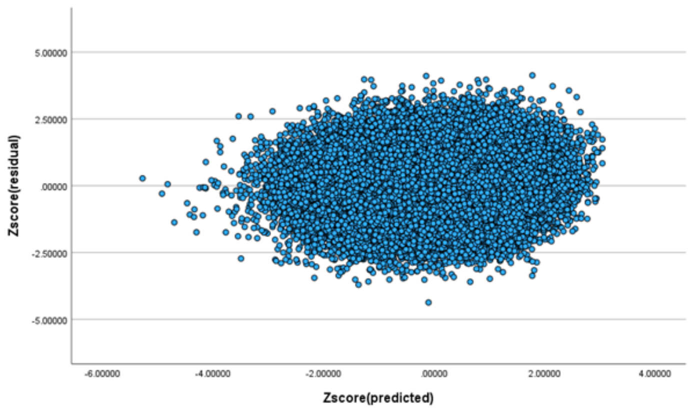

Figure 4.2. Scatterplot of ZRESID and ZPRED for natives.

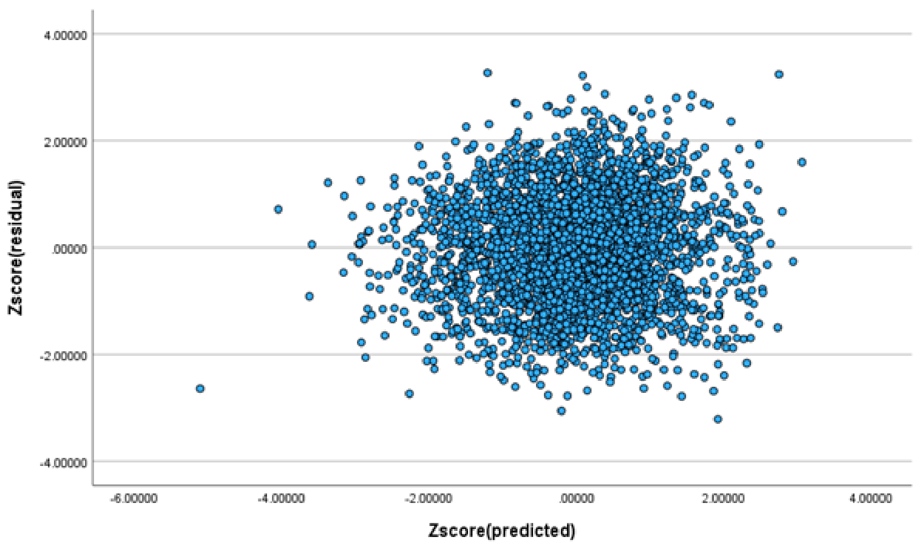

Figure 4.3. Scatterplot of ZRESID and ZPRED for 1st generation immigrants.

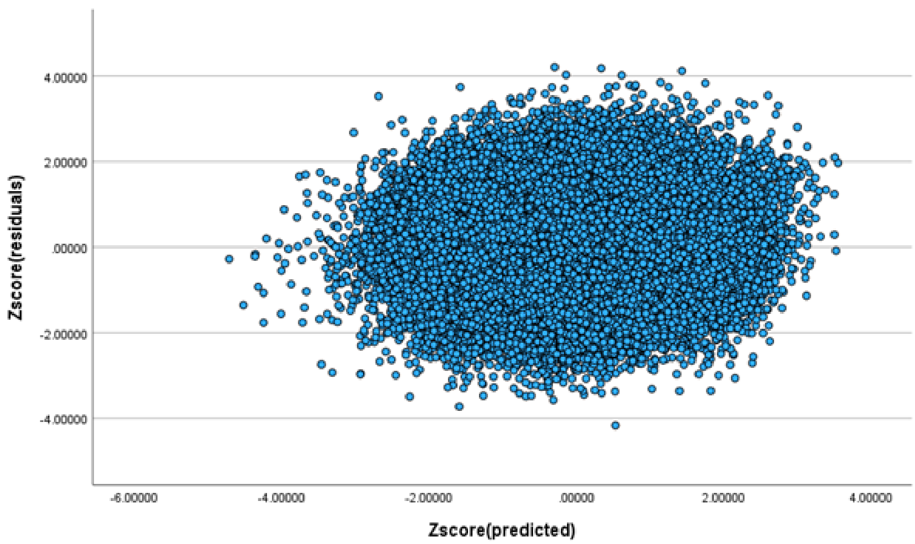

Figure 4.4. Scatterplot of ZRESID and ZPRED for 2nd generation immigrants.

4.1.2. Normality

The curves of normality for residuals were given below for each of the four analyses. It was evident from each curve that there is a normal distribution of residuals. The assumption was therefore met.

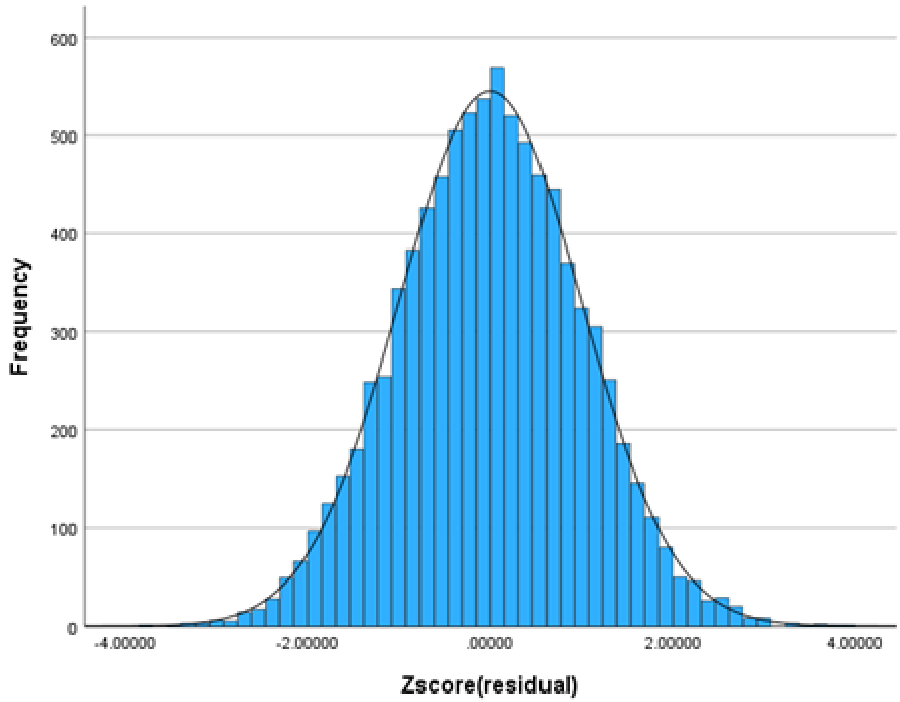

Figure 4.5. Normality curve of residuals for ımmigrants.

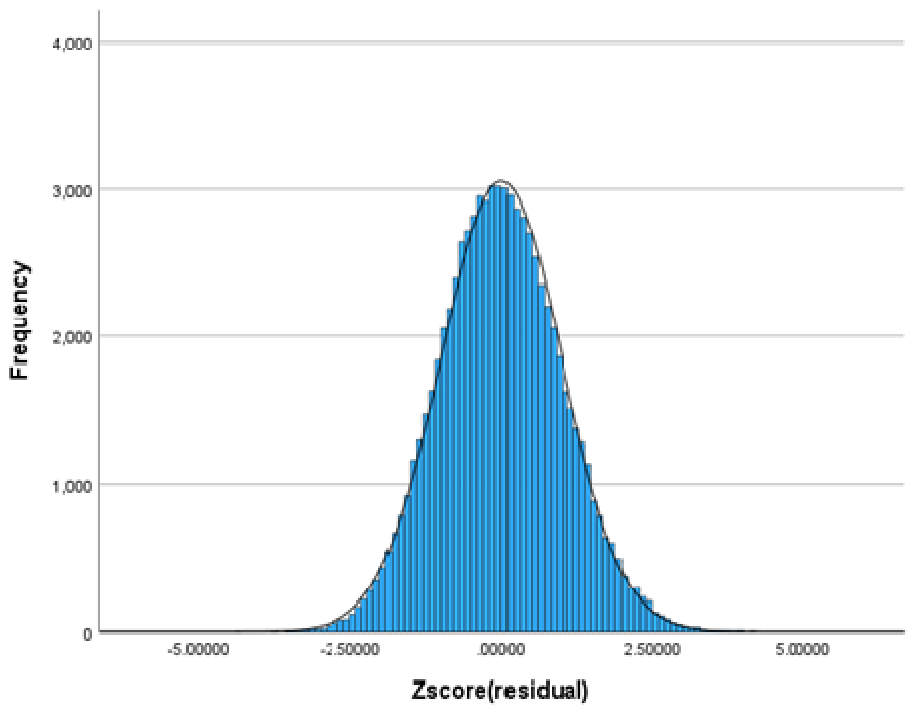

Figure 4.6. Normality curve of residuals for natives.

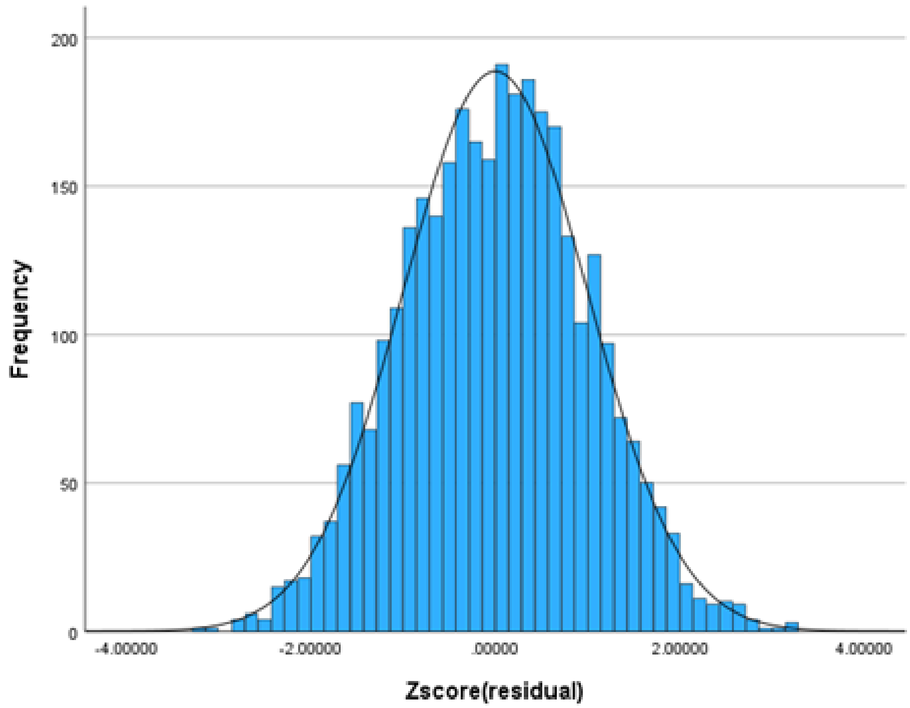

Figure 4.7. Normality curve of residuals for first-generation immigrants.

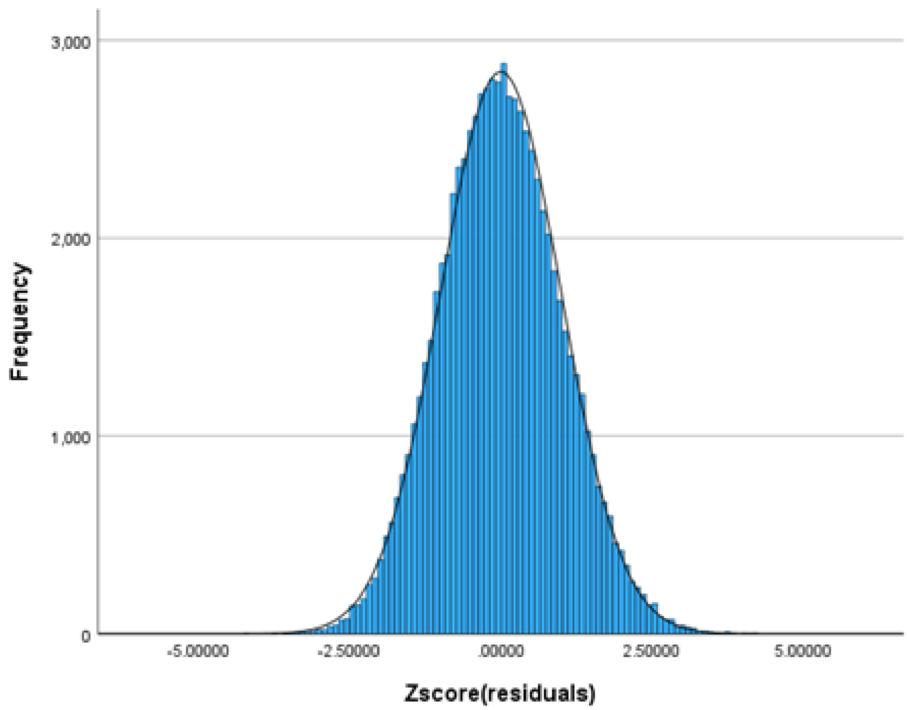

Figure 4.8. Normality curve of residuals for second-generation immigrants.

4.2. Predicting Mathematics Performance

Multilevel regression analysis was conducted to find student-level and country-level variables that predict mathematics performance. The multilevel regression analysis was used independently for each student group, such as immigrant students, native students, first-generation and second-generation immigrant students.

4.2.1. Predicting Immigrants' Mathematics Performance

Firstly, multilevel analysis was used to predict the mathematics performance of first- and second-generation immigrants from 15 countries. The countries are as follows: Croatia, Estonia, Germany, Greece, Iceland, Italy, Ireland, Latvia, Lithuania, Malta, Portugal, Serbia, Slovenia, Spain, and Switzerland.

The intraclass correlation was determined to be 0.06, indicating that 6% of the variance in mathematics performance was explained at the country level. Table 4.1 showed that resilience (β= 0.05, $p < 0.01$) students' attitudes toward immigrants (β= 0.10, $p < 0.001$), cognitive flexibility/ability (β= 0.03, $p < 0.05$), discriminating school climate (β= -0.18, $p < 0.001$), and exposure to bullying (β= -0.05, $p < 0.001$) were significant predictors of mathematics performance of first- and second-generation immigrant students at the student level. This indicated that immigrant students who are resilient, have positive attitudes toward immigrants, and are cognitively flexible tend to perform better in mathematics. Furthermore, immigrant students who were in a discriminating school climate (β= -0.18, $p < 0.001$) and exposed to bullying (β= -0.05, $p < 0.001$) tend to perform poorly in mathematics. The most important predictor at the student level was discriminating school climate (β= -0.18, $p < 0.001$). Meanwhile, the current study's student-level variables can explain 7% of the variation at the student level.

On the other hand, at the country level, government expenditure on education (β= 0.82, $p < 0.05$) was a significant predictor. Immigrant students' mathematics performance in countries with higher government education expenditures tends to be higher. The variables at the country level accounted for 39% of the variation in mathematics proficiency at the country level.

Table 4.1. Multilevel regression results for immigrants.

Variables	
Student level Variables	Standardized β weight (SE)
Motivation to Master Task	0.03 (0.03)
Resilience	0.05** (0.02)
Students' Attitudes Toward Immigrants	0.10*** (0.02)
Cognitive Flexibility/Adaptivity	0.03*(0.01)
Discriminating School Climate	-0.18*** (0.01)
Sense of Belonging	-0.01 (0.01)
Exposure to Bullying	-0.05*** (0.02)
Country level variables	
MIPEX	0.08 (0.27)
HDI	-0.38 (0.27)
Government Expenditure on Education	0.82* (0.33)
Explained variance within countries	7%
Explained variance between countries	39%
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$	

4.2.2. Predicting Natives' Mathematics Performance

Then, the analysis was repeated for native students in 15 countries. The intraclass correlation was found to be 0.06, meaning that 6% of the variation in mathematics performance was situated at the country level. As shown in Table 4.2, native students' mathematics performance was significantly predicted by motivation to master tasks (β= 0.04, $p < 0.001$), resilience (β= 0.06, $p < 0.001$), and students' attitudes toward immigrants (β= 0.09, $p < 0.001$), cognitive flexibility/ability (β= 0.05, $p < 0.001$), discriminating school climate (β= -0.19, $p < 0.001$), and exposure to bullying (β= -0.05, $p < 0.001$). It implies that native students tend to perform better in mathematics when they were motivated to master tasks, resilient, have positive attitudes toward immigrants, and were more cognitively flexible. Additionally, native students who experienced bullying and attended schools with a discriminating climate are likely to perform poorly. The discriminating school climate was the most significant predictor at the student level.

There was not any significant predictor at the country-level. The variables at the student level can account for 8% of the variance at the student level, whilst variables at the country level explained 33% of the difference in mathematics performance at the country level.

Table 4.2. Multilevel regression results for natives' mathematics performance.

Student level Variables	Standardized β weight (SE)
Motivation to Master Task	0.04*** (0.01)
Resilience	0.06*** (0.02)
Students' Attitudes Toward Immigrants	0.09*** (0.01)
Cognitive Flexibility/Adaptivity	0.05*** (0.01)
Discriminating School Climate	-0.19*** (0.01)
Sense of Belonging	-0.01 (0.01)
Exposure to Bullying	-0.05*** (0.01)
Country level variables	
MIPEX	0.28 (0.26)
HDI	0.30 (0.29)
Government Expenditure on Education	0.18 (0.30)
Explained variance within countries	8%
Explained variance between countries	33%
*p < 0.05, **p < 0.01, ***p < 0.001	

4.2.3. Predicting First-Generation Immigrants' Performance

The multilevel regression analysis was conducted for first-generation immigrant students. Only eight countries, such as Germany, Greece, Ireland, Italy, Malta, Slovenia, Spain, and Switzerland, were included, as only these countries collected data of chosen student-level variables. Since it was found that there was an intraclass correlation of 0.10, 10% of the variation in the performance of mathematics is located at the country level.

Table 4.3 indicates that resilience (β= 0.05, $p < 0.05$), students' attitudes toward immigrants (β= 0.10, $p < 0.001$), discriminating school climate (β= -0.19, $p < 0.001$), and exposure to bullying (β= -0.06, $p < 0.001$) significantly predicted mathematics performance of first-generation immigrant students at the student-level. Findings suggest that first-generation immigrant students typically performed better in mathematics when they are resilient and have a positive attitude toward immigrants. Additionally, first-generation immigrants who faced bullying and attended schools with a discriminating environment are more likely to struggle in math. Similar to the findings of variables that predict immigrant students' mathematics performance, discriminating school climate was the most significant predictor of first-generation students.

On the other hand, education scores on MIPEX (β= 0.52, $p < 0.05$) and government expenditure on education (β= 1.29, $p < 0.001$) were found to be significant predictors. The mathematics performance of first-generation immigrant students was more remarkable in countries with higher government education expenditures. Furthermore, first-generation immigrant students did better in mathematics in countries with higher MIPEX education scores. Government expenditure on education was the most significant predictor at the country level.

Variables at the student level can explain 7% of the variance at the student level. In contrast, variables at the country level can explain 83% of the variability in the performance in mathematics at the country level.

Table 4.3. Multilevel regression results of first-generation immigrants.

Student level Variables	Standardized β weight (SE)
Motivation to Master Task	0.01 (0.03)
Resilience	0.05* (0.03)
Students' Attitudes Toward Immigrants	0.10*** (0.02)
Cognitive Flexibility/Adaptivity	0.00 (0.02)
Discriminating School Climate	-0.19*** (0.02)
Sense of Belonging	-0.05 (0.03)
Exposure to Bullying	-0.06*** (0.00)
Country level variables	
MIPEX	0.52* (0.25)
HDI	-0.72 (0.42)
Government Expenditure on Education	1.29*** (0.29)
Explained variance within countries	7%
Explained variance between countries	83%
*p < 0.05, **p < 0.01, ***p < 0.001	

4.2.4. Predicting Second -Generation Immigrants' Performance

Lastly, multiple regression was conducted for second-generation students in 12 countries, such as Croatia, Estonia, Germany, Greece, Ireland, Italy, Latvia, Lithuania, Serbia, Slovenia, Spain, and Switzerland, as these countries have sufficient sample sizes of second-generation immigrants.

Given that there was an intraclass correlation of 0.05, 5% of the variation in mathematics performance was found at the country level. According to Table 4.4, at the student-level, significant factors of mathematics performance of second-generation immigrant students were students' attitudes toward immigrants (β= 0.11, $p < 0.001$), cognitive flexibility/ability (β= 0.06, $p < 0.01$), discriminating school climate (β= -

0.18, $p < 0.001$), sense of belonging (β= -0.05, $p < 0.05$), and exposure to bullying (β= -0.05, $p < 0.05$).

The findings shows that second-generation immigrant students performed better in mathematics when they have positive attitudes toward immigrants, are more flexible cognitively, and feel more sense of belonging. Additionally, second-generation immigrant students who attended schools with a discriminated climate and were bullied did worse in mathematics.

Government expenditure on education (β= 0.86, $p < 0.05$) was found to be a significant predictor at the country-level. This finding indicates that second-generation immigrant students performed better in mathematics in countries with higher government education expenditure scores.

Student-level factors can account for 7% of the variation at that level, but country-level variables can account for 76% of the variation in mathematics performance at the corresponding level.

Table 4.4. Multilevel regression results of second-generation immigrants.

Student level Variables	Standardized β weight (SE)
Motivation to Master Task	0.04 (0.03)
Resilience	0.04 (0.02)
Students' Attitudes Toward Immigrants	0.11*** (0.03)
Cognitive Flexibility/Adaptivity	0.06** (0.02)
Discriminating School Climate	-0.18*** (0.02)
Sense of Belonging	-0.05* (0.02)
Exposure to Bullying	-0.05* (0.02)
Country level variables	
MIPEX	0.33 (0.21)
HDI	-0.32 (0.23)
Government Expenditure on Education	0.86* (0.24)
Explained variance within countries	7%
Explained variance between countries	76%
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$	

To summarize, three common student-level variables were shown to be significant in all four analyses: students' attitudes toward immigrants, discriminating school climate, and exposure to bullying. Furthermore, motivation to master tasks found as significant only for native students, whereas sense of belonging tasks found as significant only for second-generation immigrant students. Resilience, on the other hand, was a significant predictor of the mathematics performance of immigrants. Cognitive flexibility was significant in all analyses except for second-generation immigrant students.

At the country level, except for natives, government expenditure on education was a significant predictor. Furthermore, education score on MIPEX has been shown to be significant in the multilevel regression analysis for only first-generation immigrant students, showing that first-generation immigrant students from countries with higher education scores on MIPEX tend to perform better.

4.3. The Mathematics Performance Gap Between Natives and Immigrants

In order to answer the second research question, independent sample t-tests were conducted for each country to compare native students and immigrants, as well as first- and second-generation immigrants' mathematics performance. Thus, Bonferroni correction was applied by dividing the p-value into 2, the number of t-test. Therefore, in this case, p-value was taken as 0.025 rather than 0.050. IDB Analyzer was used to take sampling weights and plausible values into account.

As seen in Table 4.5, there is a statistically significant difference in the mean values of mathematics performance of native and immigrant students. Native students outperformed immigrant students in countries including Austria, Belgium, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Iceland, Italy, Luxembourg, Netherlands, Norway, Portugal, Slovenia, Spain, Switzerland, and the United Kingdom. The effect size of the mathematics performance difference was large in countries such as Slovenia ($d = 0.83$), Sweden ($d = 0.80$), Finland ($d = 0.79$), Netherlands (d= 0.75), Austria (d= 0.71), and Denmark (d= 0.70). According to Table 4.6, effect size was found to be medium in the following countries: Belgium (d= 0.63), Iceland (d= 0.60), Germany (d= 0.60), France (d= 0.57), Czech Republic (0.58), Greece (d= 0.54), Switzerland (d= 0.54), Norway (d= 0.50), Portugal (d= 0.46), Spain (d= 0.46.), and Italy (d= 0.38). On the other hand, Estonia (d= 0.33), Luxembourg (d= 0.29), and the United Kingdom (d= 0.15) all had small effect sizes. For the following countries, however, there was no statistically significant result: Bosnia, Ireland, Latvia, Lithuania, Malta, Montenegro, and Serbia.

Table 4.5. *t*-test statistics for natives and immigrants.

Country's	N		Mean				
Name	Natives	Immigrants	Natives	Immigrants	SE_{diff}	t	d
Austria	5303	1407	514.13	450.81	4.81	13.17*	0.71
Belgium	6782	1473	520.71	462.93	4.19	13.78*	0.63
Bosnia	6068	177	408.34	396.41	9.49	1.26	0.15
Croatia	5882	598	466.21	453.96	6.13	2.00	0.14
Czech Republic	6674	251	500.21	471.80	10.91	4.90*	0.58
Denmark	5858	1551	516.36	460.00	4.76	11.84*	0.70
Estonia	4635	539	527.21	500.05	4.71	5.77*	0.33
Finland	5211	313	512.24	448.47	7.01	9.10*	0.79
France	5220	947	504.01	452.14	5.81	8.93*	0.57
Germany	3677	1050	518.05	462.40	6.12	9.10*	0.60
Greece	5570	702	458.28	411.15	5.36	8.79*	0.54
Iceland	3025	179	500.12	447.15	7.16	7.40*	0.60
Ireland	4454	958	501.86	496.35	3.58	1.54	0.07
Italy	10283	1071	491.71	456.09	5.33	6.68*	0.38
Latvia	4936	244	497.62	486.21	9.33	1.22	0.14
Lithuania	6558	148	482.49	470.85	11.16	1.04	0.13
Luxembourg	2318	2798	500.21	471.80	2.99	9.51*	0.29
Malta	2955	281	475.89	476.30	6.20	-0.07	0.00

Table 4.5. t-test statistics for natives and immigrants (cont.).

Country's Name	N		Mean				
	Natives	Immigrants	Natives	Immigrants	SE_{diff}	t	d
Montenegro	6121	383	430.11	432.68	5.92	-0.43	-0.03
Netherlands	3951	661	530.70	462.94	7.05	9.61*	0.75
Norway	4882	695	508.34	463.60	4.46	10.03*	0.50
Portugal	5254	342	496.73	452.02	8.81	5.07*	0.46
Serbia	5810	598	449.68	453.49	6.72	-0.57	-0.04
Slovenia	5730	573	516.10	444.19	7.70	9.34*	0.83
Spain	30727	4117	488.49	448.90	3.20	12.37*	0.46
Sweden	4283	1055	518.50	449.39	4.82	14.33*	0.80
Switzerland	3731	1926	533.54	484.48	4.32	11.36*	0.54
United Kingdom	11249	1730	506.50	492.68	4.29	3.22*	0.15
*$p < 0.025$							

4.4. The Performance Gap Between 1st and 2nd Generation Immigrants

The performance gap between first- and second-generation immigrant students was also investigated using an independent sample t-test. Second-generation immigrant students had higher mean scores than first-generation immigrant students in the following countries: Czech Republic, France, Germany, Netherlands, Portugal, Slovenia, Spain, and Sweden. Only Portugal has been found to have large effect sizes, whereas Slovenia, the Czech Republic, Sweden, Germany, and the Netherlands all have medium effect sizes (d= 0.54, d= 0.49, d= 0.43, d= 0.41, and d= 0.36, respectively).

The negative t values in Table 4.6 for Latvia and Luxembourg showed that first-generation immigrants performed better than second-generation immigrants in mathematics performance. Table 4.6 demonstrates that Luxembourg had a small effect size while Latvia had a large effect size.

Table 4.6. *t*-test statistics for first- and second-generation immigrants.

	N		Mean				
Country's Name	Second-Gen	First-Gen	Second-Gen	First-Gen	SE_{diff}	t	d
Austria	933	474	453.46	445.76	7.15	1.08	0.09
Belgium	812	661	468.11	456.15	6.00	1.99	0.13
Bosnia	87	90	406.60	386.12	15.24	1.34	0.24
Crotia	511	87	453.28	457.82	14.11	-0.32	-0.05
Czech Republic	128	123	473.37	424.90	18.71	2.59*	0.49
Denmark	1282	269	456.95	471.40	9.33	-1.55	-0.18
Estonia	498	41	501.11	486.31	19.04	0.78	0.18
Finland	134	179	455.49	443.28	11.65	1.05	0.14
France	611	336	459.83	436.39	9.19	2.55*	0.26
Germany	751	299	473.75	434.79	12.23	3.18*	0.41
Greece	529	173	414.29	402.62	9.73	1.20	0.14
Iceland	80	99	444.97	448.89	12.70	-0.31	-0.05
Ireland	429	529	492.14	499.80	6.76	-1.13	-0.10
Italy	572	499	462.45	448.47	9.92	1.41	0.16
Latvia	197	47	474.93	533.59	22.10	-2.65*	-0.67
Lithuania	114	34	467.49	483.09	29.39	-0.53	-0.16
Luxembourg	1549	1249	466.66	478.16	4.96	-2.32*	-0.11
Malta	69	212	452.13	484.08	17.13	-1.87	-0.30
Montenegro	236	147	435.66	427.81	10.90	0.72	0.09
Netherlands	518	143	469.37	437.09	14.00	2.31*	0.36
Norway	348	347	460.55	466.66	8.34	-0.73	-0.07
Portugal	202	140	481.19	412.49	15.50	4.43*	0.69
Serbia	529	69	450.97	473.60	13.72	-1.65	-0.24
Slovenia	250	323	471.86	425.03	12.61	3.71*	0.54
Spain	1610	2507	460.39	441.16	5.98	3.21*	0.22
Sweden	556	499	466.93	429.47	8.05	4.66*	0.43
Switzerland	1234	692	484.19	485.00	5.90	-0.14	-0.01
United Kingdom	853	877	489.92	496.39	9.24	-0.70	-0.07

*$p < 0.025$

4.5. Comparing Mean Scores after Controlling ESCS and Gender

In order to control the effects of ESCS and gender in mathematics performance differences, propensity score matching was employed. In order to match each immigrant student with a nearby native student based on their distance, the nearest neighbor technique was applied. This technique was applied for every country separately. Randolph and Falbe (2014) noted that R cannot match propensity scores if a value is missing in the data set. Therefore, the missing data was eliminated from the analysis so that R may execute propensity matching. The treatment group was coded as one, and the control group was coded as zero once the data have been cleaned. Following the propensity score, the independent sample t-test was used to evaluate the mean scores in mathematics performance of immigrants and native students in each country, controlling ESCS and gender. Firstly, natives and immigrants were matched. Then, first-generation immigrants and second-generation immigrants were matched.

4.5.1. Balance of Nearest Propensity Score Matching

In order to evaluate the balance of matching, variance ratios, and standardized mean differences were reported. In addition, visual data such as histograms were provided in each country. Rubin (2001) defines good matches as having a variance ratio of near 1.00 and a standardized mean difference of propensity scores of 0.50 after matching. Each country's data was reviewed using these criteria.

<u>4.5.1.1. Matching Native and Immigrant Students.</u> Firstly, the matching of natives and immigrants was examined. Most countries' standardized mean differences of propensity scores after matching are less than 0.50, except Luxembourg, as shown in Table 4.7.

On the other hand, as indicated in Table 4.8, after matching, the variance ratio was higher than one in countries such as the Czech Republic (2.20), Finland (2.01), Germany (1.61), Iceland (1.54), Netherlands (1.57), Norway (1.51), Sweden (1.70) and

Switzerland (1.77). Therefore, histograms of these countries were examined. The histograms of matched treated and matched control groups were compared. The histograms were very similar in some countries, such as the Czech Republic, Finland, Iceland, and Norway. The histogram of the Czech Republic is given in Figure 4.9. Therefore, matching was operated well. The histograms were similar but not identical in other countries such as Germany, Netherlands, Sweden, and Switzerland, indicating that matching was as good as feasible. These countries' histograms were included in Appendix A.

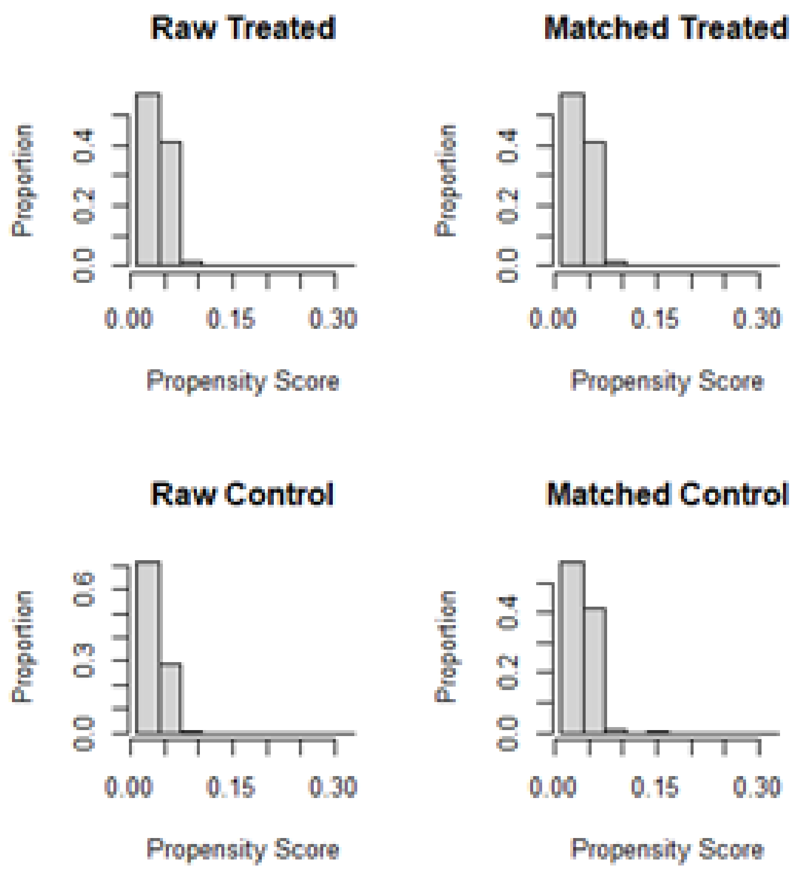

Figure 4.9. Histogram of matching natives and immigrants-czech republic.

In Luxembourg, contrary to other countries, there were more immigrant students than native students. Therefore, some immigrant students could not be matched with native students. While nearest neighbor propensity score matching was used, there were 2736 immigrants and 2272 native students initially. All 2272 immigrant students were matched with 2272 native students. As a result, 464 immigrants were unmatched. The standardized mean difference of propensity score after the matching was more than 0.50 since the value for Luxembourg is 0.89. Therefore, the propensity score was again applied by defining natives as the treatment group and immigrants as the control units. As indicated in Table 4.7, the standardized mean difference became less than 0.50, and the variance ratio was near 1. The second matching was more balanced than the first matching.

Table 4.7. Descriptives of matching natives and immigrants.

	Before Matching		After Matching	
	Std. Mean Diff	Variance Ratio	Std. Mean Diff	Variance Ratio
Austria	0.68	2.43	0.14	1.67
Belgium	0.57	2.17	0.05	1.28
Czech Republic	0.22	4.92	0.03	2.20
Denmark	0.77	2.80	0.17	1.67
Estonia	0.15	0.97	0.00	1.00
Finland	0.45	6.65	0.09	2.01
France	0.57	2.92	0.04	1.23
Germany	0.70	2.30	0.14	1.61
Greece	0.72	1.67	0.00	1.03
Iceland	0.43	6.88	0.03	1.54
Italy	0.58	1.66	0.00	1.04
Luxembourg (First)	0.59	1.64	0.89	1.09
Luxembourg (Second)	0.76	0.60	0.41	0.81
Netherlands	0.59	3.57	0.10	1.57
Norway	0.49	3.88	0.07	1.51
Portugal	0.09	1.01	0.00	0.99
Slovenia	0.58	2.28	0.01	1.22
Spain	0.60	1.80	0.00	1.04
Sweden	0.58	3.03	0.13	1.70
Switzerland	0.63	1.97	0.24	1.77
United Kingdom	0.25	1.45	0.01	1.09

4.5.1.2. Matching First and Second Generation Immigrants. Secondly, the matching of first-generation and second-generation immigrants was examined. First-generation immigrant is in the treatment group and coded as 1, while second-generation immigrants are in the control group and coded as 0. On the other hand, in Slovenia, there were more first-generation immigrants (N= 323) than second-generation immigrants (N= 248). As a result, matching was carried out by identifying second-generation immigrants as the treatment group and first-generation immigrants as the control group.

By analyzing the criteria based on standardized mean difference scores, it is possible to conclude that all countries have standardized mean difference scores less than 0.50. Furthermore, variance ratios were near 1 in all countries except the Czech Republic and Sweden. Histograms from the Czech Republic and Sweden were compared using the similarity of histograms from matched and control groups. The histograms were alike but not identical, suggesting that matching was as close to balance as possible.

Table 4.8. Descriptives of matching 1st and 2nd generation immigrants.

	Before Matching		After Matching	
	Std. Mean Diff	Variance Ratio	Std. Mean Diff	Variance Ratio
Czech Republic	0.20	1.32	0.16	1.37
France	0.04	1.06	0.00	1.01
Germany	0.04	1.20	0.00	1.01
Latvia	0.29	1.12	-0.01	1.21
Luxembourg	0.13	1.16	0.02	1.07
Netherlands	0.12	1.00	0.00	1.00
Portugal	0.30	1.02	0.04	1.14
Slovenia	0.18	0.93	0.05	1.13
Spain	0.01	0.98	0.00	1.00
Sweden	0.24	1.51	0.15	1.49

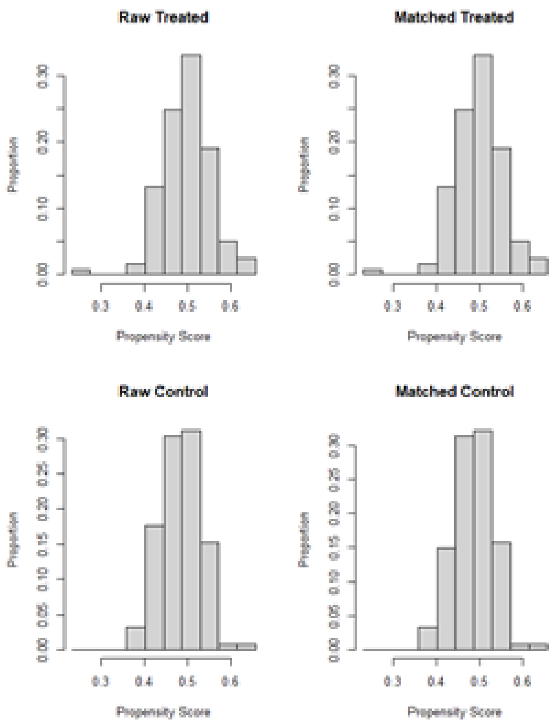

Figure 4.10. Histograms of matching first- and second-generation immigrants-czech republic.

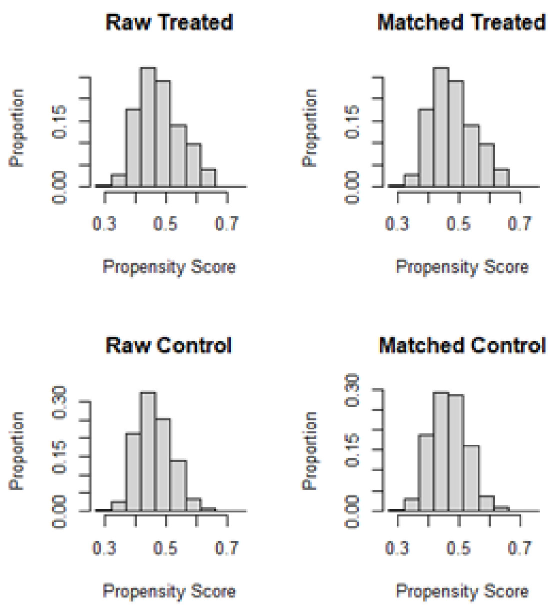

Figure 4.11. Histograms of matching first- and second-generation immigrants-sweden.

4.5.2. Gap Between Natives and Immigrants

For countries where results were statistically significant in a former t-test, propensity score matching, and an independent sample t-test were used. Therefore, 20 countries, such as Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Iceland, Italy, Luxembourg, Netherlands, Norway, Portugal, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom, were included. Immigrants represented the treatment group, whereas natives were the control group.

Table 4.9 displays the number of immigrants and natives after cleaning data and unmatched data. Since each data in the treatment group was matched with corresponding data in the control group, unmatched data

Table 4.9. Sample sizes of matching natives and immigrants.

	Before Matching		After Matching		
	Natives	Immigrants	Natives	Immigrants	Unmatched
Austria	5068	1379	1379	1379	3689
Belgium	6739	1433	1433	1433	5306
Czech Republic	6623	246	246	246	6377
Denmark	5844	1520	1520	1520	4324
Estonia	4627	536	536	536	4091
Finland	5194	309	309	309	4885
France	5184	932	932	932	4252
Germany	3615	1007	1007	1007	2608
Greece	5569	702	702	702	4867
Iceland	3018	177	177	177	2841
Italy	10254	1064	1064	1064	9190
Luxembourg	2272	2736	2736	2736	464

Table 4.9. Sample sizes of matching natives and immigrants (cont.).

	Before Matching		After Matching		
	Natives	Immigrants	Natives	Immigrants	Unmatched
Netherlands	3935	646	646	646	3289
Norway	4859	685	685	685	4174
Portugal	5236	340	340	340	4896
Slovenia	5722	571	571	571	5151
Spain	30584	4095	4095	4095	26489
Sweden	4246	1019	1019	1019	3227
Switzerland	3706	1902	1902	1902	1804
United Kingdom	10921	1601	1601	1601	9320

The results revealed that there is still a statistically significant difference in favor of native students over immigrant students in all countries except the Czech Republic, Iceland, Italy, and the United Kingdom, as indicated in Table 4.10. However, the t-value for the United Kingdom was negative since immigrant students' mean values were higher than those of their native peers. However, there was no statistically significant difference between them in the United Kingdom.

Table 4.11 shows that the effect sizes decreased from large to medium in the following nations-Austria, Belgium, Denmark, Finland, the Netherlands, Slovenia, and Sweden. Furthermore, medium effect sizes declined to small levels in countries like France, Germany, Greece, Iceland, Spain, and Switzerland. The effect sizes before and after matching were shown in Figure 4 for countries where the results are statistically significant. There is a global trend toward smaller effect sizes, except for Portugal, Luxembourg, and Estonia.

Table 4.10. t-test statistics after controlling ESCS and gender.

Country's Name	N		Mean				
	Natives	Immigrants	Natives	Immigrants	SEdiff	t	Effect Size
Austria	1379	1379	492.27	452.16	5.82	6.89*	0.45
Belgium	1433	1433	493.98	463.78	4.51	6.70*	0.33
Czech Republic	246	246	478.18	449.09	13.33	2.18	0.29
Denmark	1520	1520	495.82	460.38	5.13	6.91*	0.44
Estonia	536	536	527.22	500.30	5.92	4.55*	0.33
Finland	309	309	499.79	449.36	9.09	5.55*	0.60
France	932	932	476.39	452.75	6.93	3.41*	0.25
Germany	1007	1007	490.33	465.53	7.06	3.51*	0.26
Greece	702	702	431.47	411.15	6.80	2.99*	0.24
Iceland	177	177	470.56	447.92	10.25	2.21	0.27
Italy	1064	1064	465.43	456.15	7.19	1.29	0.10
Luxembourg	2272	2272	500.91	481.90	3.18	5.97*	0.20
Netherlands	646	646	497.34	464.17	7.17	4.63*	0.37
Norway	685	685	488.41	464.65	5.54	4.29*	0.26
Portugal	340	340	495.18	452.32	10.67	4.02*	0.43
Slovenia	571	571	488.13	444.33	8.73	5.01*	0.50
Spain	4095	4095	465.86	449.11	3.82	4.39*	0.19
Sweden	1019	1019	498.70	450.66	5.33	9.01*	0.55
Switzerland	1902	1902	519.04	485.10	4.85	6.99*	0.37
United Kingdom	1601	1601	483.95	494.94	5.70	-1.93	-0.12
*$p < 0.025$							

Table 4.11. Effect sizes before and after matching natives and immigrants.

Country's Name	Before Matching	After Matching
Austria	0.71***	0.45**
Belgium	0.63**	0.33*
Czech Republic	0.58**	0.29*
Denmark	0.70***	0.44**
Estonia	0.33*	0.33*
Finland	0.79***	0.60**
France	0.57**	0.25*
Germany	0.60**	0.26*
Greece	0.54**	0.24*
Iceland	0.60**	0.27*
Italy	0.38*	0.10*
Luxembourg	0.29*	0.20*
Netherlands	0.75***	0.37*
Norway	0.50**	0.26*
Portugal	0.46**	0.43**
Slovenia	0.83***	0.50**
Spain	0.46**	0.19*
Sweden	0.80***	0.55**
Switzerland	0.54**	0.37*
United Kingdom	0.15*	-0.12*
*Small effect size		
** Medium effect size		
***Large effect size		

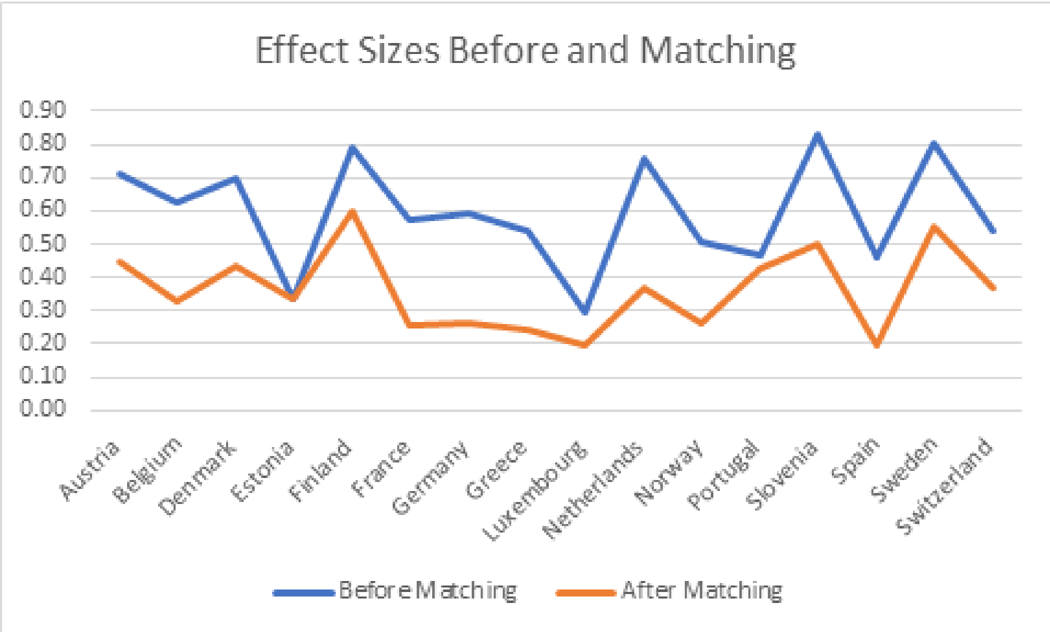

Figure 4.12. Effect sizes before and after matching of natives and immigrants.

4.5.3. Gap Between 1st and 2nd Generation Immigrants

Since there were only eight countries with a significant difference between first- and second-generation immigrants' mathematics performance, the Czech Republic, France, Germany, Netherlands, Portugal, Slovenia, Spain, and Sweden were included in the analysis.

Table 4.12 shows the number of first-generation and second-generation immigrants after eliminating missing data. Unmatched data descriptives were given as well.

The statistically significant difference was found in the following countries, the Czech Republic, Germany, Portugal, Slovenia, Spain, and Sweden. As demonstrated in Table 4.13, France and the Netherlands did not have a statistically significant difference. There was no change in effect size in countries whose result is significant except in Portugal. The effect size declined from 0.69 to 0.54, as shown in Table 4.14. Figure 5 shows the graph of effect sizes before and after in countries with significant results.

Table 4.12. Sample sizes matching first- and second-generation immigrants.

	Before Matching		After Matching		Unmatched
	Second-gen. immigrants	First-gen. immigrants	Second-gen. immigrants	First-gen. immigrants	
Czech Republic	125	121	121	121	4
France	603	329	329	329	274
Germany	728	279	279	279	449
Latvia	197	46	46	46	151
Luxembourg	1514	1222	1222	1222	292
Netherlands	509	137	137	137	372
Portugal	202	138	138	138	64
Slovenia	248	323	248	248	75
Spain	2489	1606	1606	1606	883
Sweden	541	478	478	478	63

Table 4.13. t-test statistics for immigrants after controlling ESCS and gender.

Country's Name	N		Mean				
	Second-Gen	First-Gen	Second-Gen	First-Gen	SEdiff	t	Effect size
Czech Republic	121	121	475.92	425.31	18.93	2.67*	0.52
France	329	329	455.84	437.90	10.10	1.78	0.20
Germany	279	279	485.67	440.17	13.01	3.50*	0.47
Netherlands	137	137	469.36	440.02	16.07	1.83	0.31
Portugal	138	138	466.81	412.87	16.30	3.31*	0.54
Slovenia	248	248	472.33	431.25	12.35	3.33*	0.48
Spain	1606	1606	460.41	440.02	6.39	3.19*	0.24
Sweden	478	478	466.54	430.91	7.93	4.49*	0.41
*p < 0.025							

Table 4.14. Effect sizes of matching first- and second-generation immigrants.

Country's Name	Before Matching	After Matching
Czech Republic	0.49**	0.52**
France	0.26*	0.20*
Germany	0.41**	0.47**
Netherlands	0.36*	0.31*
Portugal	0.69***	0.54**
Slovenia	0.54**	0.48**
Spain	0.22*	0.24*
Sweden	0.43**	0.41**
*Small effect size		
** Medium effect size		
***Large effect size		

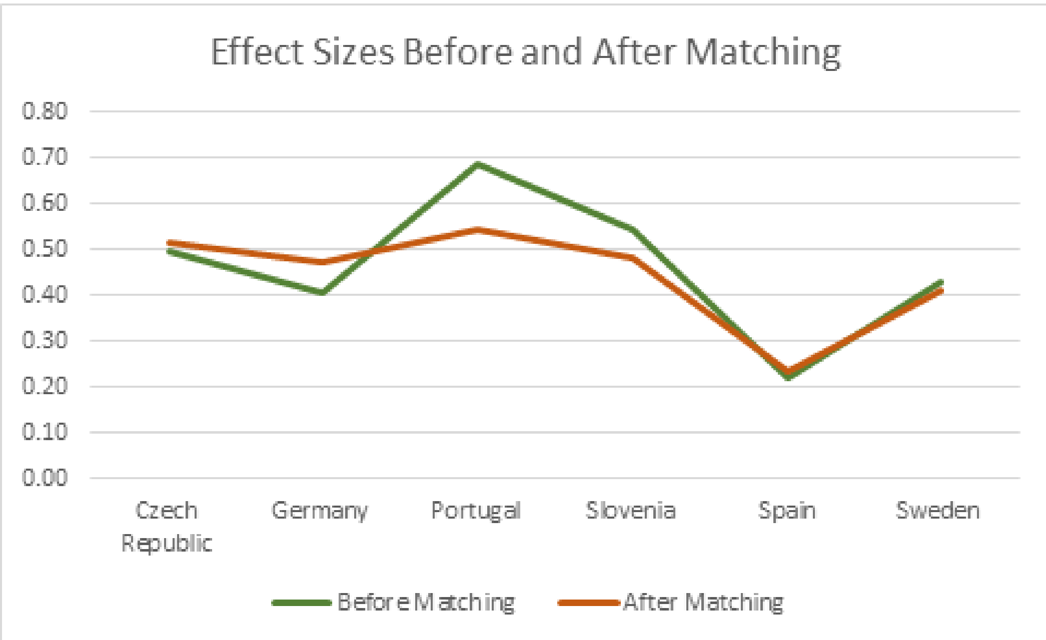

Figure 4.13. Effect sizes before and after matching of first- and second-generation immigrants.

The groups, on the other hand, were similar in terms of ESCS and gender prior to matching, as explained in the balance of matching, based on their descriptives of ESCS and balance values such as standardized mean difference and variance ratios. The reason for having no change may be that the groups were similar prior to matching.

5. DISCUSSION

5.1. Predicting the Math Performance of Immigrants and Native

One of the aims of this study was to investigate student-level and country-level variables that can predict the mathematics performance of immigrant and native students. The results indicated that at the country-level, government expenditure on education was a significant variable in the mathematics performance of immigrant and native students. The MIPEX was found to be an important factor only for first-generation immigrant students. Depending on whether the group consists of immigrants or natives, variables that predict mathematics performance can differ. The result supported Walberg's theory of academic achievement by providing that resilience, students' attitudes toward immigrants, and cognitive flexibility were significant at the student aptitude dimension, and discriminating school climate and exposure to bullying were significant at the psychological dimension.

5.1.1. Country-Level Variables

The Migrant Integration Policy Index (MIPEX) was found to be a significant factor in the mathematics performance of first-generation immigrant students. In the literature, MIPEX was found to be a significant predictor of achievement (Arıkan *et al.*, 2017; Arikan *et al.*, 2020; He *et al.*, 2017), whilst Dronkers and de Heus (2016) claimed that the performance of immigrants was not related to MIPEX. Based on these mixed results, only for first-generation immigrants, education scores on MIPEX could predict the mathematics performance in PISA 2018.

The Human Development Index (HDI) could not predict the mathematics performance of immigrant and native students. This result supports earlier research (Arıkan *et al.*, 2017). On the other hand, Campbell *et al.* (2021) have found a strong correlation between the human development index and PISA scores. In contrast, change in the HDI over 12 years of PISA assessment is not a significant predictor of PISA scores (Campbell *et al.*, 2021). This implies that, despite the HDI perhaps serving as a measure of overall growth, student performance couldn't be predicted by it. Further research is needed to understand its relation to students' performance.

Government expenditure on education was a statistically significant predictor for immigrant, first-generation immigrant, and second-generation immigrant students separately, but it was not statistically significant for native students. He *et al.*, (2017) claims that governments whose spending on education is more have higher levels of educational accomplishment. As a result, immigrant students who lived in a country that invests more in education were better at mathematics. On the other hand, Mazurek *et al.* (2021) have investigated the relationship between students' performance in PISA 2018 and socio-economic inequality using a variety of indices and found expenditure on education significant but negative, resulting in lower for students who reside in countries with higher spending on education performance in PISA 2018. However, as 70 of the 78 countries in their study were economically different, they could not develop an all-encompassing strategy to improve student performance (Mazurek *et al.*, 2021). As the researchers in this study did not categorize students according to their immigration status or native status, their findings cannot be generalized to immigrants.

Government spending on education, however, was determined to be considerable and beneficial for immigrants in the current study, which only included European countries. The conclusion is that government spending on education in European nations significantly predicted the mathematics performance of immigrants, including first and second-generation. Further research can be done to investigate the causality between allocating money to educational resources and the performance of immigrants.

Some scholars examined the causality between spending per student and their achievement. Jackson *et al.*, (2015) examined the causal relationship between external school spending rises and academic achievement in the United States. Increasing spending per student resulted in significant gains for children from low-income families regarding educational achievement (Jackson *et al.*, 2015). In addition, it was found that higher school investment has less of an impact on children from non-poor families in terms of their future educational success and family income (Jackson *et al.*, 2015). Furthermore, Jackson *et al.* (2015) claimed that their findings highlight how better access to educational resources can have a major impact on the life outcomes of economically disadvantaged children.

Further research can be done to investigate the effect of educational resources on the performance of immigrant students and which category of government expenditure predicts the performance of immigrants. For instance, Farayibi and Folarin (2021) have claimed that governments in developing nations, sub-Saharan African countries were included in their study, often focusing on spending on the lower levels of education, such as elementary school enrollment rates.

5.1.2. Student-Level Variables

In this section, the interpretation of student-level variables is made based on the results.

<u>5.1.2.1. Motivation to Master Task.</u> Motivation to master tasks was found significant for natives but not for immigrants. In the literature, motivation indicates its role in achievement. Motivation is defined in terms of mastery goals and performance goals. Mastery goals are mastering a task and developing performance (Elliot and McGregor, 2001). However, a performance goal is defined as whether someone is performed better than their peers (Elliot and McGregor, 2001). In this regard, the motivation to master tasks variable can be considered mastery goals since the items included in PISA indicate students' fulfillment based on their own success. It was found that mastery goal is correlated with achievement in some studies (Moore *et al.*, 2010; Huang, 2012) but not in others (Korpershoek, 2016). On the other hand, mastery goal is also correlated with school commitment (Korpershoek, 2016; McInerney, 2008). Stweart (2007) has found that school commitment is a significant predictor of higher education students,

indicating that students who commit to schoolwork have higher GPA scores. Therefore, motivation to master tasks was a significant predictor for native students, which may relate to their school commitment.

5.1.2.2. Resilience. Cerna *et al.* (2021) stated a difference in the share of academically resilient students by immigrant groups in OECD countries in PISA 2018; specifically, the gap between first-generation immigrants and native students is the greatest, while the gap between second generations and natives is the lowest. However, their study does not indicate any prediction between resilience and performance.

In addition to the fact that immigrants are, on average, more resilient than their native counterparts (OECD, 2019c), the current study found that resilience is also a significant predictor of immigrants' performance in mathematics. It is possible that this is related to the feeling of cultural identity. Cardoso and Thompson (2010) mentioned that cultural traditions help people feel more connected to their cultural heritage, which fosters academic and psychological resilience.

On the other hand, PISA 2018 results revealed that, in general, academically resilient children showed stronger ability to develop and follow goals as well as an increased tendency to love reading more (OECD, 2019c). There was a slight difference between the two groups of students in their expression of positive emotions. However, academically resilient students reported feeling less of a sense of meaning in life than children who were not (OECD, 2019c). On the other hand, there was no statistically significant difference between the well-being of resilient and non-resilient differences (OECD, 2019c). Furthermore, Cerna *et al.* (2021) claimed that socioeconomic status plays a key moderating role in the association between immigrant origin and being academically resilient. This might be related to the immigrant paradox. The immigrant paradox is characterized by the academic success of immigrant students in the lack of socioeconomic benefits, such as highly educated parents, linked to better school performance (Crosnoe and Turley, 2011).

On the other hand, the analysis showed that resilience was not a significant factor in predicting the mathematics performance of second-generation immigrant students. Furthermore, a sense of belonging was significant for only second-generation immigrant students indicating that emotional resilience might be a significant factor for second-generation immigrants. Cerna *et al.* (2021) defined socially resilient students by using the variables of sense of belonging, life satisfaction, and fear of failure. It appears that socially resilient second-generation immigrant students become better at mathematics performance. Sirin and Rogers-Sirin (2004) have also found that students with higher school engagement are better in terms of academic success. School engagement is measured to reflect students' sense of belonging (Sirin and Rogers-Sirin, 2004).

<u>5.1.2.3. Cognitive Flexibility/Adaptivity.</u> Cognitive flexibility is typically described in the literature as a component of the brain's executive function associated with emotional and social growth (Diamond and Lee, 2011; Zelazo and Carlson, 2012). The definition of the variable, however, explains cognitive flexibility/adaptivity as being able to adapt to new circumstances and across cultural boundaries (OECD, 2018b). On the other hand, adaptability is defined as an individual's ability to manage productively in the face of new, unpredictable, and changing settings and conditions (Martin *et al.*, 2012). Burns *et al.*, (2018) argued that adaptability and personal goal-setting affect improvements in the academic performance of students positively. Furthermore, students' mathematics achievement was positively correlated with adaptability (Collie and Martin, 2017). The current study also emphasized that cognitive flexibility/adaptivity predicted natives' and immigrants' mathematics performance. Martin *et al.* (2013) claimed that adaptability is included in self-regulated learning so that students develop strategies to adapt to new circumstances. Therefore, the inclusion of self-regulation activities in the classroom is encouraged.

<u>5.1.2.4. Exposure to Bullying.</u> In the PISA 2018 study, Karakus *et al.* (2022) found that bullying significantly affected immigrant reading results, mathematics performance, and science performance. Furthermore, students who experience bullying per-

form poorly on PISA and TIMSS (Karakus *et al.*, 2022; Ponzo, 2013). Similar to the literature, it was found that get exposed to bullying negatively related to mathematics performance for both native-born and first- and second-generation immigrant students in PISA 2018. According to Juvonen and Graham (2014), bullying students are often misconstrued, while their victims suffer unfavorable consequences that might change their feelings about their fellow students. Interestingly, Graham (2006) found that less vulnerability to bullying is associated with a more diverse classroom and school. Because exposure to bullying leads to worse performance, it is essential to take action to prevent bullying from happening in the future. Evidence showed that more diverse classrooms lead to less bullying (Graham, 2006). Therefore, diversity in classrooms and schools is encouraged because it may reduce bullying, which may lessen its negative influence on students' performance. Investing in programs that promote students' knowledge of different cultures and backgrounds is suggested.

5.1.2.5. Sense of Belonging. The sense of belonging predicted mathematics performance for second-generation immigrant students but not first-generation immigrants or native students. The previous comments mentioned that resilience did not predict mathematics performance among second-generation immigrants. There is a lower sense of belonging among immigrant students than their native peers (Cerna *et al.*, 2021). Rodríguez *et al.* (2020) found that first-generation immigrant students have the lowest sense of belonging, while natives have the highest. Second-generation immigrant students display a stronger sense of belonging than first-generation immigrant students, implying a lesser likelihood of participating in antisocial or risky activities (Rodríguez *et al.*, 2020). According to research (OECD, 2013b; Sirin and Rogers-Sirin, 2004), self-esteem, motivation, and achievement are related to the sense of belonging. This shows that fostering a feeling of community among students in the classroom might enhance their performance in school. Therefore, schools are encouraged to make an effort to provide a setting where children feel a sense of community.

5.1.2.6. Discriminating School Climate. Several scholars have discussed the school climate over the past few decades. Its effect on students and their learning was discussed first by Perry (1908). The school climate is linked to many areas, such as motivation (Goodenow and Crady, 1997) and cooperation (Finnan *et al.*, 2003). On the other hand, Cohen *et al.*, (2009) suggest that the school climate either fosters or hinders students' capacity to study and accomplish academically. A nurturing school climate can foster academic achievement (Cohen, 2009). Hoge *et al.*, (1990) studied the effect of school climate on self-esteem in 6th and 7th graders and found that school climate and positive teacher evaluation positively affect academic self-esteem. However, their study did not investigate the relationship between school climate, self-esteem, and achievement. In the current study, discriminating school climate predicted mathematics performance negatively for immigrants and natives, indicating that more discrimination reduces mathematics performance. The variable was measured by students' ratings of their teachers' attitudes toward diversity in the classroom. Cohen *et al.* (2009) stated that most teacher education initiatives lack knowledge about school climate. Since discriminating school climate leads to worse performance, teacher education programs may need to include school climate. Additionally, understanding and addressing school climate could help create a beneficial learning experience for all students, including immigrants and natives. The National School Climate Center offers planning and mentoring in addition to climate assessment (National School Climate Center, 2021). Pickeral, Evans, Hughes and Hutchison (2009) provided guidance to policymakers and education leaders and suggested that school-based teams can be created to evaluate the alignment of practice and policy so that school climate can be sustained.

5.1.2.7. Students' Attitudes Toward Immigrants. In order to explain students' attitudes toward immigrants, concepts of diversity and multiculturalism can be applied. A study by Chang and Le (2010) found that multiculturalism in schools is correlated with empathy, which leads to better academic performance among Hispanic students. Based on their findings, students' attitudes toward various cultures predict academic achievement (Chang and Le, 2010). The current study found that native students' attitudes toward immigrant students predicted their mathematics performance. Figlio *et* *al.*, (2021) have found a significant and positive correlation between academic achievement and the presence of immigrant students in classrooms in the U.S. On the other hand, their study did not investigate the attitudes of native students toward their immigrant peers.

In addition, attitudes toward immigrant students predicted mathematics performance among first- and second-generation immigrant students. Students who were aware of their rights as immigrant students performed better academically. This may be due to feeling a sense of belonging to a particular community, so a sense of equality among immigrant students may lead to better performance. Further research is necessary to investigate immigrant students' sense of cultural identity and their performance.

5.1.3. Performance Gap Between Immigrant Students and Natives

It has been systematically attempted to understand the phenomenon across national boundaries, taking into account factors such as academic subject, age, and grade level (Andon *et al.*, 2014). It was found that there was a significant achievement gap between immigrants and natives in all three of the assessment areas (math, science, reading) of PISA, TIMSS, and PIRLS from 2000 to 2009 (Andon *et al.*, 2014). In addition to this, the extent of the performance gaps between native students and immigrants varies greatly in worldwide comparisons, even though immigrant pupils typically demonstrate excellent learning requirements (Schleicher, 2006). Azzolini *et al.*, (2012) have investigated the performance gap between immigrants and natives in Italy and Spain using data from PISA 2009 and have found that natives outperformed their immigrant counterparts in mathematics and reading. The current study examined which countries have a performance gap between natives and immigrants, as well as between first-generation and second-generation immigrants. It was found that immigrant students perform poorly in mathematics in 21 counties out of 28 countries: Austria, Belgium, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Iceland, Italy, Luxembourg, Netherlands, Norway, Portugal, Slovenia, Spain, Switzerland, and the United Kingdom in PISA 2018. This suggests that immigrants may need more robust educational support in countries where they perform lower than natives. Policymakers are encouraged to provide more excellent resources and assistance to immigrant children in order for them to achieve their full potential.

On the one hand, Solano and Huddleston (2020) mentioned that education strategies are usually more effective in countries with many immigrant students. Accordingly, the Nordic countries adopt a tailored, needs-based approach (Solano and Huddleston, 2020). Australia, Canada, and New Zealand have built effective and tailored education systems through diversity, while the US focuses additional resources on vulnerable ethnic and social groups (Solano and Huddleston, 2020). In comparison, Austria, France, Germany, and Luxembourg have less sensitivity to the needs of their students, who are comparatively many immigrants (Solano and Huddleston, 2020). Similarly, the current study has been able to demonstrate that in countries such as Austria, the mean difference was larger, also in countries such as France and Germany, the mean difference is medium as well, supporting the idea that these three countries may place less value on the needs of immigrants. Furthermore, Schleicher (2006) has found performance gaps between native students and immigrants in PISA 2003 by indicating that Austria, Belgium, Denmark, France, Germany, the Netherlands, and Switzerland have the most noticeable differences. Taking into account the results of PISA 2003 and PISA 2018, it can be stated that the results of both surveys support and complement one another, so the mean difference was larger in the following countries Austria, Denmark, and the Netherlands, followed by middle effect sizes for Belgium, Germany, France, and Switzerland. Based on the results of these studies, there is an unchanged trend in these countries between 2003 and 2018 as well. This suggests that the educational practices in these countries may need to sufficiently prioritize the needs of immigrant students, which may affect their academic performance.

In addition to this, Solano and Huddleston (2020) stated that intercultural education and diversity are included in teacher education programs in Belgium, Korea, Luxembourg, Lithuania, Netherlands, New Zealand, Norway, and Switzerland. It appears that these countries recognize the importance of intercultural education and have implemented policies to ensure that teachers are trained to teach diverse student populations. Therefore, it is necessary to conduct further research to determine how teacher education correlates with student achievement. Furthermore, it is well known that education systems are difficult to get responses (Solano and Huddleston, 2020). It may be necessary to conduct longitudinal studies for this reason.

5.1.4. Performance Gap Between 1st and 2nd Generation Immigrants

In most nations, second-generation students demonstrate greater competency levels than first-generation students in PISA 2003 (Schleicher, 2006). Moreover, second-generation students outperform first-generation students in Canada, Luxembourg, Sweden, Switzerland, and Hong Kong-China (Schleicher, 2006). In these nations, there is a perceived narrowing of the achievement gap between native and immigrant students with each successive generation of immigrants (Schleicher, 2006). This pattern could be attributed to the impacts of integration policies and practices that lessen achievement gaps over time and between generations; however, it could also result from variations in the demographic composition of first- and second-generation students (Schleicher, 2006). On the other hand, Karakus *et al.* (2022) also found that first generations were mostly associated with worse academic outcomes in PISA 2018. Similar to this finding, the current study has found that second-generation have higher scores than their first-generation peers in countries that are the Czech Republic, France, Germany, Netherlands, Portugal, Slovenia, Spain, and Sweden. On the other hand, first-generation immigrant students outperformed second-generation immigrants in Latvia and Luxembourg. This implies that first-generation immigrant students outperform second-generation immigrant students in mathematics performance in some countries. Additional research is required to fully comprehend why this might be the case.

Moreover, Azzolini *et al.*, (2012) found that, in PISA 2009, first-generation immigrant students underperformed relative to native peers, while the gap between second-generation immigrant and native students was smaller in Italy and Spain. However, in

Italy, the performance gap between natives and first-generation immigrant students is smaller than in Spain, and it has been suggested that this may be due to school-related factors (Azzolini *et al.*, 2012).

5.1.5. Performance Gap After Controlling Gender and ESCS

Wang, Perry, Malpique, and Ide (2023) conducted a systematic literature review on PISA and found that 135 factors predict PISA mathematics performance. Mathematic performance has consistently been positively correlated with students' grade level and economic and social status (Wang *et al.*, 2023). According to several studies, after adjusting for background factors, such as socioeconomic status, the achievement gap between immigrant students and mainstream students narrowed or disappeared (Arikan *et al.*, 2017; Rangvid, 2007; Sakellariou, 2017). There seems to be a trend toward decreasing the effect size of the performance gap between immigrants and natives. Portugal, Luxembourg, and Estonia were the only exceptions to this trend in the current study. Notably, these three exceptions showed that a significant gap in favor of native students still exists. Therefore, these countries may need to make considerable efforts to close the achievement gap. Further research is needed to identify the underlying causes and develop effective interventions. For instance, language support is suggested because Schleicher (2006) claimed that countries with well-established language support programs with relatively clearly defined goals and standards tend to have a relatively small performance gap between immigrant and native students.

Moreover, Azzolini *et al.* (2012) found that after accounting for parental education and employment, home belongings, and family structure, the performance gap between first-generation immigrant and native students from Italy and Spain is diminished. On the other hand, the achievement gap between first- and second-generation immigrant students remained the same, except in Portugal. Figure 14 shows that Portugal is the only country with an increased effect size. This could be explained immigrant paradox. Crosnoe and Turley (2011) mentioned that when socioeconomic status is taken into account, the gap between first-generation immigrants and their third-plus peers increases, supporting the immigrant paradox. This means that first-generation immigrants in Portugal are more likely to perform better than their second-generation peers in educational outcomes. Portugal's educational system is responsible for fostering student inclusion in the Portuguese curriculum, as well as reinforcing Portuguese language instruction (Migrants: Education in Portugal for children, youth and adults, 2023). Therefore, the Portuguese as a second language course is taken by first-generation immigrants (Migrants: Education in Portugal for children, youth and adults, 2023). In addition to this, supports such as transportation and food facilities were given to refugees in Portugal (Migrants: Education in Portugal for children, youth and adults, 2023).

On the other hand, Karakus *et al.* (2022) have found that male immigrants outperformed their female immigrant peers in mathematics, reading, and science in PISA 2018. Moreover, Arıkan *et al.* (2017) have found that gender is a significant predictor of students' mathematics performance, indicating that boys were better at mathematics performance. The current study did not investigate the performance gap in between gender. However, gender was included as the control variable, and the results indicated that when socio-economic status and gender were both controlled, there was a tendency to decrease the performance gap between natives and immigrants, with some exceptions in some countries. However, no change in effect sizes was seen in the performance gap between first- and second-generation immigrants after controlling gender and economic and social status.

6. LIMITATIONS

The current study has some limitations. Firstly, the study has the limitation of only including data from the PISA 2018 application at the time of analysis. Furthermore, several countries had a limited number of immigrants, so they were excluded from the multi-level regression analysis. More countries could have been considered if more immigrants were available. Also, data was collected at a single point in time in this study, which represents one of the study's limitations. As a result, the performance of immigration over time could not be assessed based on this, which is why longitudinal research is necessary. Moreover, the results were limited to European countries, so they cannot be generalized. Additionally, missing values were excluded from the nearest neighbor matching, which may have affected the sample's representativeness.

REFERENCES

Andon, A., C. G. Thompson and B. J. Becker, 2014, “A Quantitative Synthesis of the Immigrant Achievement Gap Across OECD Countries”, *Large-scale Assessments in Education*, Vol. 2, No. 1, pp. 1-20.

Arikan, S., F. J. Van de Vijver and K. Yagmur, 2017, “PISA Mathematics and Reading Performance Differences of Mainstream European and Turkish Immigrant Students, Educational Assessment”, *Evaluation and Accountability*, Vol. 29, No. 3, pp. 229-246.

Arikan, S., F. J. Van de Vijver and K. Yagmur, 2020, “Mainstream and Immigrant Students’ Primary School Mathematics Achievement Differences in European Countries”, *European Journal of Psychology of Education*, Vol. 35, No. 4, pp. 819-837.

Azzolini, D., P. Schnell and J. R. Palmer, 2012, “Educational Achievement Gaps Between Immigrant and Native Students in Two “ New” Immigration Countries: Italy and Spain in Comparison”, *The Annals of the American Academy of Political and Social Science*, Vol. 643, No. 1, pp. 46-77.

Beaton, A. E., T. N. Postlethwaite, K. N. Ross, D. Spearritt and R. M. Wolf, 1999, *The Benefits and Limitations of International Educational Achievement Studies, United Nations Educational*, Scientific and Cultural Organization (UNESCO) International Institute for Educational Planning, Paris, France.

Burns, E. C., A. J. Martin and R. J. Collie, 2018, “Adaptability, Personal Best (PB) Goals Setting, and Gains in Students’ Academic Outcomes: A Longitudinal Examination From a Social Cognitive Perspective”, *Contemporary Educational Psychology*, Vol. 53, pp. 57-72.

Campbell, J. A., J. McIntyre and N. Kucirkova, 2021, "Gender Equality, Human Development, and PISA Results Over Time", *Social Sciences*, Vol. 10, No. 12, pp. 480-499.

Cardoso, J. B. and S. J. Thompson, 2010, "Common Themes of Resilience among Latino Immigrant Families: A Systematic Review of the Literature", *Families in Society*, Vol. 91, No. 3, pp. 257-265.

Cerna, L., O. Brussino and C. Mezzanotte, 2021, "The Resilience of Students with an Immigrant Background", *Organisation for Economic Co-operation and Development*, Vol. 1, pp. 261-278.

Cicchetti, D. V., 1994, "Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology", *Psychological Assessment*, Vol. 6, No. 4, pp. 284-290.

Chang, J. and T. T. Le, 2010, "Multiculturalism as a Dimension of School Climate: The Impact on the Academic Achievement of Asian American and Hispanic Youth", *Cultural Diversity and Ethnic Minority Psychology*, Vol. 4, No. 16, pp. 485-492.

Clarke, M. and D. Luna-Bazaldua, 2021, *Primer on Large-Scale Assessments of Educational Achievement*, World Bank Publications, Washington, District of Columbia.

Cohen, J., E. M. Mccabe, N. M. Michelli and T. Pickeral, 2009, "School Climate: Research, Policy, Practice and Teacher Education", *Teachers College Record*, Vol. 111, No. 1, pp. 180-213.

Collie, R. J. and A. J. Martin, 2017, "Students' Adaptability in Mathematics: Examining Self-Reports and Teachers' Reports and Links with Engagement and Achievement Outcomes", *Contemporary Educational Psychology*, Vol. 49, pp. 355-366.

Crosnoe, R. and R. N. L. Turley, 2011, "K-12 Educational Outcomes of Immigrant Youth", *The Future of Children*, Vol. 21, No. 1, pp. 129-145.

Dennis, J. P. and J. S. Vander Wal, 2010, "The Cognitive Flexibility Inventory: Instrument Development and Estimates of Reliability and Validity", *Cognitive Therapy and Research*, Vol. 34, No. 3, pp. 241-253.

Diamond, A. and K. Lee, 2011, "Interventions Shown to Aid Executive Function Development in Children 4 To 12 Years Old", *Science*, Vol. 6045, No. 333, pp. 959-964.

Dronkers, J. and M. De Heus, 2016, "The Educational Performance of Children of Immigrants in Sixteen OECD Countries", *Adjusting to a World In Motion: Trends in Global Migration and Migration Policy*, Vol. 23, pp. 264-290.

Elliot, A. J. and H. A. McGregor, 2001, "A 2x2 Achievement Goal Framework", *Journal of Personality and Social Psychology*, Vol. 80, No. 3, pp. 501-519.

Farayibi, A. O. and O. Folarin, 2021, "Does Government Education Expenditure Affect Educational Outcomes? New Evidence from Sub-Saharan African Countries", *African Development Review*, Vol. 33, No. 3, pp. 546-559.

Field, A., 2009, *Discovering Statistics Using SPSS: And Sex and Drugs and Rock N Roll*, Sage Publications, Los Angeles, London.

Field, A., J. Miles and Z. Field, 2012, *Discovering Statistics Using R.*, Sage Publications, Research and Development (RAND) Corporation, United States of America.

Figlio, D. N., P. Giuliano, R. Marchingiglio, U. Ozek and P. Sapienza, 2021, *Diversity in Schools: Immigrants and the Educational Performance of US Born Students*, National Bureau of Economic Research, Cambridge, Massachusetts.

Finnan, C., K. Schnepel and L. Anderson, 2003, "Powerful Learning Environments: The Critical Link between School And Classroom Cultures", *Journal of Education for Students Placed at Risk*, Vol. 8, pp. 391-418.

Greaney, V. and T. Kellaghan, 2008, *Assessing National Achievement Levels In Education*, World Bank Publications, Washington, District of Columbia.

Goodenow, C. and K. E. Crady, 1997, "The Relationship of School Belonging and Friends' Values to Academic Motivation Among Urban Adolescent Students", *Journal of Experimental Education*, Vol. 62, pp. 60-71.

Graham, S., 2006, "Peer Victimization in School: Exploring the Ethnic Context", *Current Directions in Psychological Science*, Vol. 15, No. 6, pp. 317-321.

Guerra, R., R. B. Rodrigues, C. Aguiar, M. Carmona, J. Alexandre and R. C. Lopes, 2019, "School Achievement and Well-Being of Immigrant Children: The Role of Acculturation Orientations and Perceived Discrimination", *Journal of School Psychology*, Vol. 75, pp. 104-118.

Hanushek, E. A. and D. D. Kimko, 2000, "Schooling, Labor-Force Quality, and the Growth of Nations", *American Economic Review*, Vol. 90, No. 5, pp. 1184-1208.

He, J., F. J. R. V. De Vijver and A. Kulikova, 2017, "Country-Level Correlates of Educational Achievement: Evidence from Large-Scale Surveys", *Educational Research and Evaluation*, Vol. 23, No. 5-6, pp. 163-179.

Heiman, G., 2011, *Basic Statistics for the Behavioral Sciences*, Cengage Learning, Belmont, United States.

Ho, D., K. Imai, G. King and E. A. Stuart, 2011, "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference", *Journal of Statistical Software*, Vol. 42, No. 8, pp. 1-28.

Hoge, D. R., E. K. Smit and S. L. Hanson, 1990, "School Experiences Predicting Changes in Self-Esteem of Sixth-And Seventh-Grade Students", *Journal of Educational Psychology*, Vol. 82, No. 1, pp. 117-127.

Hox, J. J., 2013, "Multilevel Regression and Multilevel Structural Equation Modeling", *The Oxford Handbook of Quantitative Methods*, Vol. 2, No. 1, pp. 281-294.

Hox, J. J., M. Moerbeek and R. Van De Schoot, 2017, *Multilevel Analysis: Techniques And Applications*, Routledge, New York.

International Energy Agency (IEA), 2022, *Help Manual for the International Energy Agency (IEA)*, International Development Bank (IDB) Analyzer, Hamburg, Germany.

International Organization of Migration, 2020, "World Migration Report", *Journal of Technical Writing and Communication*, Vol. 52, No. 1, pp. 57-93.

Jackson, C. K., R. C. Johnson, and C. Persico, 2015, "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms", *National Bureau of Economic Research*, Vol. 131, No. 1, pp. 157-218.

Juvonen, J. and S. Graham, 2014, "Bullying In Schools: The Power of Bullies and the Plight of Victims", *Annual Review of Psychology*, Vol. 65, pp. 159-185.

Karakus, M., M. Courtney and H. Aydin, 2022, "Understanding the Academic Achievement of the First-and Second-Generation Immigrant Students: A Multi-Level Analysis of PISA 2018 Data", *Educational Assessment, Evaluation and Accountability*, Vol. 23, pp. 1-46.

Korpershoek, H., 2016, “Relationships among Motivation, Commitment, Cognitive Capacities, and Academic Achievement in Secondary Education”, *Frontline Learning Research*, Vol. 4, No. 3, pp. 28-43.

Martin, M. M. and R. B. Rubin, 1995, “A New Measure of Cognitive Flexibility”, *Psychological Reports*, Vol. 76, No. 2, pp. 623-626.

Martin, A. J., G. A. Liem, M. Mok and J. Xu, 2012, “Problem Solving and Immigrant Student Mathematics and Science Achievement: Multination Findings from the Programme for International Student Assessment (PISA)”, *Journal of Educational Psychology*, Vol. 104, No. 4, pp. 1054-1075.

Martin, A. J., H. Nejad, S. Colmar and G. A. D. Liem, 2012, “Adaptability: Conceptual and Empirical Perspectives on Responses to Change, Novelty and Uncertainty”, *Australian Journal of Guidance and Counselling*, Vol. 22. No. 1, pp. 58-81.

Martin, A. J., H. G. Nejad, S. H. Colmar and G. A. D. Liem, 2013, “Adaptability: How students’ Responses to Uncertainty and Novelty Predict Their Academic and Nonacademic Outcomes”, *Journal of Educational Psychology*, Vol. 105, pp. 728-746.

Mazurek, J., C. F. García and C. P. Rico, 2021, “Inequality and Students’ PISA 2018 Performance: A Cross-Country Study, Comparative Economic Research”, *Central and Eastern Europe*, Vol. 24, No. 3, pp. 163-183.

McClelland, D. C., 1985, “Public Perception of the Motives that Lead Political Leaders to Launch Interstate Armed Conflicts: A structural and Cross-Cultural Study”, *Universitas Psychologica*, Vol. 12, No. 2, pp. 327-346.

McInerney, D. M., 2008, “Personal Investment, Culture and Learning: Insights into School Achievement Across Anglo, Aboriginal, Asian and Lebanese Students in Australia”, *International Journal of Psychology*, Vol. 43, pp. 870-879.

Moore, L. L., D. K. Grabsch and C. Rotter, 2010, "Using Achievement Motivation Theory to Explain Student Participation in a Residential Leadership Learning Community", *Journal of Leadership Education*, Vol. 9, No. 2, pp. 22-34.

Migrants: "Education in Portugal for children, youth and adults", 2023, https:// eportugal. gov. pt/ en/ migrantes- viver- e - trabalhar- em-portugal/ migrantes- ensino-em- portugal- para- criancas- jovens- e- adultos, accessed on August 11, 2023.

Napierala, M. A., 2012, "What is the Bonferroni correction?", *Aaos Now*, Vol. 16, pp. 40-41.

National School Climate Center, 2023, "School Climate Services. Retrieved from", https:// schoolclimate. org/, accessed on June 21, 2023.

Organization for Economic Cooperation and Development (OECD), 2006, *Where Immigrant Students Succeed: A Comparative Review of Performance and Engagement in PISA 2003, Program for International Student Assessment (PISA)* Organization for Economic Cooperation and Development (OECD) Publishing, Paris.

Organization for Economic Cooperation and Development (OECD), 2010, *PISA 2009 Results: Overcoming Social Background - Equity in Learning Opportunities and Outcomes*, Overcoming Social, Germany.

Organization for Economic Cooperation and Development (OECD), 2013a, *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy, Program for International Student Assessment (PISA)*, Organization for Economic Cooperation and Development (OECD) Publishing, Paris.

Organization for Economic Cooperation and Development (OECD), 2013b. *PISA 2012 Results: Ready to Learn (volume III): Students' Engagement, Drive, and Self-Beliefs, Program for International Student Assessment (PISA)*, Organization for

Economic Cooperation and Development (OECD) Publishing, Germany.

Organization for Economic Cooperation and Development (OECD), 2014, PISA 2012 Technical Report, Organization for Economic Cooperation and Development (OECD) Publishing, Paris.

Organization for Economic Cooperation and Development (OECD), 2018a, *Working Together for Local Integration of Migrants and Refugees*, Organization for Economic Cooperation and Development (OECD) Publishing, Paris.

Organization for Economic Cooperation and Development (OECD), 2018b, "PISA 2018 Technical Report, Organization for Economic Cooperation and Development (OECD) Publications", https://https://www.oecd.org/pisa/data/ pisa2018 technicalreport, accessed on May 25, 2023.

Organization for Economic Cooperation and Development (OECD), 2019a, *PISA 2018 Results (Volume I): What Students Know and Can Do, Program for International Student Assessment (PISA)*, Organization for Economic Cooperation and Development (OECD) Publishing, Paris.

Organization for Economic Cooperation and Development (OECD), 2019b, *PISA 2018 Results (Volume III): What School Life Means for Students' Lives, Program for International Student Assessment (PISA)*, Organization for Economic Cooperation and Development (OECD) Publishing, Paris.

Organization for Economic Cooperation and Development (OECD), 2019c, *PISA 2018 Results (Volume II): Where All Students Can Succeed, Program for International Student Assessment (PISA)*, Organization for Economic Cooperation and Development (OECD) Publishing, Paris.

Organization for Economic Cooperation and Development (OECD), 2019d, *PISA 2018*

Assessment and Analytical Framework, Program for International Student Assessment (PISA), Organization for Economic Cooperation and Development (OECD) Publishing, Paris.

Organization for Economic Cooperation and Development (OECD), 2020, *PISA 2018 Results (Volume VI): Are Students Ready to Thrive in an Interconnected World?, Program for International Student Assessment (PISA)*, Organization for Economic Cooperation and Development (OECD) Publishing, Paris.

Perry, A., 1908, *The Management Of A City School*, Macmillan, New York.

Pickeral, T., L. Evans, W. Hughes and D. Hutchison, 2009, *School Climate Guide for District*, Policymakers and Educational Leaders, New York.

Pivovarova, M. and J. M. Powers, 2019, "Generational Status, Immigrant Concentration and Academic Achievement: Comparing First and Second-Generation Immigrants with Third-Plus Generation Students", *Large-scale Assessments in Education*, Vol. 7, No. 1, pp. 1-18.

Ponzo, M., 2013, "Does Bullying Reduce Educational Achievement? An Evaluation Using Matching Estimators", *Journal of Policy Modeling*, Vol. 35, No. 6, pp. 1057-1078.

Randolph, J. J. and K. Falbe, 2014, "A Step-By-Step Guide to Propensity Score Matching in R. Practical Assessment", *Research and Evaluation*, Vol. 19, pp. 45-67.

Rangvid, B. S. 2007, "Sources of Immigrants' Underachievement: Results from PISA-Copenhagen", *Education Economics*, Vol. 15, No. 3, pp. 293-326.

Rodríguez, S., A. Valle, L. M. Gironelli, E. Guerrero, B. Regueiro and I. Estévez, 2020, "Performance and Well-Being of Native and Immigrant Students, Comparative

Analysis Based on PISA 2018", *Journal of Adolescence*, Vol. 85, pp. 96-105.

Rosenbaum, P. R. and D. B. Rubin, 1983, "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, Vol. 70, No. 1, pp. 41-55.

Rubin, D. B., 2001, "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation", *Health Services and Outcomes Research Methodology*, Vol. 2, No. 3, pp. 169-188.

Rust, J. P., M. A. Jackson, J. G. Ponterotto and F. C. Blumberg, 2011, "Biculturalism and Academic Achievement of African American High School Students", *Journal of Multicultural Counseling and Development*, Vol. 39, No. 3, pp. 130-140.

Rutkowski, L., E. Gonzalez, M. Joncas and M. Von Davier, 2010, "International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting", *Educational Researcher*, Vol. 39, No. 2, pp. 142-151.

Sakellariou, C., 2017, "The Immigrant-Native Student Educational Achievement Gap in Greece Using PISA 2012", Athens *Journal of Business and Economics*, Vol. 3, No. 3, pp. 221-238.

Schafer, J. L., 1999, "Multiple Imputation: A primer", *Statistical Methods in Medical Research*, Vol. 8, No. 1, pp. 3-15.

Schleicher, A., 2006, "Where Immigrant Students Succeed: A Comparative Review of Performance and Engagement in PISA 2003: Organization for Economic Cooperation and Development (OECD) 2006", *Intercultural Education*, Vol. 17, No. 5, pp. 507-516.

Schulz, W., J. Ainley and J. Fraillon, 2011, *ICCS 2009 Technical Report, Amsterdam*, The Netherlands: International Association for the Evaluation of Educational

Achievement (IEA), Roma.

Sirin, S. R. and L. Rogers-Sirin, 2004, "Exploring School Engagement of Middle-Class African American Adolescents", *Youth and Society*, Vol. 35, No. 3, pp. 323-340.

Solano, G. and T. Huddleston, 2020, *Migrant Integration Policy Index 2020, Barcelona/ Brussels: Barcelona Centre for International Affairs (CIDOB)*, Miles Per Gallon (MPG), North, Macedonia.

Stewart, E. B., 2007, "School Structural Characteristics, Student Effort, Peer Associations, and Parental Involvement", *Education and Urban Society*, Vol. 2, No. 40, pp. 179-204.

Tabachnick, B. G. and L. S. Fidell, 2007, U*sing Multivariate Statistics*, Pearson, Boston, Massachusetts, United States.

United Nations Development Programme (UNDP), 2014, *Human Development Report 2014: Sustaining Human Progress: Reducing Vulnerabilities*, Building Resilience, New York.

Walberg, H. J., 2004, "Improving Educational Productivity: An Assessment of Extant Research", *The Last Song Syndrome (LSS) Review*, Vol. 3, No. 2, pp. 11-14.

Wang, X. S., L. B. Perry, A. Malpique, and T. Ide, 2023, "Factors Predicting Mathematics Achievement in PISA: A Systematic Review", *Large-Scale Assessments in Education*, Vol. 11, No. 1, pp. 24-42.

World Bank 2023a, "GDP per capita", https:// data. worldbank. org/ indicator/, accessed on July 21, 2023.

Wentzel, K. R., S. Jablansky and N. R. Scalise, 2021, "Peer Social Acceptance and Academic Achievement: A Meta-Analytic Study", *Journal of Educational Psychology*, Vol. 113, No. 1, pp. 157-177.

World Bank, 2022b, "Governmental Expenditure on Education Retrieved from", https:// data. worldbank. org/ indicator/, accessed on July 28, 2023.

Zelazo, P. D. and S. M. Carlson, 2012, "Hot and Cool Executive Function in Childhood and Adolescence: Development and Plasticity", *Child Development Perspectives*, Vol. 6, No. 4, pp. 354-360.

APPENDIX A: SAMPLE PAGES

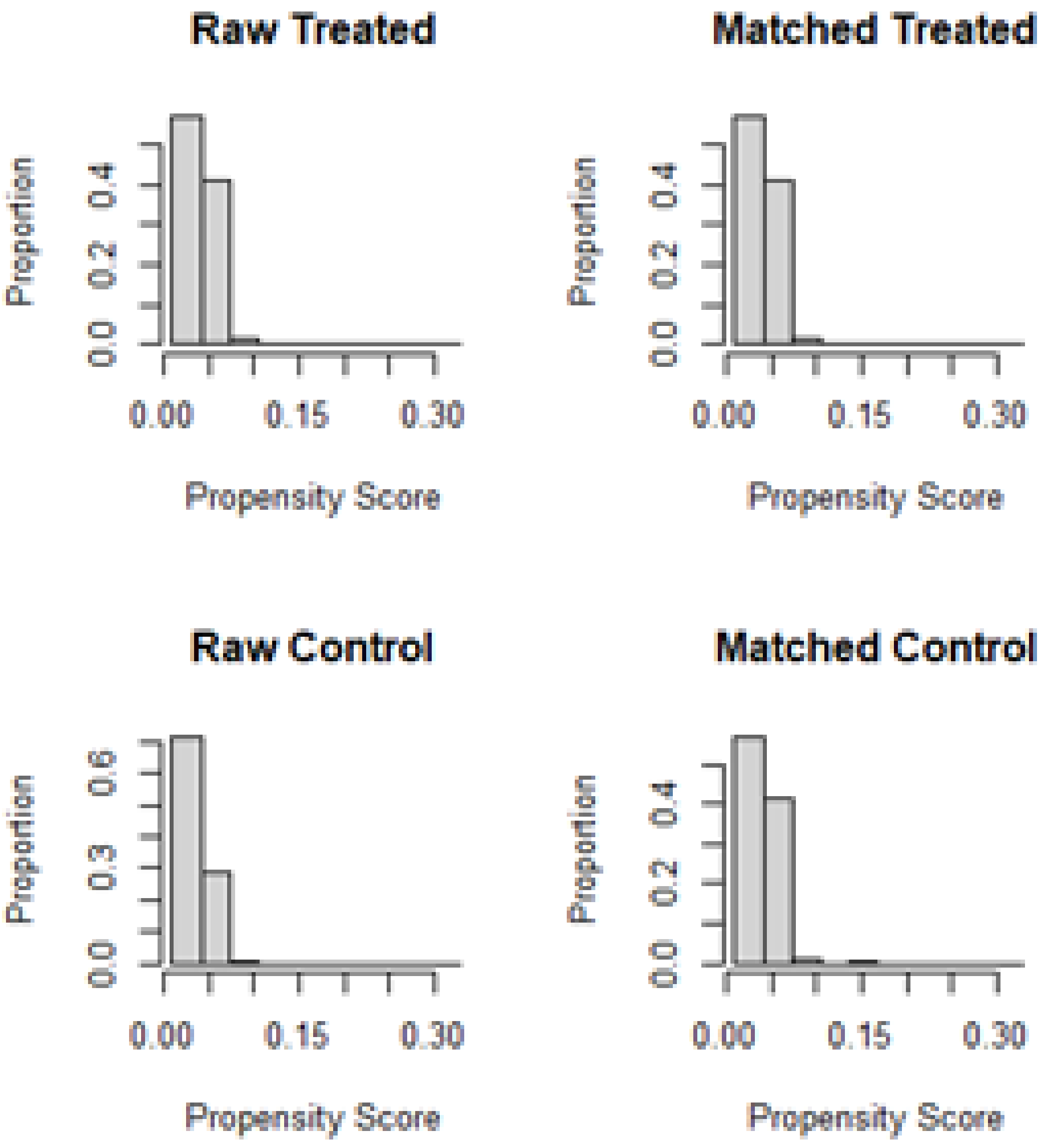

Figure A.1. Czech republic.

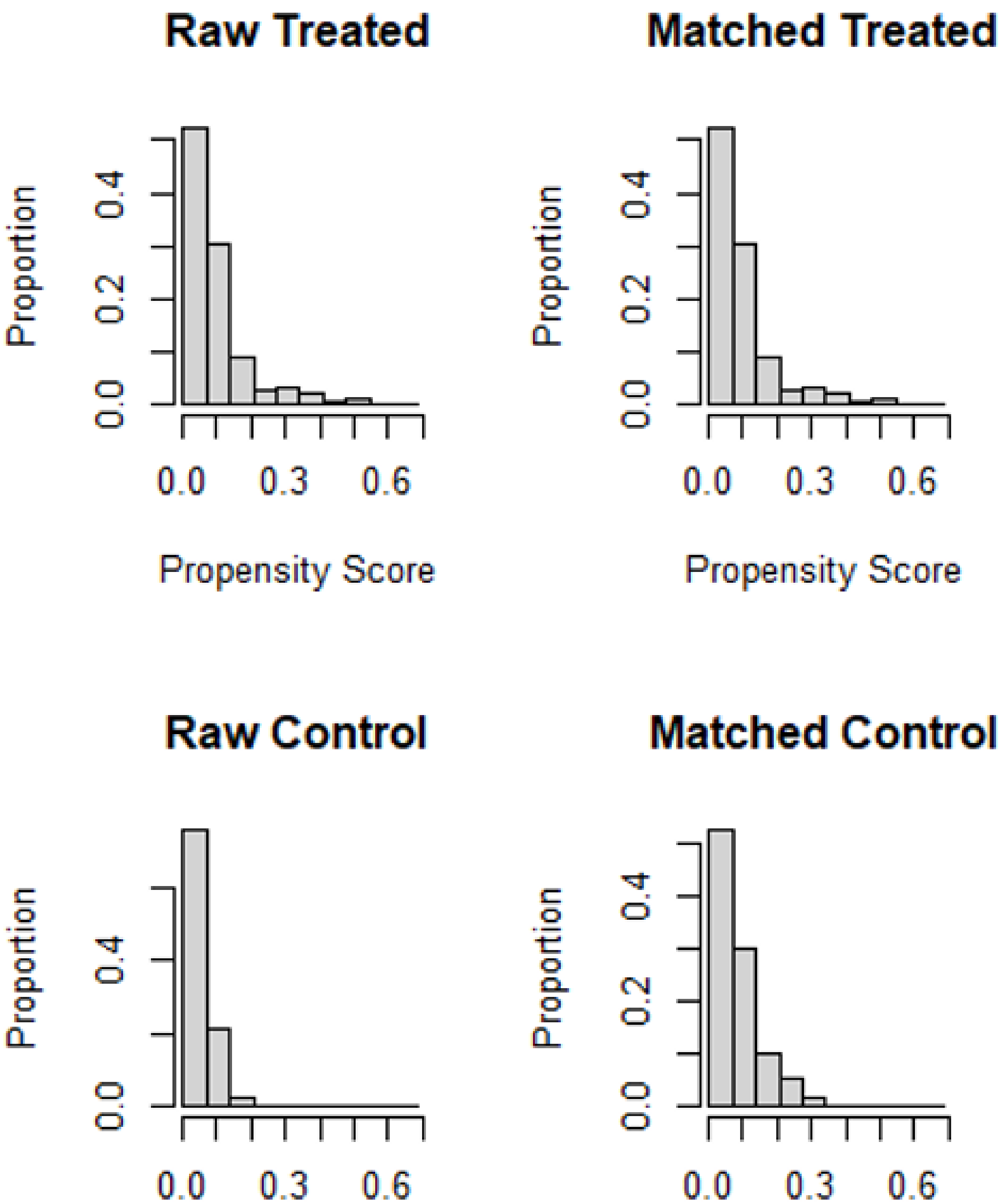

Figure A.2. Finland.

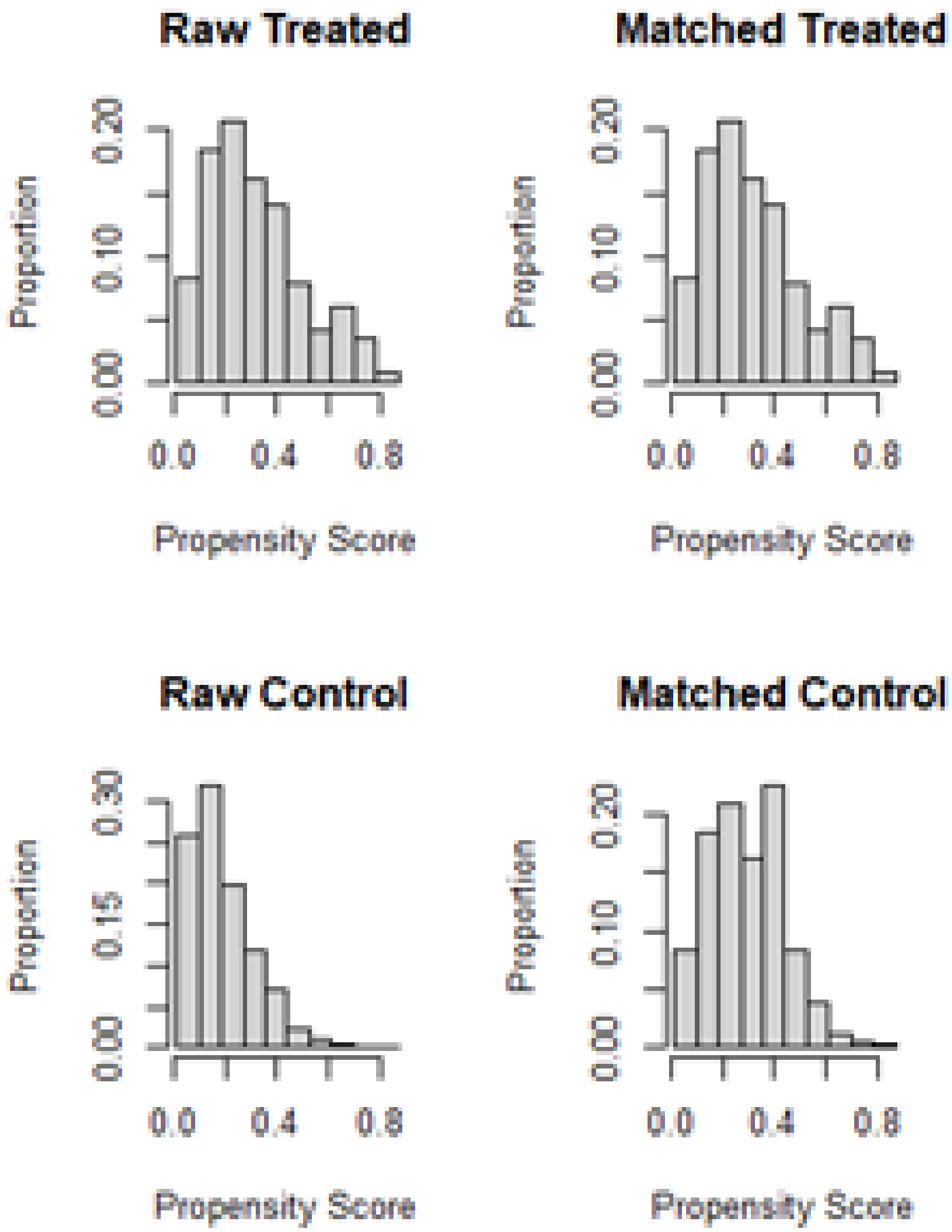

Figure A.3. Germany.

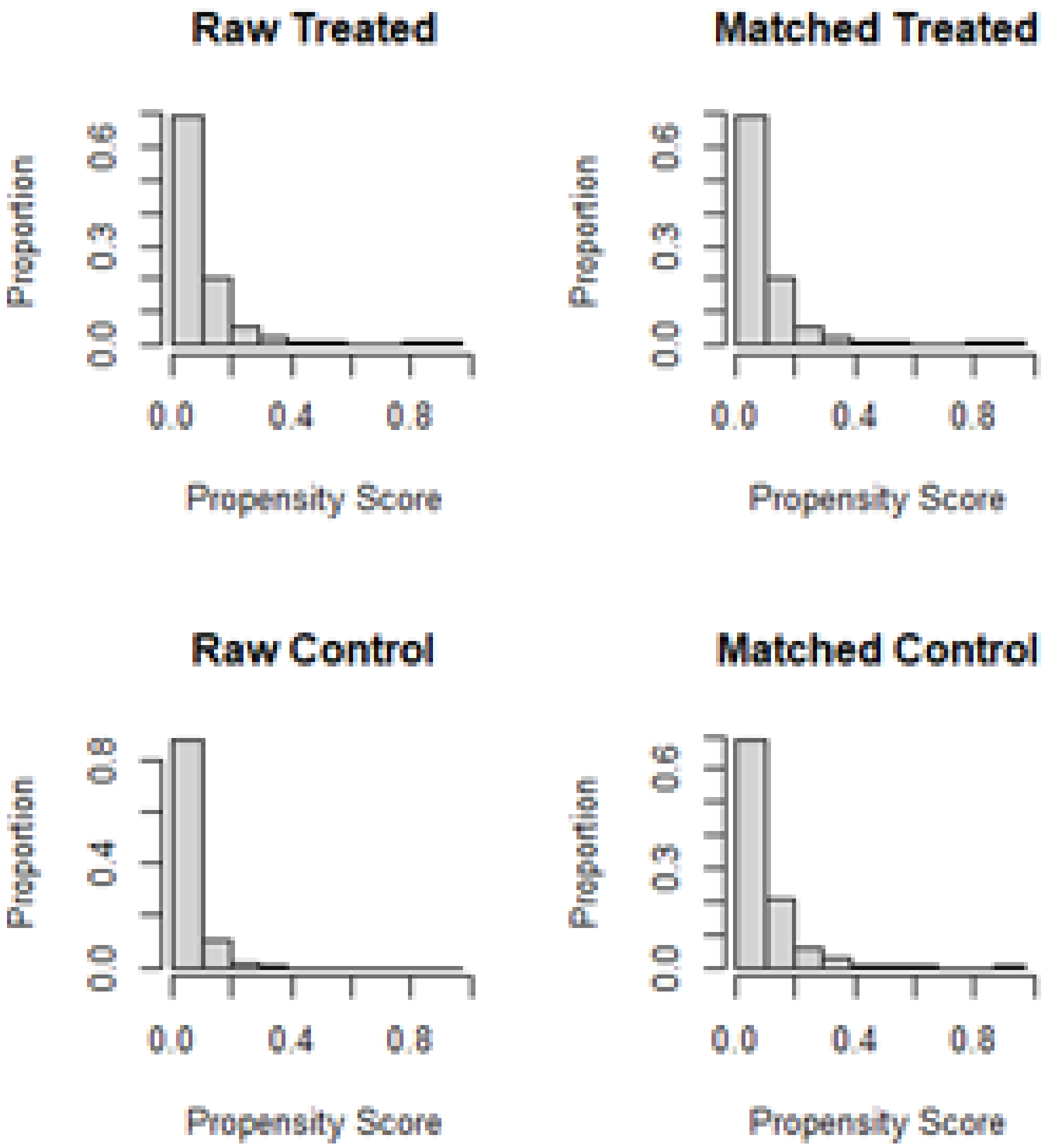

Figure A.4. Iceland.

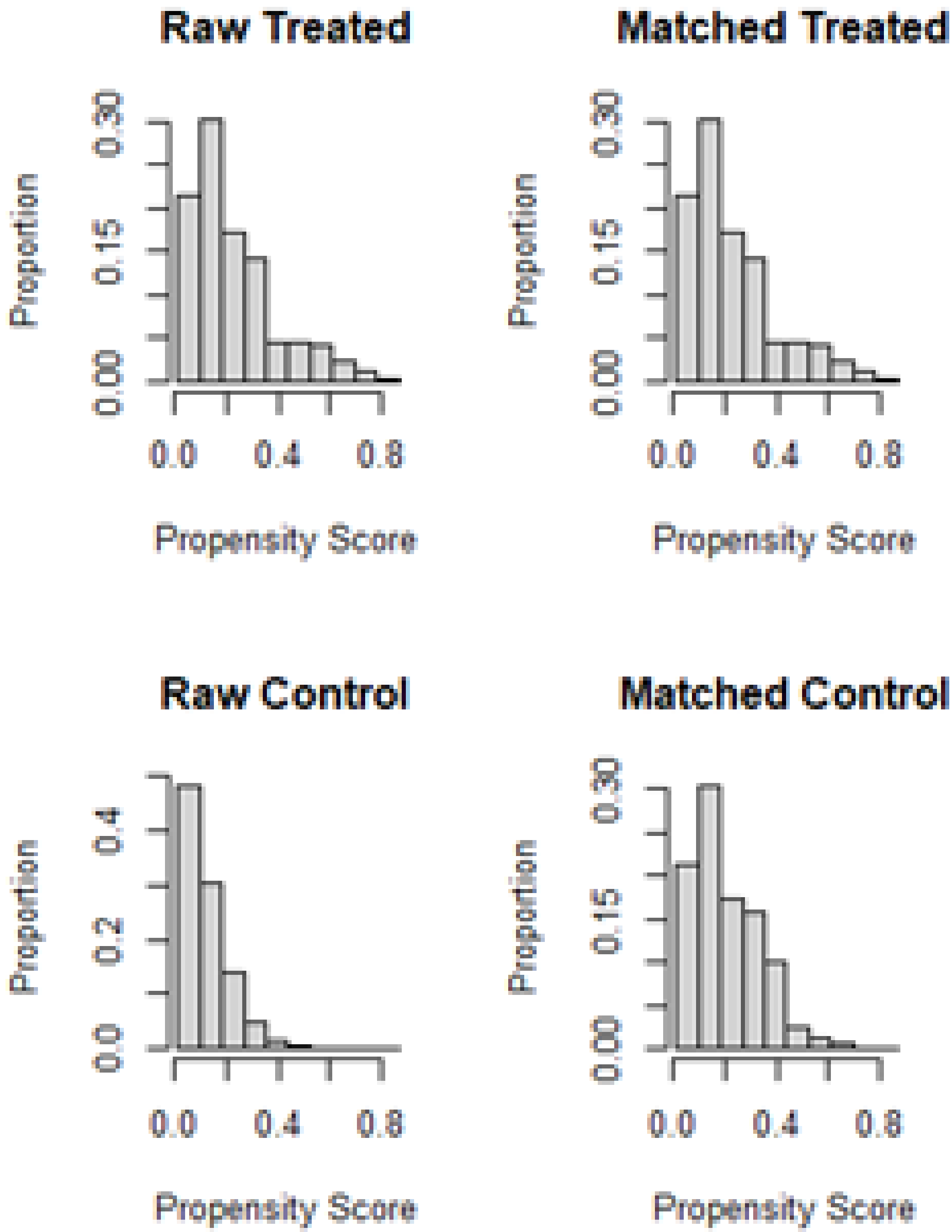

Figure A.5. Netherlands.

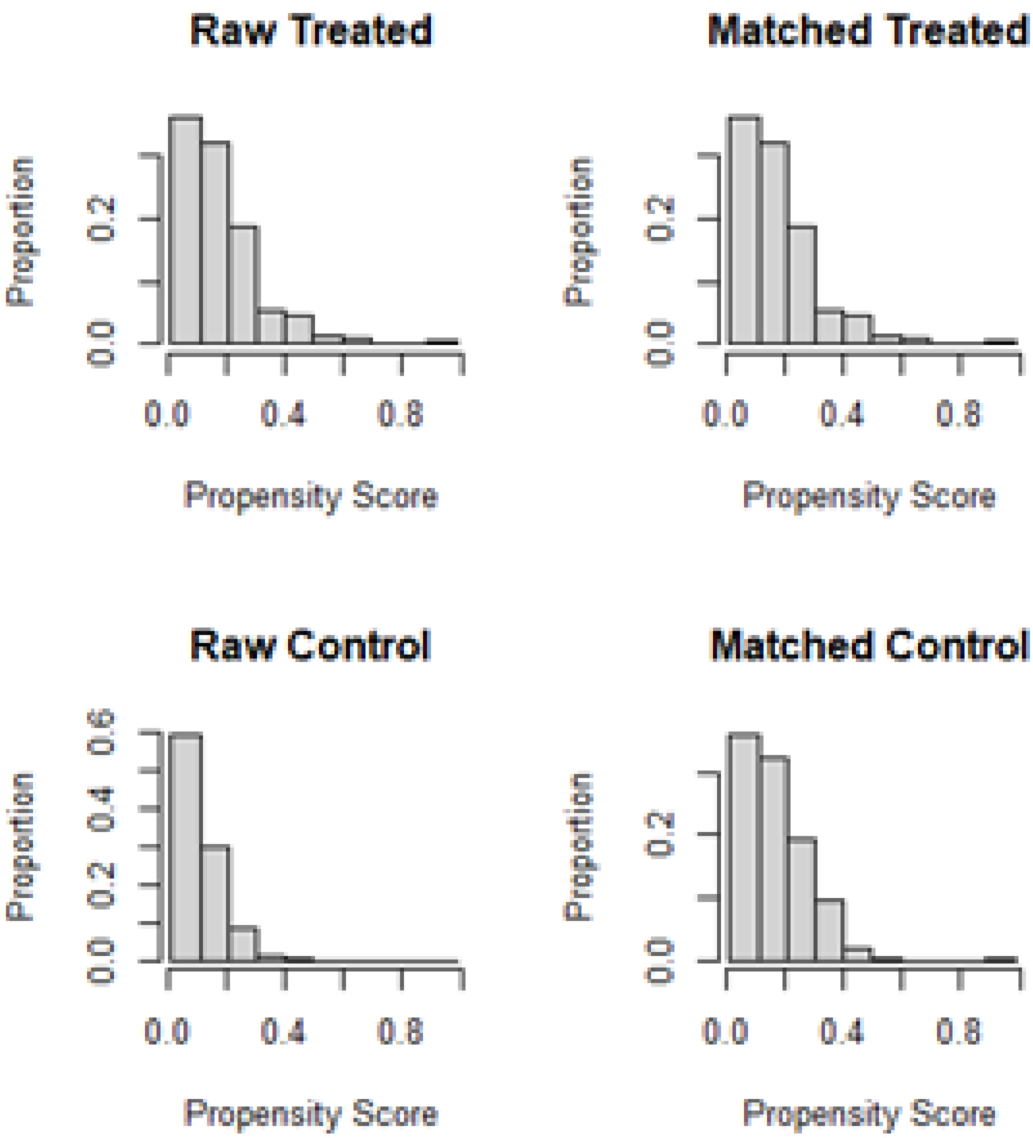

Figure A.6. Norway.

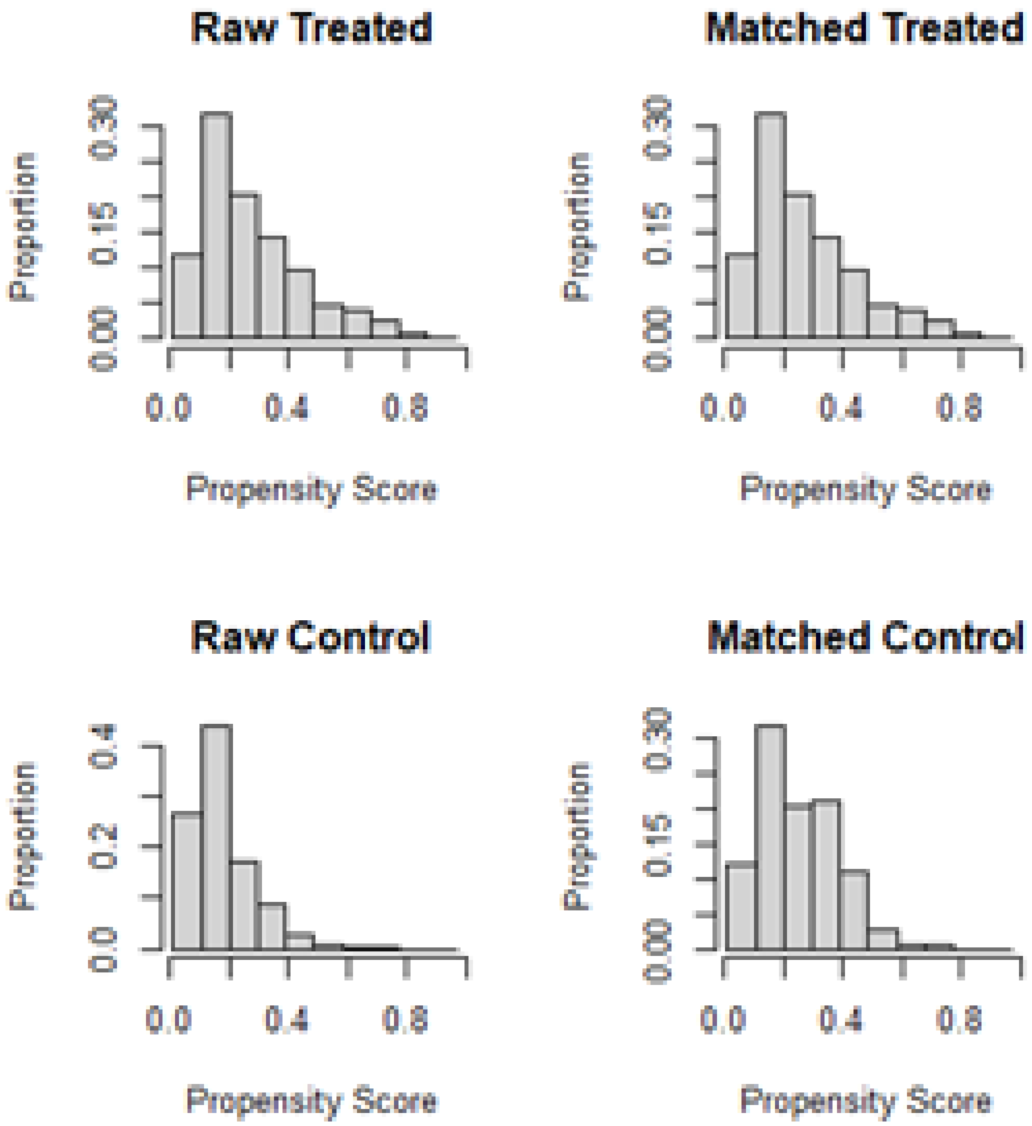

Figure A.7. Sweden.

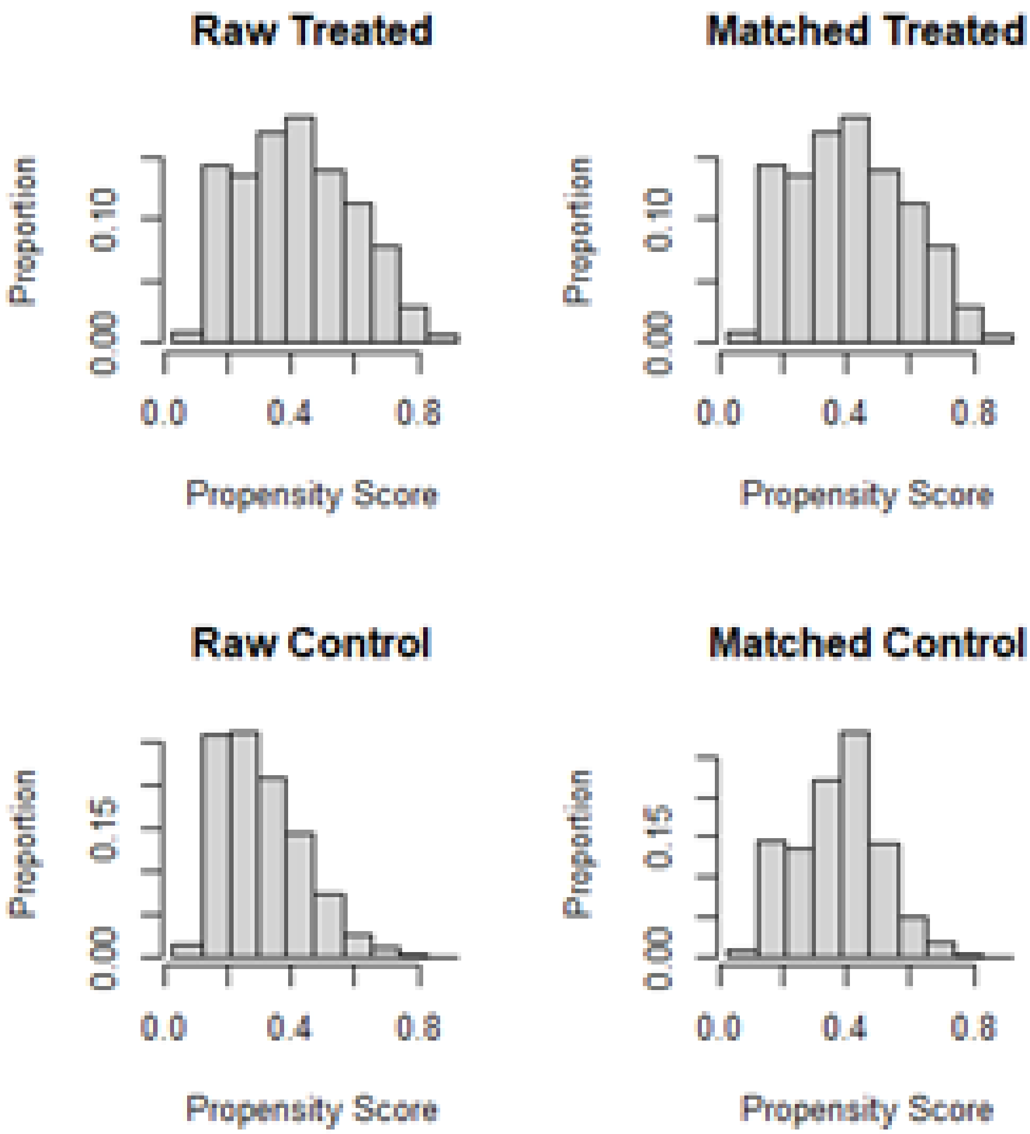

Figure A.8. Switzerland.