university of groningen / faculty of behavioural and social sciences

Social integration through sporting for immigrants in the Netherlands

A quantitative analysis of the effect of sporting on the social integration of men and women of the immigrant community

Bachelor Thesis Sociology

Student: Christel Meijer (S4778758)
c.m.meijer.2@student.rug.nl

Supervisor: F. Giardini

Second reader: R. Smaniotto

Course: Bachelorwerkstuk (SOBA313A)

Date: August 19th 2024

Abstract

The well-being of immigrants is, on average, lower than the well-being of the local population in the Netherlands (Hendriks & Burger, 2019). Since the well-being of people is linked to their social integration, it is important to research how the social integration of immigrants into a host society could be stimulated. In this research, the effect of doing sports on the social integration of men and women in the immigrant community is examined. This is formulated in the following research question: *What is the effect of sporting on the social integration of immigrants and what role does gender play*? The following two hypotheses have been formulated according to the existing literature: *'Sporting has a positive effect on the social integration of immigrants' and 'The positive effect of sporting on the social integration of immigrants is larger for men than for women'.* A hierarchical linear regression analysis has been done to test the hypotheses using data from the LISS Immigrant Panel with a sample size of 1232 respondents. The results of the linear regression analysis find evidence that supports the first hypothesis. Not enough evidence is found to support the second hypothesis. This research concludes that sporting has a positive effect on the integration of immigrants into a host society. No significant differences were found between men and women and how sports positively contribute to their social integration. In future research, a more precise definition of sporting and social integration of immigrants should be used in order to test for more specific hypotheses. An important policy recommendation includes the advice to focus on informing immigrants about existing sports programs and to invest in sports organisations that actively work to attract immigrants.

Table of Contents

1. Introduction

People moving to a new country represents a global trend nowadays. The Netherlands seems to be a popular destination for immigrants and experiences the growing cultural diversity in its population (Tuturea et al., 2023). People from all over Europe and beyond have come to the Netherlands, whether it is for work, education, or many other reasons. As a result, in 2022 about 27 percent of the Dutch population is either born abroad or has a parent who is born abroad (Statistics Netherlands, 2022).

The Netherlands has a rich immigration history with many migration streams adding to this in the last century. After the Second World War, many immigrants arrived in the Netherlands from Indonesia, the Moluccas, and Suriname (Jennissen et al., 2022). Additionally, many people from the Dutch Antilles migrated to the Netherlands in the late twentieth century due to favourable economic prospects (Jennissen et al., 2022). Furthermore, some migration streams include immigrants from Iraq, Afghanistan and Iran who came to the Netherlands from the 1990s onwards (McGinnity & Gijsberts, 2017). Currently, most of the immigrants who come to the Netherlands are from Poland, Turkey, Germany, Spain and Romania (Statistics Netherlands, n.d.-a). Some of these countries have gained EU membership in the past two decades, which allows the free flow of people to other member states, including the Netherlands (European Union, n.d.). Most of these contemporary immigrants are here for employment, family reunification and education. The Netherlands has also implemented policies that favour the inflow of immigrants, for instance, to attract highly qualified immigrants (Vermeulen & Verweel, 2009; Ager & Strang, 2008). Such immigration developments have shaped the Dutch population to the present. The Dutch government must design policies that facilitate good integration of immigrants into society, for the well-being of immigrants and the broader Dutch population.

The well-being of immigrants is impacted by their social integration into the host society, as seen in earlier academic work (Doherty & Taylor, 2007; Haslam et al., 2020). In a study by Ager and Strang (2008), being socially integrated is defined as having social contacts, cultural knowledge and feeling welcome in society. Doherty and Taylor (2007) add to this definition the feeling of being part of the community where they reside and fitting into the environment of the host society. It is hereby important for immigrants to have a diverse set of social networks, which includes social connections with the broader society, not solely having connections with a particular social group (Schinkel, 2013; Lee et al., 2019). A common one-sided view on this matter expects newcomers to adapt to the dominant standards regarding norms and behaviour in the host society (Li, 2003; Schinkel, 2013; Smith et al., 2018). As a part of the adaptation process, immigrants are expected to put effort into

learning the local language, becoming economically integrated, acquiring values and participating in social activities (Li, 2003; Grip, 2019). However, existing language barriers, insufficient cultural knowledge, and prejudice can for example inhibit immigrants from integrating into the host society successfully (McGinnity & Gijsberts, 2017). To achieve good well-being and health among the immigrant community, governmental measures that stimulate their social integration are important. Therefore, it is relevant to further investigate how social integration can be stimulated and what activities can contribute to this process.

One activity that is known to create opportunities to connect with other people and to invest in one's health and well-being is doing sports (Vermeulen & Verweel, 2009). The Dutch government has, to a certain extent, addressed the importance of doing sports in its societal plans to stimulate social integration and enhance public health (Smith et al., 2018). To start with, being physically active is beneficial for the cognitive and physical health of individuals (Waring & Mason, 2010). Besides that, doing sports can facilitate building or strengthening connections with other people in society. It offers an opportunity to exchange cultural knowledge and learn about behavioural norms. Such benefits are key to stimulating social integration, but not much is known about the effects on the social integration of male and female immigrants.

Men and women show differences in how they become socially integrated into a host society. For example, men seem to create friendships more easily than women in the context of doing shared physical activities like sports (Hall, 2011). Besides that, women have a greater need for intimacy than men in their social relationships, which can for example be found in bonding over conversations rather than doing a shared physical activity (Hall, 2011; Athenstaedt et al., 2008). Therefore, doing sports could have greater benefits to men in comparison to women when it comes to their social integration process.

In short, the immigrants that reside in the Netherlands could face difficulties in becoming socially integrated. Being socially integrated is associated with health benefits and greater well-being and is therefore an important objective within national policy. To enhance the integration of immigrants into society, doing sports could add to this objective. As men and women have different needs and preferences, doing sports might have various outcomes in their process of becoming socially integrated. This research will therefore take on the following research question:

What is the effect of sporting on the social integration of immigrants and what role does gender play?

Figure 1 shows the research model and the concepts that are involved in this research.

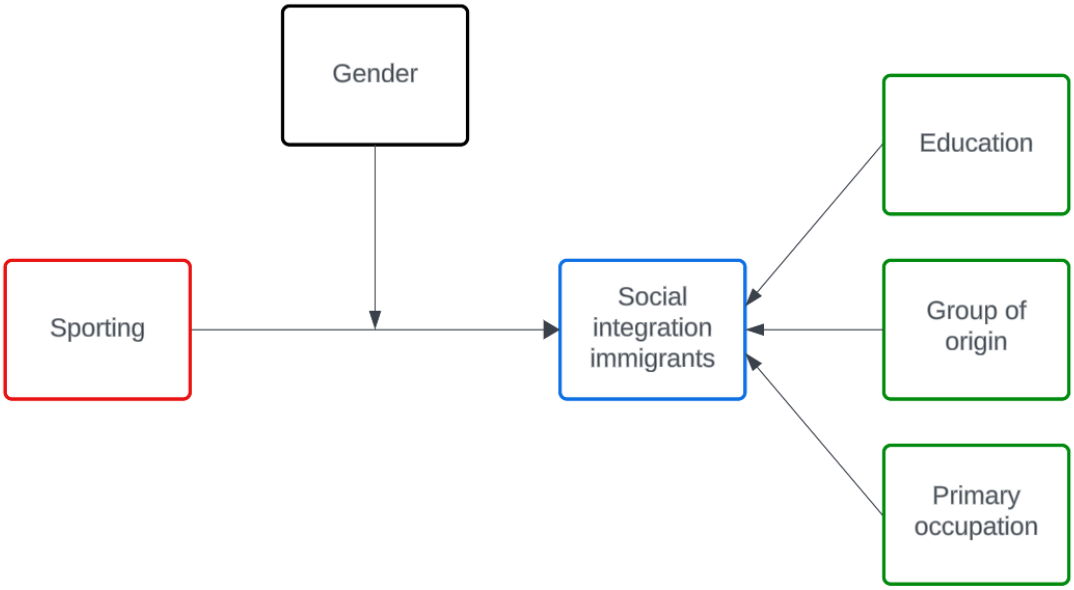


Figure 1 Research model with the independent variable, dependent variable, three control variables and moderator

2. Theoretical framework

Sporting can be either done individually or in a group context. Although both possibilities can contribute to integrating into a new society, sporting in a group context however offers people to gain new knowledge, skills and social contacts from interacting with others. These benefits are of great importance, as they make up the essential parts of being socially integrated. These positive effects of both doing individual sports or sports in a group context are elaborated further. Moreover, the differences in doing sports between men and women are discussed and the control variables in this research will be explained briefly.

2.1 Sporting and integrating into society

A well-known benefit of doing sports is that it adds to people's physical and cognitive well-being. Much evidence can be found on the relationship between sporting and health benefits (Doherty & Taylor, 2007; Holt et al., 2016; Waring & Mason, 2010). To name a few health benefits, sporting enhances positive feelings and mood states, improves cognitive functions, can reduce stress levels and enhances cardiovascular health (Bailey, 2016). As a result of doing sports, people feel healthier and are in a better physical state (Doherty & Taylor, 2007). Relating to integration into society, Barker-Ruchti et al. (2013) argue that maintaining your health through sports can stimulate social integration and reduce the constraints of health-related issues. The positive effect of sports on one's physical health supports the mobilisation of people and therefore facilitates socialization (Kim et al., 2019).

Practising sports can also assist in creating a positive image of one's own identity (Doherty & Taylor, 2007; Waring & Mason, 2010; Videon, 2002). Sporting is a way to challenge one's self to become good at something and have the control to achieve goals about one's physique. Engagement in sports can create opportunities to develop self-esteem and a positive image of one's self (Doherty & Taylor, 2007; Waring & Mason, 2010; Cherney & London, 2006). Doherty & Taylor (2007) argue that doing sports can help immigrants feel less isolated and can help with depression. Positive self-identity is seen to benefit creating and strengthening social contacts. Therefore, a positive self-image is beneficial for immigrants to integrate into society.

Moreover, immigrants can create and strengthen social connections and benefit from the social support of others to better navigate society (Holt et al., 2016). Earlier academic work suggests that doing sports can facilitate building and maintaining social relationships (Morela et al., 2013; Doherty & Taylor, 2007; Pringle & Liu, 2023). Sporting can provide a context where people meet each other regularly and perform a shared activity, which stimulates building and maintaining friendships (Mollenhorst et al., 2014). It is argued that socialising with friends leads to greater happiness and improved health (Shea & Beausoleil, 2011). Besides that, social relationships shape the behaviour of

people, thus encouraging immigrants for example to conform to social norms that exist in the host society (Holt-Lunstad et al., 2010). Additionally, being part of a social network creates meaningful roles for people. This enhances self-esteem and gives people a purpose in life (Holt-Lunstad et al., 2010). Certain social contacts, such as friends, can lower the threshold to seek help when learning to navigate into society. Immigrants can for example benefit from asking others about institutional practises that exist in the host country (Holt-Lunstad et al., 2010).

Furthermore, when done in a group context, sports can stimulate immigrants to become familiar with the mainstream culture. Engaging in leisure activities such as sports can facilitate the process of 'fitting in' to mainstream society, which encapsulates the dominant culture, attitudes and practises (Doherty & Taylor, 2007; Smith et al., 2018). Stodolska and Alexandris (2004) also argue that doing sports is beneficial for bringing immigrants in contact with the mainstream culture. For example, immigrants face cultural practises that are embedded into the culture of playing sports. Seeing what clothes team members wear, being involved in small talk before a spinning class or learning about group dynamics within team sports are examples of nuances of mainstream culture that immigrants could learn through practising sports in a group context. Previous work also finds that developing a skill or comprehension of the culture of the host society is crucial for fitting into the dominant culture (Doherty & Taylor, 2007; Stodolska & Alexandris, 2004).

Moreover, sports can improve one's process of learning a new language or improving a language skill. Experiencing difficulties with the language of the host country can lead people to feel isolated and anxious as it becomes a threshold to communicate with others (Doherty & Taylor, 2007). People need to be able to express their feelings, opinions and ideas to effectively communicate with others in society. The lack of language proficiency can therefore lead to difficulties 'fitting in' in areas like work and school. Having the ability to speak the country's (main) language can enhance participation in social settings and therefore add to the process of integrating into society (McGinnity & Gijsberts, 2017). When doing sports in a group context, communication is essential and therefore learning and improving a language skill is stimulated (Doherty & Taylor, 2007). Immigrants get to learn new vocabulary and improve their pronunciation.

Lastly, sports can inhibit prejudice and discrimination against social groups (Morela et al., 2013). Doing sports in a group context can teach people about cultural diversity, as argued by Elling & Claringbould (2005). Creating mutual understanding between immigrants and natives can reduce stereotypes and discrimination (Morela et al., 2013).

Based on the existing literature about a possible relationship between sporting and social integration, the following hypothesis has been formulated:

Hypothesis 1: Sporting has a positive effect on the social integration of immigrants.

2.2 The role of gender in sporting and integrating into society

Previous literature has argued that men and women differ in their approach to social exclusion. Social exclusion contains the mechanisms that detach groups of people from the social mainstream (Doherty & Taylor, 2007). Women report higher feelings of social exclusion than men, therefore leading them to more often stop participating in sports. Sports have traditionally been associated with masculinity, which causes stereotypes to involve women not being capable or less interested in sports. It could lead to environments that do not encourage women to do sports (Doherty & Taylor, 2007). On the whole, women are also less likely to participate in mixed-gender groups due to their perceptions of social exclusion (Doherty & Taylor, 2007). Both men and women from the immigrant community experience social exclusion due to language barriers, prejudice and insufficient knowledge about practises, but women have more difficulties in overcoming this in comparison to men (Doherty & Taylor, 2007).

Also, differences due to religious or cultural traditions could explain the division in attitudes between men and women. For instance, as studied by De Knop et al. (1996) Muslim women can feel constrained in their behaviour due to the assignment of gender roles that restrict their access to sports. Dress codes for women, perspectives on the body about modesty and sporting mixed-gender settings could for example constrain women in their sports participation (Doherty & Taylor, 2007).

Furthermore, men and women differ in their needs and preferences in terms of forming social connections. Men tend to create friendships more easily over shared activities like doing sports in comparison to women (Hall, 2011). As doing sports can enhance building and strengthening social relations that facilitate social integration, men tend to have an advantage when doing sports compared to women. Hall (2011) argues that women expect and validate intimacy within a friendship more than men do. Besides that, males' friendships are more activity-focused than females' friendships. On the whole, sports could therefore be more beneficial for men to enhance their integration into a host society in comparison to women.

Based on this literature, the following hypothesis is formulated regarding the difference in the positive effect of sporting on social integration for men and women:

Hypothesis 2: The positive effect of sporting on the social integration of immigrants is larger for men than for women.

2.3 Control variables

Three control variables are added to the research model, namely educational background, group of origin and primary occupation. These things can influence the process of becoming socially integrated into a host society, irrespective of whether or not someone practises sports. To control for the influence of these things, all three variables have been added to the research model as control variables.

Firstly, the effect of the educational background of immigrants is being controlled. People who have obtained a higher level of education have on average more skills and resources that favour their integration into a new society (Fong, 2023). Having more years to study could increase the chances of improving a foreign language, for example English for many people. On the whole, higher-educated people are more exposed to other cultures, values and norms and could therefore have an advantage in adapting to a new environment in comparison to lower-educated people (Fong, 2023). Higher-educated immigrants also have more contact with the native population in the host society in comparison to lower-educated immigrants, which positively contributes to their social integration process (Martinović et al., 2008).

Furthermore, the effect of an immigration background on the social integration of immigrants is taken into account. People who have an immigration background on the whole have fewer resources, skills and cultural knowledge about the host society than people without such a background (Smith et al., 2018). Furthermore, Doherty and Taylor (2007) argue that children from immigrant parents could face the struggle of their parents trying to adjust to a new family life and work situation. The stress that comes with it can negatively influence the parents' ability to support their children's social integration process. Besides that, second-generation immigrants could have built more cultural capital than first-generation immigrants, as they possibly have more connections with the host society through various institutions (Smith et al., 2018).

Lastly, the effect of immigrants' primary occupation on their social integration process is being controlled. People who are employed or who engage in volunteer work can more easily connect with others in the broader society compared to those who are unemployed or do not participate in volunteer activities (Grimm et al., 2023; Sveen et al., 2022). Working or doing volunteer work can facilitate social interaction through different contexts and therefore enhance the creation or maintenance of social connections (Teije et al., 2013). Regarding economic integration, people who are more engaged in the workforce have a higher participation in community life (Putnam, 2000). Through work life, immigrants can connect with different people and possibly get to practise the local language more. Therefore, being employed or doing volunteer work can positively impact the social integration of immigrants.

3. Methods

This chapter discusses how the study has been conducted. A description of the dataset is given with information on the selection of respondents and non-response. Furthermore, the operationalisations of the variables are explained in detail and the analysis design is shown. References are made to Appendix 1 where more information on the operationalisations of the variables are given.

3.1 Dataset

The data that has been used in this research is derived from the panel of the Longitudinal Internet Studies for the Social Sciences (LISS) (LISS panel, n.d.-a). The LISS panel was set up in 2007 and is part of the project Measurement and Experimentation in the Social Sciences (MESS) (LISS panel, n.d.-b). The data that is being collected through the LISS panel is used for scientific research and policy design. It aims to gather representative information about the Dutch population. The topics that are included in the LISS panel questionnaires belong to the social, economic and behavioural sciences.

The LISS Core Study is a longitudinal study that measures changes in the lives of panel members (LISS panel, n.d.-c). It collects data on eight different topics, out of which the 'Social Integration and Leisure' data will be used for this research. Apart from the regular LISS panel, an Immigrant panel was set up from October 2010 until December 2014 (LISS panel, n.d.-d). For this research, the data from the first wave of the LISS Immigrants panel on 'Social Integration and Leisure' will be analysed.

LISS panel members fill in monthly questionnaires for which they receive financial compensation. Completing the questionnaires takes about an hour per month on average (LISS panel, n.d.-a). The sample that the LISS Immigrant panel uses is focused on the household level. This design allows more than one person from a household to fill in the questionnaires.

3.1 Respondents

The dataset Wave 1 is the starting point for this research and this data was collected in March 2011. The respondents to the survey are members of the LISS Immigrants panel who are at least 16 years old. 2557 household members have been approached to fill in the survey. This number can be divided into 1822 (71.3%) people who have completed the survey, 13 (0.5%) people who did not complete the survey and 722 (28.2%) people who have not filled in the questionnaire. Due to the presence of different migration backgrounds, the research has made efforts to include representative group sizes of respondents with different migration backgrounds. Thus, the total immigrant population in the Netherlands is reflected in the data. People of Turkish-Dutch origin and other non-Western origins are

slightly underrepresented in the panel. The group of Antillean-Dutch origin is a bit overrepresented due to the attempt to achieve a specific absolute number of respondents within this group (LISS panel, n.d.-e).

The LISS panel dataset includes a random sample that has been drawn from population registers (LISS panel, n.d.-f). People needed to be invited by the LISS panel to be able to participate in the study. Furthermore, the respondents that were selected have been approached by receiving a letter, followed by a telephone call and/or a house visit. Around 28% of the list of invited people for the research have not responded to the invitation. A large part of the non-response is caused by some groups being difficult to reach.

The sample size that is used in this research is smaller than the original dataset as certain respondents did not fill in all the relevant questions. After selecting the relevant questions for this research, combining the chosen variables into the final model and leaving out respondents who show missing values on one or more questions, 1232 respondents were left to incorporate in the final analysis. The decrease in sample size from 2227 to 1232 respondents is mainly because many respondents did not answer the question regarding whether or not they do sports.

3.2 Research design and procedure

The data was collected through an online structured survey with various pathways depending on the answers of respondents. The first part of the survey consists of questions relating to social relations and the leisure time of respondents. This includes questions about satisfaction with the current leisure time of respondents, how they spend their leisure time and whether or not they participate in certain activities. Subsequently, respondents are asked about their social contacts. These include statements about their feelings towards others and how often they participate in given activities that are related to socialising. The survey ends with a few short questions about how the respondent experienced filling in the questionnaire. To enhance the response rate and minimise selection bias, people who were invited and who did not have the means to complete the survey were given a computer and internet connection (LISS panel, n.d.-5).

3.3 Operationalisations

To test the hypotheses mentioned earlier, a range of questions from the questionnaire has been selected and used to define the concepts in the research model. The operationalisation of all the variables is explained in this section. An extensive explanation of the operationalisations is listed in Appendix 1.

The independent variable is *sporting*. This variable measures the amount of hours per week spent on doing sports. Respondents could answer the question 'Do you practise sports?' with 1 (*yes*) or 2 (*no*). All respondents who answered this question with 1 (*yes*) were given the question 'How many hours do you spend on sports per week, on average?'. This question could be answered by filling in a discrete amount of hours spent on sport per week. A small recoding has been done to change the answer possibilities of the first question to 0 (*no*) and 1 (*yes*). Furthermore, the respondents who answered 0 (*no*) on the first question were given the answer 0 for the second question, implying that they engage in sports for 0 hours per week.

The dependent variable is *social integration of immigrants*. The concept of social integration consists of many aspects that are listed in the literature and cannot be measured using one question from the dataset. To measure this concept as best as possible, the variable is computed using six statements about respondents' feelings about the community and their position among others. It is arguably the best option to operationalise this variable, as social integration is defined as a feeling of 'fitting' into society and having social contact with the broader society (Ager & Strang, 2008; Doherty & Taylor, 2007). Despite the limitations of questions in the dataset, *social integration of immigrants* is operationalised using the following six statements: 'I have a sense of emptiness around me', 'there are enough people I can count on in case of a misfortune', 'I know a lot of people that I can fully rely on', 'there are enough people to whom I feel closely connected', 'I miss having people around me' and 'I often feel deserted' (Cronbach's alpha = 0.82). All six statements have answer possibilities 1 (*yes*), 2 (*more or less*) and 3 (*no*). A high score on the variable *social integration of immigrants* is associated with positive feelings of being socially integrated. The second, third and fourth questions from the list mentioned earlier have high scores that imply not being well integrated into society. Therefore, the scale of these three questions was mirrored in order to aggregate the six items from the dataset into the variable *social integration of immigrants*.

The interaction variable in the model is *gender*. This variable measures the category that fits the gender of a respondent. Respondents could answer this question by selecting 1 (*male*) or 2 (*female*). To make the interpretation of the final results easier, the answers possibilities are recoded to 0 (*male*) and 1 (*female*).

The first control variable is *education*. This variable measures the highest level of education that a respondent obtained with a diploma. Respondents were able to choose between nine categories: 1 (*primary school*), 2 (*vmbo*), 3 (*havo/vwo*), 4 (*mbo*), 5 (*hbo*), 6 (*wo*), 7 (*other*), 8 (*not (yet) completed any study*) and 9 (*not yet started any education*). Answer category 7 is unspecified and cannot be placed on an ordinal scale between the other answer possibilities. Therefore, this answer is

recoded into a missing value. Answer possibilities 8 and 9 formally indicate that a respondent has not obtained any degree. These situations are ranked below other educational backgrounds and were recoded into the value 0.

The second control variable is *group of origin*. This variable measures the category that fits the respondents' migration background, including the division between Western and non-Western countries. Respondents fit into one of the following categories: 0 (*Dutch background*), 101 (*first generation foreign, Western background*), 102 (*first generation foreign, non-western background*), 201 (*second generation foreign, Western background*), 202 (*second generation foreign, non-western background*) and 999 (*origin unknown or part of the information unknown (missing values*)). These categories are defined by Statistics Netherlands (Statistics Netherlands, n.d.-b). According to the literature, second-generation immigrants have on average built up more resources, skills and knowledge that favour their process of becoming socially integrated in comparison to first-generation immigrants (Smith et al., 2018). To control for the effect of being a first or second-generation immigrant, the categories are divided into two groups and recoded as follows: 0 (*Dutch background, second generation foreign, Western background and second generation foreign, non-western background and) and 1 (first generation foreign, Western background and first generation foreign, non-western background).* The last category 999 is classified as a missing value, as the category is too vague to incorporate in the variable *group of origin*.

The third control variable is *primary occupation*. This variable measures the category that fits the occupation status of a respondent. The original variable divides primary occupations into 14 categories, namely 1 (*paid employment*), 2 (*works or assists in family business*), 3 (*autonomous professional, freelancer, or self-employed*), 4 (*job seeker following job loss*), 5 (*first-time job seeker*), 6 (*exempted from job seeking following job loss*), 7 (*attends school or is studying*), 8 (*takes care of the housekeeping*), 9 (*is pensioner*), 10 (*has (partial) work disability*), 11 (*performs unpaid work while retaining unemployment benefit*), 12 (*performs voluntary work*), 13 (*does something else) and 14 (is too young to have an occupation*). Earlier academic research suggests that being employed or doing volunteer work creates opportunities to build social relationships and take part in the pursuit of common collective goals (Grimm et al., 2023; Sveen et al., 2022). Having such an occupational status could therefore enhance the process of becoming socially integrated. Based on this assumption, respondents' primary occupations were categorised into groups of occupations that do and do not facilitate socializing. Answer options 4, 5, 6, 8, 9, 10, 13, and 14 were recoded into the new value 0 (occupation does not facilitate socializing), while answer options 1, 2, 3, 7, 11, and 12 were recoded into the new value 1 (occupation facilitates socializing).

3.4 Analysis design

In order to test the hypotheses formulated earlier, a hierarchical linear regression analysis has been done. Before executing the final regression analysis, the univariate and bivariate statistics of all variables involved are inspected.

The univariate statistics are shown in Appendix 1 and contain information on the distribution of data within all variables. Some respondents showed missing values on one or more variables that are included in this research. For the final analysis, these respondents have been removed from the sample size. The details on this are shown in Appendix 1.

The bivariate statistics are separated into investigating associations between categorical, continuous and dummy variables. The Pearson correlation is used to measure the association between continuous and dummy variables. To measure the association between categorical variables, the Cramer's V statistic is used. Lastly, the ANOVA test is used to measure the association between categorical and continuous variables. More details on the bivariate statistics are shown in Appendix 2.

Furthermore, the four assumptions of linear regression analysis have been checked: independent samples, linearity, homoscedasticity, and normality. Multicollinearity has also been examined using VIF scores. To check for potential outliers, the leverage, Cook's distance, standardised residuals and DDFIT have been examined. The details of this procedure can be found in Appendix 3.

The hierarchical regression analysis contains 4 models. The first model consists of the dependent variable *social integration of immigrants* and the control variables *education, group of origin and primary occupation*. The independent variable *sporting* will be added to the second model. The moderator *gender* will be added to the third model and the fourth and last model will include the interaction term *gender*sporting*.

4. Results

To test the previously mentioned hypotheses, a hierarchical linear regression was conducted. First, the univariate and bivariate statistics will be discussed to examine the distribution of the data and the correlations between the variables. Following this, the regression analysis and the final model are discussed to evaluate model fit. The role of outliers, multicollinearity and the linear regression assumptions are taken into account and are explained briefly. The final section discusses the hypotheses using the outcomes of the linear regression.

4.1 Descriptive statistics

4.1.1. Univariate statistics

Table 1 shows the univariate descriptives of all the variables that are included in the research. An interesting finding is that the independent variable *sporting* has a very right-skewed distribution. The mean of *2.17* is smaller than the standard deviation of *3.15* and the mean lies very close to the minimum value of *0* of this variable. Besides that, the variable *social integration of immigrants* has a very left-skewed distribution. The mean of *9.64* is close to the maximum value of *12* of this variable. This variable also has a small standard deviation of *2.66* in comparison to its mean. The fact that many respondents have a high score on this variable indicates that many respondents feel that they are well-integrated into society. The gender ratio is fairly balanced with 53.6% of all respondents being female. Furthermore, the variable *education* has an unequal distribution of respondents across all the groups. Most respondents reported having earned either a MBO or HBO diploma or an equivalent to this. Only a small number indicated that they had not pursued any education or had only completed primary school. When it comes to the variable *group of origin*, two-thirds of all respondents reported having a Dutch background or being a second-generation immigrant. Lastly, according to the statistics for the variable *primary occupation*, two out of three respondents fit the category of having an occupation that facilitates socialising. Therefore, most respondents report being employed or doing volunteer work.

Table 1: Univariate statistics of all the variables

Variable		Mean (standard deviation)	Minimum	Maximum	N
Sporting (scale 0-27)		2.17 (3.15)	0.00	27.00	1232
Social integration immigrants (scale 0-12)		9.64 (2.66)	0.00	12.00	1232
Gender					1232
	Men	46.4%			
	Women	53.6%			
Education					1232
	Not (yet) completed/started any education	3.3%			
	Primary school	5.6%			
	Secundary school	16.6%			
	Havo/vwo	12.9%			
	Mbo	21.7%			
	Hbo	24.8%			
	Wo	15.0%			
Group of origin					1232
	Dutch background/second generation immigrant	65.5%			
	First generation immigrant	34.5%			
Primary occupation					1232
	Occuaption does not facilitate socializing	34.6%			
	Occupation facilitates socializing	65.4%			

**For categorical variables, the frequencies are given as a percentage of the total*

4.1.2. Bivariate statistics

The bivariate statistics for the variables in this study are presented in Table 2. A large part of all correlations are significant, even at the p<0.01 level. A few interesting correlations will be discussed further, as they are relevant for testing the hypotheses in this research.

To begin with, there appears to be a strong association between *sporting* and *social integration of immigrants* (*r=0.12, p<0.01*). This finding is likely to significantly impact the results of testing the first hypothesis. Furthermore, there seems to be an association between *gender* and *sporting* (*r=0.06, p<0.05*). It could indicate a difference between the amount of sports men and women do per week. No significant difference is found between men and women and their feelings of being socially integrated (*r=0.03, p>0.05*).

The control variables also show some interesting findings in relation to the dependent variable. All three control variables show significant correlations with the variable *social integration of immigrants*. The first control variable *education* has a significant correlation with *social integration of immigrants* (*r=0.36, p<0.05*). This could indicate that immigrants with different educational backgrounds significantly differ in their feelings of being socially integrated. The variable *group of origin* shows a strong correlation with *social integration of immigrants* too (*r=0.19,p<0.01*). This strong correlation indicates that significant differences exist between immigrants with a Dutch background or second-generation immigrants and those who are first-generation immigrants in their feelings of being socially integrated. The last control variable *primary occupation* also shows a strong and significant correlation with *social integration of immigrants* (*r=0.11, p<0.01*). This finding suggests that there is a difference between immigrants who do and do not have a primary occupation that facilitates socializing and their feelings of being socially integrated. On the whole, immigrants who are employed or do volunteer work differ from those who do not in terms of their feelings about being socially integrated.

Lastly, all correlations between the control variables are significant. *Education* has a strong correlation with both *group of origin* and *primary occupation* (respectively *r=0.14, p<0.01 and r=0.22, p<0.01*). Besides that, *group of origin* and *primary occupation* also show a significant correlation (*r=0.09, p<0.01*). This indicates that there are differences between both groups with different immigration backgrounds and their educational attainment. Furthermore, immigrants who do and do not have a primary occupation that facilitates socializing also differ in their educational background. Lastly, both groups with different immigration backgrounds seem to differ in their current primary occupation status.

Table 2: Bivariate statistics of all the variables (N=1232)

	Sporting	Social integration immigrants	Gender	Education	Group of origin	Primary occupation
Sporting	-	-	-	-	-	-
Social integration immigrants	0.12**[a]	-	-	-	-	-
Gender (0= male; 1= female)	0.06*[a]	0.03[a]	-	-	-	-
Education	0.08[c]	0.36*[c]	0.09[b]	-	-	-
Group or origin (0 =Dutch background/second generation immigrant, 1 = first generation immigrant)	0.05[a]	0.19**[a]	0.00[b]	0.14**[b]	-	-
Primary occupation (0= facilitates socializing; 1= does not facilitate socializing	0.09**[a]	0.11**[a]	0.06*[b]	0.22**[b]	0.09**[b]	-

[a]Pearson correlation, [b]Cramer's V, [c]based on ANOVA R^2

**Significant at p<0.05, **significant at p<0.01*

4.2. Model evaluation

A hierarchical linear regression has been done and the fit of all four models will be examined in this section. A reference is made to figure 3 which contains the details of the regression. Further on, the assumptions for linear regression analysis, multicollinearity and outliers are discussed.

4.2.1. Model fit

To evaluate the model fit, we take a look at the amount of variance in immigrants' feelings of being socially integrated which is explained using the variables that are added to the model. It is important

to investigate whether adding the explanatory variable, the moderator and the interaction term results in more explained variance in *social integration of immigrants*.

The first model explains a small but significant part of the variance in *social integration of immigrants* (R^2*=0.06;* $R^2_{adjusted}$*=0.05; F=23,89*). This model contains the control variables *education*, *group of* origin and *primary occupation* all show a significant amount of explained variance in *social integration of immigrants*.

Sporting has been added to the second model and shows that it explains a significant extra amount of variance in *social integration of immigrants* (R^2*=0.07;* $R^2_{adjusted}$*=0.06; F=14.61*). Together with the control variables, *sporting* has a significant parameter (*b=0.09; p<0.01*). This indicates that the amount of hours that immigrants spend on sporting during the week has a significant impact on their feelings of being socially integrated.

The moderator *gender* has been added to the third model and does not explain a significant extra amount of variance in *social integration of immigrants* (R^2*=0.06;* $R^2_{adjusted}$*=0.06; F=2,34*). The parameter of *gender* does therefore not show a significant influence on predicting immigrants' feelings of being socially integrated (*b=0.23; p>0.05*).

The fourth and final model contains all the variables, together with the interaction term *gender*sporting*. This model does not explain a significantly bigger amount of variance in *social integration of immigrants* compared to the previous model (R^2*=0.07;* $R^2_{adjusted}$*=0.06; F=1.20*). Both the moderator and interaction term do not show a significantly better prediction of *social integration of immigrants* (respectively *b=0.23, p>0.05* and *b=0.05, p>0.05*). On the whole, the addition of a new variable to every new model has not changed any parameters from previous models.

Table 3: Results of a hierarchical regression analysis with social integration of immigrants as the dependent variable, sporting as the independent variable, gender as the moderator and control variables.

**Significant at p<0.05, **significant at p<0.01*

Regression	Model 1	Model 2	Model 3	Model 4
	B (SE)	B (SE)	B (SE)	B (SE)
Intercept	9.14**	8.98**	8.84**	8.84**
	(0.21)	(0.21)	(0.23)	(0.23)
Education	0.16**	0.16**	0.16**	0.16**
	(0.05)	(0.05)	(0.05)	(0.05)
Group origin (0: Dutch background/second generation immigrant; 1: first generation immigrant)	-1.08**	-1.06**	-1.06**	-1.06**
	(0.16)	(0.16)	(0.16)	(0.16)
Primary occupation (0: occupation does not facilitate socializing; 1: occupation facilitates socializing)	0.40*	0.35*	0.36*	0.37*
	(0.16)	(0.16)	(0.16)	(0.16)
Sporting		0.09**	0.09**	0.10**
		(0.02)	0.02)	(0.02)
Gender			0.23	0.23
(0: male; 1: female)			(0.15)	(0.15)
Gender*Sporting				0.05
				(0.05)
R^2	0.06	0.07	0.07	0.07
R^2 adjusted	0.05	0.06	0.06	0.06
F change	23.89**	14.61**	2.34	1.20
N	1232	1232	1232	1232

4.2.2. Model assumptions, multicollinearity and outliers

Firstly, the assumptions of linear regression analysis will be discussed in this section. After that, possible multicollinearity and outliers in the data will be examined.

The four assumptions of linear regression analysis will be examined briefly. For a more detailed look, a reference is made to Appendix 3. The first assumption is about independent samples. As mentioned earlier, the LISS Immigrant panel is used to gather the data of immigrants in the Netherlands. This data has been gathered at the household level. Several people from a single household were able to participate in the study. People from the same household can influence one another, which violates the assumption of independent samples. Moreover, the dependent variable *social integration of immigrants* has been regarded as a continuous variable, although it contains a limited amount of score possibilities. When having a look at the univariate statistics of the variables, it is noticeable that the data in some variables do not show a normal distribution. Therefore, the assumptions of normality, linearity and homoscedasticity are slightly violated and have influenced the outcomes of this research to a certain extent.

Multicollinearity has been checked, as some variables might correlate strongly with each other. The scores on the VIF test range from *1.009* to *1.051*, which is below the criteria of VIF < 4. Therefore, no multicollinearity has been found among the independent variables. In other words, the independent variables are not strongly correlated and can therefore predict the dependent variable effectively.

4.3. Hypothesis testing

Two hypotheses have been formulated based on the research question '*What is the effect of sporting on the social integration of immigrants and what role does gender play?'.* These will be discussed using the relevant models from the regression analysis.

The first hypothesis is formulated as follows: *Sporting has a positive effect on the social integration of immigrants.* The effect of sporting on the social integration of immigrants is shown in model 2. According to the statistics, *sporting* has a significant positive effect on predicting *social integration of immigrants* (*b=0.009, p<0.01*). Every extra hour that an immigrant practises sports during the week increases the score of feeling socially integrated by *0.09*. This indicates that the more hours immigrants spend on sports each week, the more they report feeling socially integrated. This is in accordance with the formulated hypothesis. Therefore, evidence has been found in this research that supports the positive effect of sporting on the social integration of immigrants.

The second hypothesis is stated as follows: *The positive effect of sporting on the social integration of immigrants is larger for men than for women.* The effect of *gender* and *gender*sporting* on *social integration of immigrants* is shown in models 3 and 4. As seen in model 3, *gender* does not have a significant effect on predicting *social integration of immigrants* (*b=0.23,*

p>0.05). This indicates that there is no significant difference between men and women reporting their feelings of being socially integrated. Furthermore, model 4 shows the interaction term *sporting*gender*, which does not have a significant effect on *social integration of immigrants* (*b=0.05,p>0.05*). This finding suggests that there is no significant difference in the positive effect of sporting on social integration of men and women in the immigrant community. In other words, no evidence has been found that supports the second hypothesis.

Lastly, the control variables will be discussed. When having a look at model 1, it becomes clear that all three control variables have a significant influence on predicting social integration of immigrants. *Education* has a significant positive effect on *social integration of immigrants* (*b=0.16, p<0.01*). This indicates that, on average, an increase in the category of education background will lead to a *0.16* higher score on reporting feeling socially integrated. *Group of origin* has a significant negative effect on predicting *social integration of immigrants* (*b=-1.08, p<0.01*). This finding suggests that immigrants with a Dutch background or who are second-generation immigrants on average score *1.08* higher than first-generation immigrants on reporting feeling socially integrated. The third and final control variable *primary occupation* has a significant positive effect on predicting *social integration of immigrants* (*b=-0.40, p<0.05*). This indicates that immigrants with occupations that facilitate socializing on average score *0.40* higher on reporting feelings of being socially integrated in comparison to immigrants without such occupations. The statistics and significance of control variables barely change when adding the independent variable, moderator and interaction term to the other models.

5. Conclusion

This research focuses on the role of sporting on the social integration of men and women in the immigrant community. Through a hierarchical linear regression analysis based on data from the LISS Immigrant panel, the following research question was addressed: *What is the effect of sporting on the social integration of immigrants and what role does gender play?* In line with existing literature, two hypotheses have been formulated and tested. The effect of sporting on immigrants' social integration was examined, along with the effect of gender on the relationship between sporting and immigrants' social integration. The conclusions on the hypotheses will be discussed in this section.

The first hypothesis is formulated as follows: *'Sporting has a positive effect on the social integration of immigrants'.* The linear regression analysis has found evidence that supports this hypothesis. An increase in the number of hours spent on doing sports per week leads to a significant increase in immigrants reporting feelings of being socially integrated. This finding supports the arguments that are made explicit in the theoretical framework. It is previously argued that doing sports is beneficial for both cognitive and physical health (Bailey, 2016; Doherty & Taylor, 2007; Ruchti et al., 2013). It therefore impacts one's ability to mobilize and to be active in diverse aspects of everyday life. Doing sports in a group context offers opportunities for immigrants to build and maintain social relationships (Mollenhorst et al., 2014; Doherty & Taylor, 2007; Pringle & Liu, 2023). From interacting with others during sports, immigrants face cultural norms and attitudes and learn about cultural practises and the mainstream culture (Smith et al., 2018; Stodolska & Alexandris, 2004). Besides that, immigrants can improve their language skills to further enhance their opportunities to communicate with others in society (McGinnity & Gijsberts, 2017; Doherty & Taylor, 2007).

The second hypothesis contains the moderation effect: *'The positive effect of sporting on the social integration of immigrants is larger for men than for women'.* The analysis in this research found no significant effect of gender on the social integration of immigrants. No evidence was found that supports this hypothesis, as the effect of sporting on the social integration of immigrants does not seem to differ between men and women. This finding is not in line with the arguments that are made explicit in the theoretical framework. Here, it was discussed that men and women differ in their needs in creating social relationships. Women have a greater need for intimacy in social relationships in comparison to men (Hall, 2011; Athenstaedt et al., 2008). Besides that, doing shared activities could be more beneficial for men compared to women in terms of creating new social relationships (Hall, 2011). Furthermore, it was argued that women and men both experience social exclusion in doing sports, but women experience more difficulties in overcoming this sense of social exclusion

compared to men (Doherty & Taylor, 2007). These arguments could have explained a possible difference in the effects of sporting on social integration of men and women in the immigrant community.

6. Discussion

Some aspects of this research are limiting the results of the analysis and overall research. For instance, these limitations include the operationalisations of variables, the items from the questionnaire that are used and the influences that were unable to take into account for this research.

To begin with, the variable *social integration of immigrants* is created using items from the questionnaire that do not fully reflect the true meaning of this concept. The items that are used include questions that measure feelings of loneliness and sense of belonging. However, due to the restrictions of the questionnaire, these questions are most suitable to use for defining the concept of social integration among immigrants.

Moreover, this research solely measures the number of hours per week that respondents spend doing sports and does not provide further details about this activity. No distinction is made in the type of sports, whether the activity is done individually or within a group context and in what context the activity is done. Specifying the kind of sports activity could have led to more concrete and detailed outcomes in this research.

Some of the variables in this research have violated the assumption of normality, which has influenced the results of the regression analysis. It is important to work with data that shows a normal distribution of respondents, as it impacts the outcomes of statistical analyses and limits the power of drawing conclusions about the population at large. For example, the variables *sporting* and *social integration of immigrants* have many respondents with either relatively high or low scores. These skewed distributions of data have therefore limited the analysis in this research. For future research, it is recommended to increase the sample size to help reduce the impact of non-normal distributions. Besides that, other research methods could be considered to mitigate the effect of non-normality in the data.

Lastly, the definition of social integration of immigrants has been approached from a somewhat narrow perspective. Social integration is not limited to only immigrants putting effort into adapting to the host society. It also involves how the local population interacts with immigrants. This aspect is not reflected in the operationalisations used in this research. Moreover, while it is desirable for social integration to occur from both sides (newcomers and locals), the expectation of the host society is often placed solely on newcomers to adapt as seen in previous research (Phillips, 2009).

Based on this research, some advice on policy design can be given to improve national policy. A positive effect was found between the amount of sports immigrants do every week and their

feelings of being socially integrated. A focus on the relevance of sports for the immigrant community has already been implemented in policy strategies. However, this could be enhanced by investing in national sports programs that lower the threshold for immigrants to engage in sports. Some immigrants might be unfamiliar with existing organisations that offer opportunities to become active in sports. It is therefore important to further inform immigrants on the possibilities regarding sports clubs and programs and to provide opportunities for them to do sports in an organised setting. This strategy will enhance the overall health of both the immigrant community and the society at large.

Moreover, the importance of immigrants becoming socially integrated is seen in their understanding of public health interventions and the structure of the economy of the host society (Pringle & Liu, 2023). Immigrants who are better integrated into society have a greater understanding of national policies and have more resources to navigate society. For example, immigrants can have a better understanding of governmental benefits, the structure of the Dutch healthcare system and the welfare programs offered by the government. As seen in the theoretical framework, immigrants can benefit from having a diverse set of social networks to increase their understanding of national policies and interventions (Pringle & Liu, 2023; Schinkel, 2013; Lee et al., 2019).

Literature

Ager, A., & Strang, A. (2008). Understanding Integration: a Conceptual framework. *Journal Of Refugee Studies*, *21*(2), 166–191. https://doi.org/10.1093/jrs/fen016

Athenstaedt, U., Mikula, G., & Bredt, C. (2008). Gender role Self-Concept and leisure activities of adolescents. *Sex Roles*, *60*(5–6), 399–409. https://doi.org/10.1007/s11199-008-9543-y

Bailey, R. (2016). Sport, physical activity and educational achievement – towards an explanatory model. *Sport in Society*, *20*(7), 768–788. https://doi.org/10.1080/17430437.2016.1207756

Barker-Ruchti, N., Barker, D., Sattler, S., Gerber, M., & Pühse, U. (2013). Sport—'It's Just Healthy': Locating Healthism within Discourses of Social Integration. *Journal Of Ethnic And Migration Studies*, *39*(5), 759–772. https://doi.org/10.1080/1369183x.2013.756674

Cherney, I. D., & London, K. (2006). Gender-linked differences in the toys, television shows, computer games, and outdoor activities of 5- to 13-year-old children. *Sex Roles*, *54*(9–10), 717–726. https://doi.org/10.1007/s11199-006-9037-8

De Knop, P., Theeboom, M., Wittock, H., & De Martelaer, K. (1996). Implications of Islam on Muslim Girls' Sport Participation in Western Europe. Literature Review and Policy Recommendations for Sport Promotion. *Sport Education And Society*, *1*(2), 147–164. https://doi.org/10.1080/1357332960010202

Doherty, A., & Taylor, T. (2007). Sport and physical recreation in the settlement of immigrant youth. *Leisure*, *31*(1), 27–55. https://doi.org/10.1080/14927713.2007.9651372

Elling, A., & Claringbould, I. (2005). Mechanisms of Inclusion and Exclusion in the Dutch Sports Landscape: Who Can and Wants to Belong? *Sociology Of Sport Journal*, *22*(4), 498–515. https://doi.org/10.1123/ssj.22.4.498

European Union. (n.d.). *EU enlargement*. Retrieved August 8, 2024, from https://european-union.europa.eu/principles-countries-history/eu-enlargement_en

Fong, E. (2023). Social and economic integration of highly educated/high-skilled migrants in East and Southeast Asia: Overview of the Special Issue. *Asian and Pacific Migration Journal, 32*(1), 3-9. https://doi.org/10.1177/01171968231172526

Grimm, N., Hense, A., & Vogel, B. (2023). How Employment Generates Social Integration: Trends Towards Disintegration and Over-Integration in the Hyper-Work Society. *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, *75*(S1), 331–356. https://doi.org/10.1007/s11577-023-00903-5

Grip, L. (2019). Knocking on the Doors of Integration: Swedish Integration Policy and the Production of a National Space. *Journal Of International Migration And Integration*, *21*(3), 861–877. https://doi.org/10.1007/s12134-019-00691-y

Hall, J. A. (2011). Sex differences in friendship expectations: A meta-analysis. *Journal Of Social And Personal Relationships*, *28*(6), 723–747. https://doi.org/10.1177/0265407510386192

Haslam, C., Dane, S., Lam, B. C. P., Jetten, J., Liu, S., Gallois, C., & Tran, T. L. N. (2020). Ageing well in a foreign land: group memberships protect older immigrants' wellbeing through enabling social support and integration. *Ageing And Society*, *42*(7), 1710–1732. https://doi.org/10.1017/s0144686x20001695

Hendriks, M., & Burger, M. J. (2019). Unsuccessful Subjective Well-Being Assimilation among Immigrants: The role of faltering perceptions of the host society. *Journal of Happiness Studies*, *21*(6), 1985–2006. https://doi.org/10.1007/s10902-019-00164-0

Holt, N. L., Neely, K. C., Slater, L. G., Camiré, M., Côté, J., Fraser-Thomas, J., MacDonald, D., Strachan, L., & Tamminen, K. A. (2016). A grounded theory of positive youth development through sport based on results from a qualitative meta-study. *International Review Of Sport And Exercise Psychology*, *10*(1), 1–49. https://doi.org/10.1080/1750984x.2016.1180704

Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social Relationships and Mortality Risk: A Meta-analytic Review. *PLoS Medicine*, *7*(7), 1-19. https://doi.org/10.1371/journal.pmed.1000316

Jennissen, R., Bovens, M., Engbersen, G., & Bokhorst, M. (2022). The Netherlands as a country of immigration. In *Research for policy* (pp. 17–41). https://doi.org/10.1007/978-3-031-14224-6_2

Kim, A. C. H., Park, S. H., Kim, S., & Fontes-Comber, A. (2019). Psychological and social outcomes of sport participation for older adults: a systematic review. *Ageing and Society*, *40*(7), 1529–1549. https://doi.org/10.1017/s0144686x19000175

Lee, J., Hong, J., Zhou, Y., & Robles, G. (2019). The Relationships Between Loneliness, Social Support, and Resilience Among Latinx Immigrants in the United States. *Clinical Social Work Journal*, *48*(1), 99–109. https://doi.org/10.1007/s10615-019-00728-w

Li, P. S. (2003). Deconstructing Canada's discourse of immigrant integration. *Journal Of International Migration And Integration*, *4*(3), 315–333. https://doi.org/10.1007/s12134-003-1024-0

LISS panel. (n.d.-a). About the panel. Retrieved August 9, 2024 from https://www.lissdata.nl/how-it-works

LISS panel. (n.d.-b). About us. Retrieved August 9, 2024 from https://www.lissdata.nl/about-us

LISS panel. (n.d-c). LISS Core Study. Retrieved August 9, 2024 from https://www.lissdata.nl/liss-core-Study

LISS panel. (n.d.-d). Immigrant panel. Retrieved August 9, 2024 from https://www.dataarchive.lissdata.nl/study-units/view/162

LISS panel. (n.d.-e). Methodology: Sampling and recruitment of the MESS Immigrant panel. Retrieved August 9, 2024 from https://www.lissdata.nl/methodology

LISS panel. (n.d.-f). Methodology. Retrieved August 9, 2024 from https://www.lissdata.nl/methodology

Martinović, B., Van Tubergen, F., & Maas, I. (2008). Dynamics of Interethnic Contact: A Panel Study of Immigrants in the Netherlands. European Sociological Review, 25(3), 303-318. https://doi.org/10.1093/esr/jcn049

McGinnity, F., & Gijsberts, M. (2017). The experience of discrimination among newly arrived Poles in Ireland and the Netherlands. *Ethnic and Racial Studies, 41*(5), 919-937. https://doi.org/10.1080/01419870.2017.1332376

Mollenhorst, G., Volker, B., & Flap, H. (2014). Changes in personal relationships: How social contexts affect the emergence and discontinuation of relationships. *Social Networks*, *37*, 65–80. https://doi.org/10.1016/j.socnet.2013.12.003

Morela, E., Hatzigeorgiadis, A., Kouli, O., Elbe, A., & Sanchez, X. (2013). Team cohesion and ethnic–cultural identity in adolescent migrant athletes. *International Journal Of Intercultural Relations*, *37*(5), 643–647. https://doi.org/10.1016/j.ijintrel.2013.05.001

Phillips, D. (2009). Minority Ethnic Segregation, Integration and Citizenship: A European Perspective. *Journal Of Ethnic And Migration Studies*, *36*(2), 209–225. https://doi.org/10.1080/13691830903387337

Pringle, R., & Liu, L. (2023). Mainland Chinese first-generation immigrants and New Zealanders' views on sport participation, race/ethnicity and the body: Does sport participation enhance cultural understandings? *International Review For The Sociology Of Sport*, *58*(4), 725–745. https://doi.org/10.1177/10126902231156278

Putnam, R. D. (2000). Bowling Alone: The collapse and revival of American community. New York: Simon & Schuster. 10.1145/358916.361990

Schinkel, W. (2013). The imagination of 'society' in measurements of immigrant integration. *Ethnic And Racial Studies*, *36*(7), 1142–1161. https://doi.org/10.1080/01419870.2013.783709

Shea, J. M., & Beausoleil, N. (2011). Breaking down 'healthism': barriers to health and fitness as identified by immigrant youth in St. John's, NL, Canada. *Sport Education And Society*, *17*(1), 97–112. https://doi.org/10.1080/13573322.2011.607914

Smith, R., Spaaij, R., & McDonald, B. (2018). Migrant Integration and Cultural Capital in the Context of Sport and Physical Activity: a Systematic Review. *Journal Of International Migration And*

Integration, *20*(3), 851–868. https://doi.org/10.1007/s12134-018-0634-5

Statistics Netherlands. (2022). Integratie en Samenleven 2022 [Integration and coexistence 2022]. Retrieved August 3, 2024 from https://longreads.cbs.nl/integratie-en-samenleven-2022/bevolking/

Statistics Netherlands, (n.d.-a). *How many people immigrate to the Netherlands?* Retrieved August 3, 2024 from https://www.cbs.nl/en-gb/dossier/asylum-migration-and-integration/how-many-people-immigrate-to-the-netherlands-

Statistics Netherlands. (n.d.-b). *Migration background*. Retrieved August 10, 2024 from https://www.cbs.nl/en-gb/our-services/methods/definitions/migration-background

Stodolska, M., & Alexandris, K. (2004). The Role of Recreational Sport in the Adaptation of First Generation Immigrants in the United States. *Journal Of Leisure Research*, *36*(3), 379–413. https://doi.org/10.1080/00222216.2004.11950029

Sveen, S., Anthun, K. S., Batt-Rawden, K. B., & Tingvold, L. (2022). Volunteering: A Tool for Social Inclusion and Promoting the Well-Being of Refugees? A Qualitative Study. *Societies*, *13*(1), 12. https://doi.org/10.3390/soc13010012

Teije, I. T, Coenders, M., & Verkuyten, M. (2013). The Paradox of Integration. Social Psychology, 44(4), 278-288. https://doi.org/10.1027/1864-9335/a000113

Tuturea, L. E., Hahn, E., Mavituna, S., Eillinghoff, L., Loan, T. T., Böge, K., & Ta, T. M. T. (2023). Cultural immersion, acculturation strategies, and depressive symptoms among first-generation Vietnamese migrants in Germany. *International Journal of Social Psychiatry*, *69*(8), 2048–2058. https://doi.org/10.1177/00207640231188036

Vermeulen, J., & Verweel, P. (2009). Participation in sport: bonding and bridging as identity work. *Sport in Society*, *12*(9), 1206–1219. https://doi.org/10.1080/17430430903137886

Videon, T. M. (2002). Who plays and who benefits: gender, interscholastic athletics, and academic

outcomes. *Sociological Perspectives*, *45*(4), 415–444. https://doi.org/10.1525/sop.2002.45.4.415

Waring, A., & Mason, C. (2010). Opening doors: promoting social inclusion through increased sports opportunities. *Sport in Society*, *13*(3), 517–529. https://doi.org/10.1080/17430431003588192

Appendix 1 - Operationalisations and univariate statistics

All operationalisations of the variables and their missing data statistics are displayed in this section, using SPSS syntax and output with brief explanations. The descriptive statistics are provided for each variable. A boxplot and bar chart is shown for all the continuous variables after recoding the original variables. For all the discrete variables, a bar chart and frequency table is given. Respondents with missing values on one or more variables were analysed in the missing values analysis and were excluded from the final linear regression.

1.1 Sporting

The original items from the dataset are shown below.

- fb14b097: Do you practise sports?

  1=yes

  2=no

- If fb14b097=1, then fb14b098: How many hours do you spend on sports per week, on average?

  Respondents could fill in a discrete number of hours

```
FREQUENCIES VARIABLES=fb14b097
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN
  /ORDER=ANALYSIS.

FREQUENCIES VARIABLES=fb14b098
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN
  /ORDER=ANALYSIS.

GRAPH
  /HISTOGRAM=fb14b098.
```

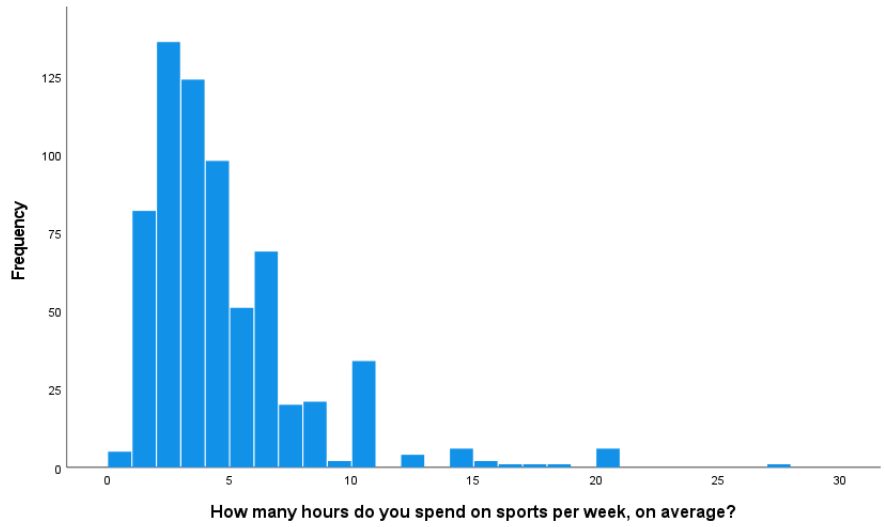
Do you practice sports?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	664	29,8	51,2	51,2
	no	634	28,5	48,8	100,0
	Total	1298	58,3	100,0	
Missing	System	929	41,7		
Total		2227	100,0		

Statistics

How many hours do you spend

N	Valid	664
	Missing	1563
Mean		4,26
Median		3,00
Std. Deviation		3,290
Minimum		0
Maximum		27



For the question 'How many hours do you spend on sports per week, on average?' the scores of the respondents vary from 0 to 27 with a mean score of 4.26 hours per week of sporting. The data does not show a normal distribution, as most respondents score low on this variable and few respondents report high scores.

All respondents who answered the question 'Do you practise sports?' with 2 (*no*) have been recoded into a score of 0 on the question 'How many hours do you spend on sports per week, on average?'. The system missing part of the second question consists of people who have not answered the first question and who answered the first question with 2 (*no*). To only work with respondents that answered the first question, the respondents that had a missing value on the first question were removed from the dataset. As a result, all respondents with a missing value on the second question must have answered the first question with 2 (*no*). Therefore, all respondents who indicated that they

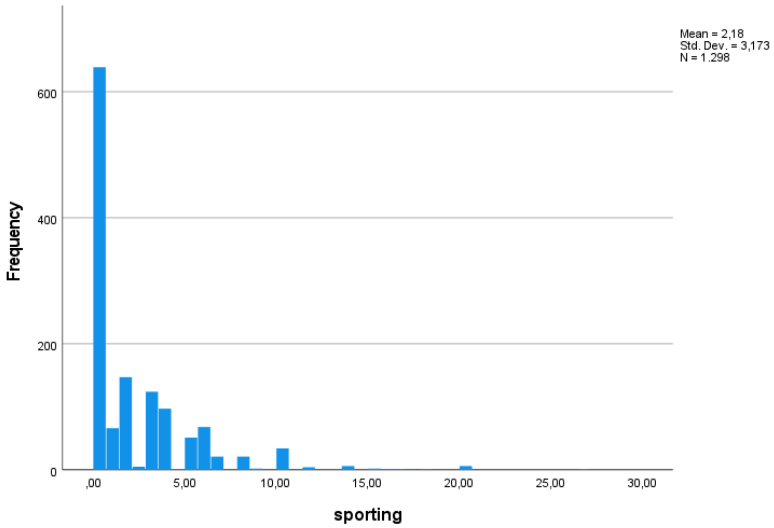
do not sport can be given the value 0 in the second question. 1298 respondents were left to compute the final variable *sporting*.

```
RECODE fb14b098 (SYSMIS=0) (0 thru 27=Copy) INTO sporting.
EXECUTE.
```

Statistics

sporting

N	Valid	1298
	Missing	0
Mean		2,1784
Median		1,0000
Std. Deviation		3,17316
Range		27,00
Minimum		,00
Maximum		27,00



The final *sporting* has a right-skewed distribution. Most respondents score relatively low on this variable and the distribution of the scores shows some gaps. There are a few respondents that have an extreme score on this variable and are located on the right tail of the distribution. Most respondents indicate that they do not sport. We can conclude that this variable violates the assumption of normality.

1.2 Social integration immigrants

The original items from the dataset are shown below.

- fb14b307: I have a sense of emptiness around me
  - 1= yes
  - 2= more or less
  - 3= no
- fb14b308: there are enough people I can count on in case of a misfortune
  - 1= yes
  - 2= more or less
  - 3 = no

- fb14b309: I know a lot of people that I can fully rely on

  1= yes

  2= more or less

  3= no

- fb14b310: there are enough people to whom I feel closely connected

  1= yes

  2= more or less

  3= no

- fb14b311: I miss having people around me

  1= yes

  2= more or less

  3= no

- fb14b312: I often feel deserted

  1= yes

  2= more or less

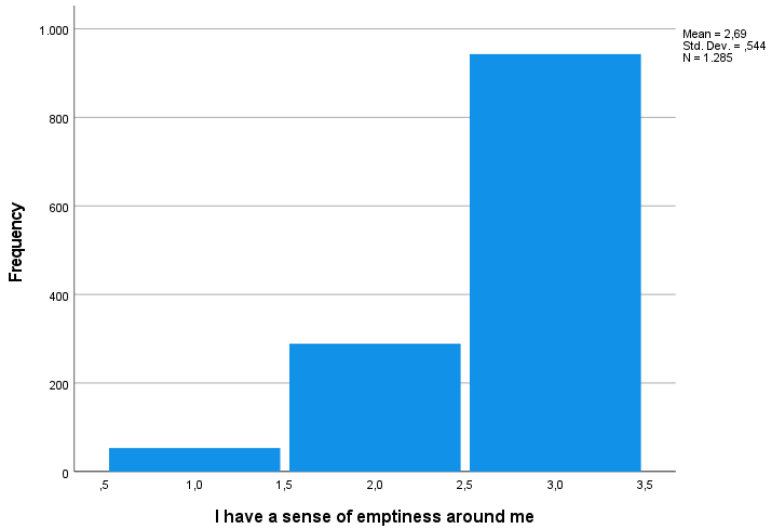
  3= no

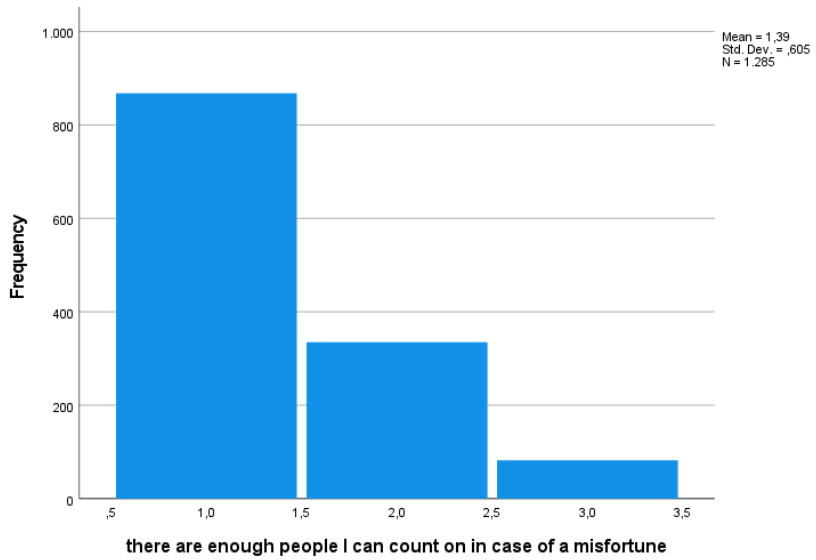
```
FREQUENCIES VARIABLES=fb14b307 fb14b308 fb14b309 fb14b310 fb14b311 fb14b312
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN
  /ORDER=ANALYSIS.

GRAPH
  /HISTOGRAM=fb14b307.

GRAPH
  /HISTOGRAM=fb14b308.

GRAPH
  /HISTOGRAM=fb14b309.

GRAPH
  /HISTOGRAM=fb14b310.

GRAPH
  /HISTOGRAM=fb14b311.
```

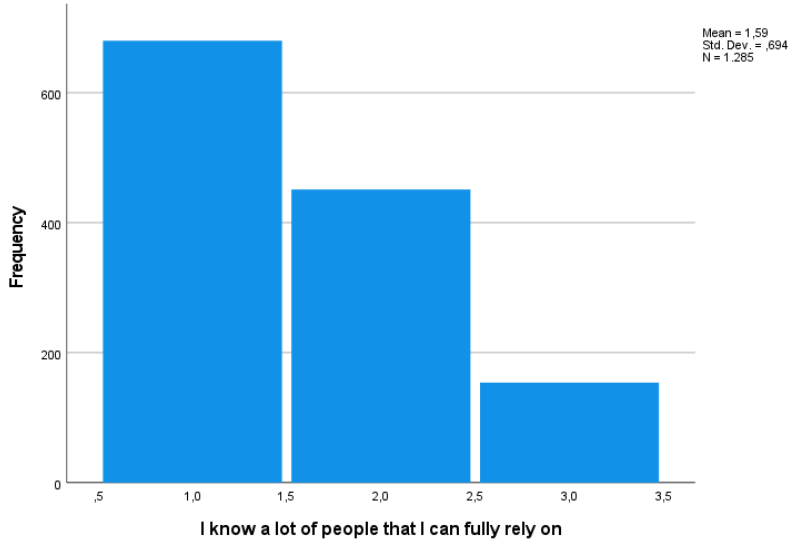
```
GRAPH
  /HISTOGRAM=fb14b312.

RECODE fb14b308 fb14b309 fb14b310 (1=3) (2=2) (3=1) INTO rfb14b308 rfb14b309 rfb14b310.
EXECUTE.

RELIABILITY
  /VARIABLES=rfb14b308 rfb14b309 rfb14b310 fb14b307 fb14b311 fb14b312
  /SCALE('ALL VARIABLES') ALL
  /MODEL=ALPHA
  /STATISTICS=DESCRIPTIVE SCALE CORR
  /SUMMARY=TOTAL.
```

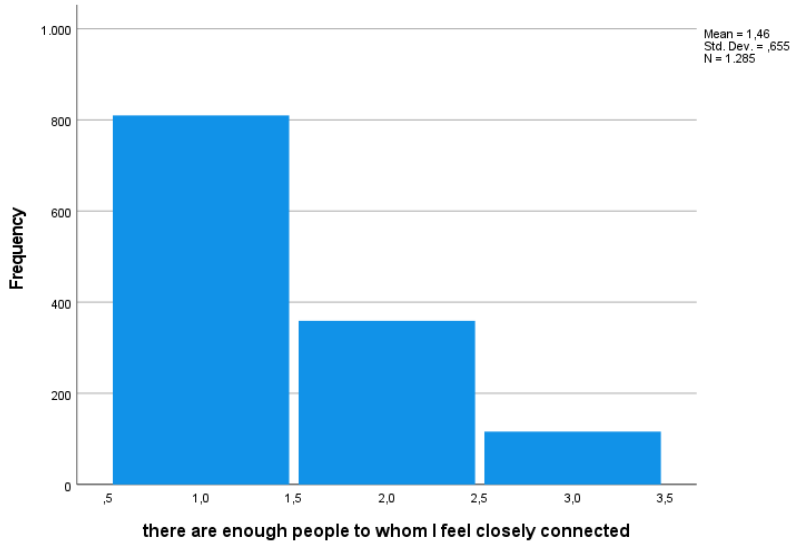
Statistics

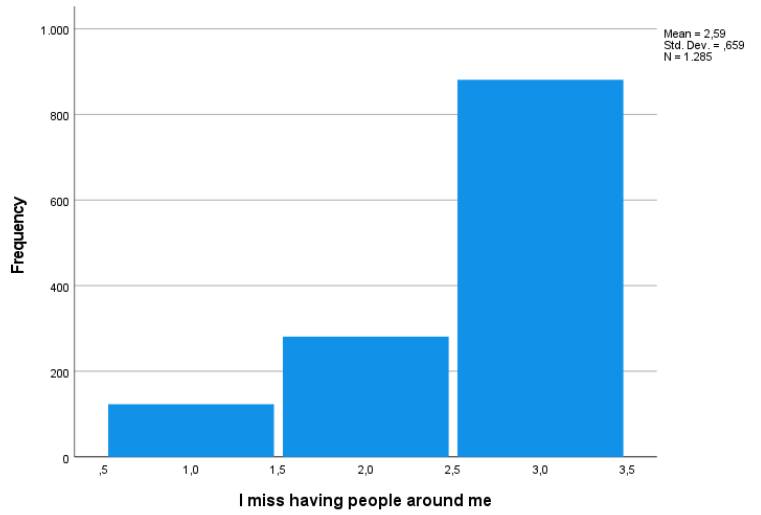
		I have a sense of emptiness around me	there are enough people I can count on in case of a misfortune	I know a lot of people that I can fully rely on	there are enough people to whom I feel closely connected	I miss having people around me	I often feel deserted
N	Valid	1285	1285	1285	1285	1285	1285
	Missing	13	13	13	13	13	13
Mean		2,69	1,39	1,59	1,46	2,59	2,78
Median		3,00	1,00	1,00	1,00	3,00	3,00
Std. Deviation		,544	,605	,694	,655	,659	,508
Minimum		1	1	1	1	1	1
Maximum		3	3	3	3	3	3

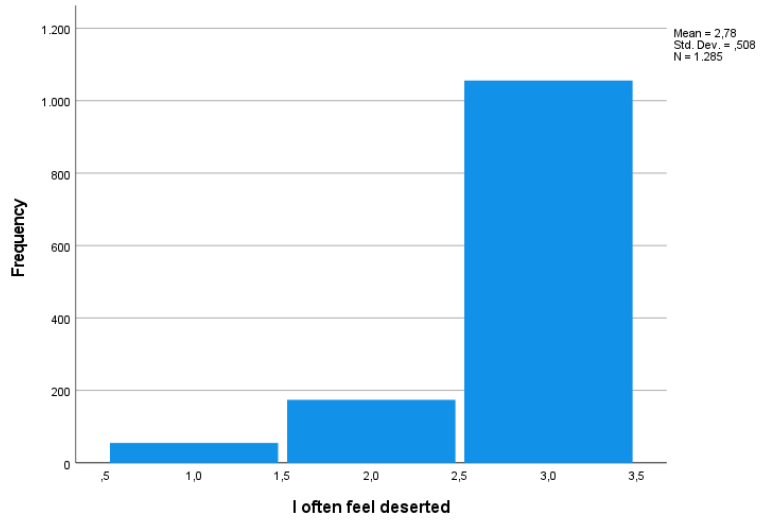





Mean = 1,59
Std. Dev. = ,694
N = 1.285
Frequency
600
400
200
0
,5
1,0
1,5
2,0
2,5
3,0
3,5
I know a lot of people that I can fully rely on


Mean = 1,46
Std. Dev. = ,655
N = 1.285
Frequency
1.000
800
600
400
200
0
,5
1,0
1,5
2,0
2,5
3,0
3,5
there are enough people to whom I feel closely connected





The univariate descriptives of the individual items are displayed above. The six statements all have 3 answer possibilities. A high score on some statements indicates a feeling that is associated with a low level of social integration, whereas for other statements a high score indicates a high level of social integration. On the whole, the scores of most respondents indicate high levels of social integration.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,815	,819	6

The figure above shows the Cronbach's alpha for all the six items (Cronbach's alpha = 0.82). The high value on this test tells that the items measure a similar construct. Therefore the six items are used to measure social integration of immigrants.

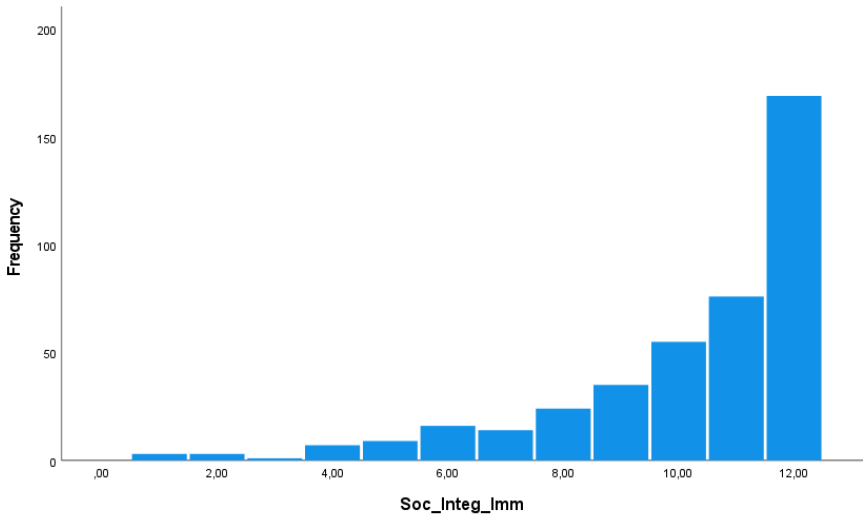
A new variable has been computed that sums up the scores of respondents on the statements. Before this procedure, the scales of some variables were changed into a high score indicating positive feelings of being socially integrated.

```
RECODE fb14b307 rfb14b308 rfb14b309 rfb14b310 fb14b311 fb14b312 (1=0) (2=1) (3=2) INTO
Emptiness
   Count_on Rely_on Connected Miss_people Deserted.
EXECUTE.

COMPUTE Soc_Integ_Imm=Emptiness + Count_on + Rely_on + Connected + Miss_people +
Deserted.
EXECUTE.

FREQUENCIES VARIABLES=Soc_Integ_Imm
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN
  /ORDER=ANALYSIS.

GRAPH
  /HISTOGRAM=Soc_Integ_Imm.
```

Statistics

Soc_Integ_Imm

N	Valid	412
	Missing	0
Mean		10,1942
Median		11,0000
Std. Deviation		2,31911
Minimum		1,00
Maximum		12,00



The final variable *social integration of immigrants* has a left-skewed distribution. Most respondents have a maximum score on this variable, therefore a peak is seen in the frequency on the right tail. This tells that most respondents have a good feeling about their social integration. There are a few (extreme) outliers on the left tail of the distribution. The data clearly shows a violation of the assumption of normality.

1.3 Gender

The original item from the dataset is shown below.

- Geslacht: Gender

  1= male

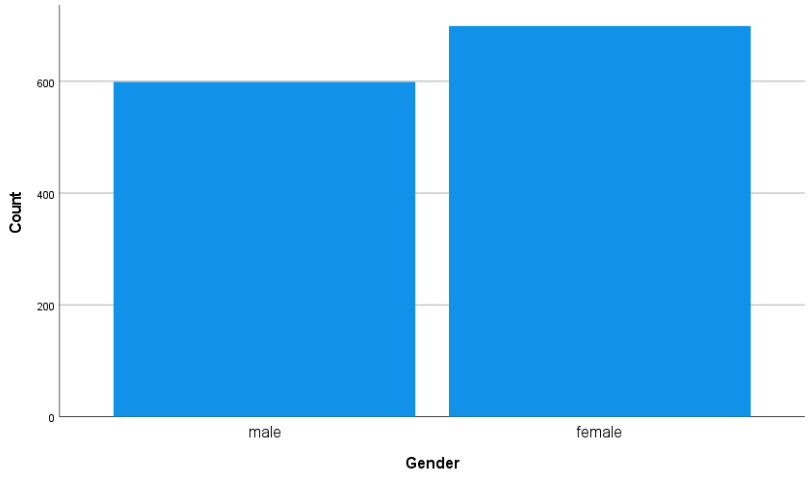
  2= female

To make the interpretation of gender easier, the answer possibilities 1 (*male*) and 2 (*female*) are recoded into 0 (*male*) and 1 (*female*).

```
FREQUENCIES VARIABLES=geslacht
  /ORDER=ANALYSIS.

RECODE geslacht (1=0) (2=1) INTO Gender.
EXECUTE.

GRAPH
  /HISTOGRAM=Gender.

FREQUENCIES VARIABLES=Gender
  /ORDER=ANALYSIS.

GRAPH
  /BAR(SIMPLE)=COUNT BY Gender.
```

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	599	46,1	46,1	46,1
	Female	699	53,9	53,9	100,0
	Total	1298	100,0	100,0	



The distribution of men and women in the dataset is quite equal. There are slightly more women than men in this dataset.

1.4 Education

The original item from the dataset is shown below.

- Oplmet: Highest level of education with diploma

  1= primary school

  2= vmbo (intermediate secondary education, US: junior high school)

  3= havo/vwmo (higher secondary education/preparatory university education, US: senior high school)

  4= mbo (intermediate vocational educaiton, US: junior college)

  5= hbo (higher vocational education, US: college)

  6= wo (university)

  7= other

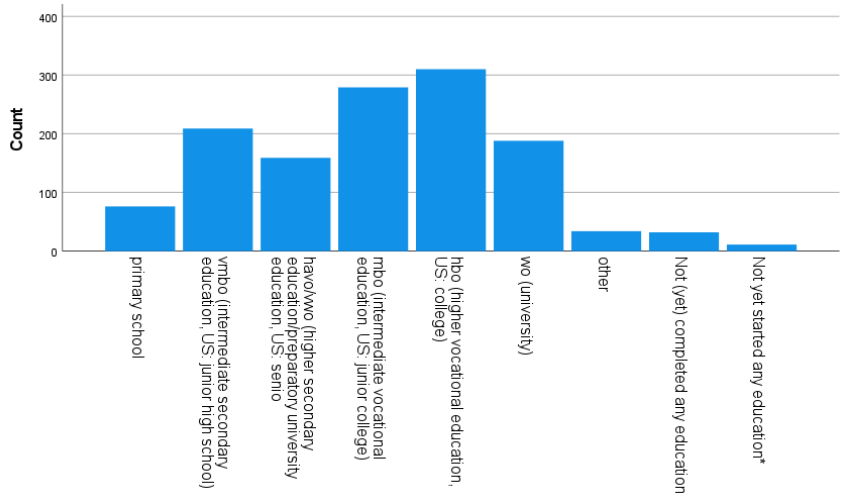
  8= not (yet) completed any study

  9= not yet started any education

```
FREQUENCIES VARIABLES=oplmet
  /ORDER=ANALYSIS.

GRAPH
  /BAR(SIMPLE)=COUNT BY oplmet.
```

Highest level of education with diploma

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	primary school	76	5,9	5,9	5,9
	vmbo (intermediate secondary education, US: junior high school)	209	16,1	16,1	22,0
	havo/vwo (higher secondary education/preparatory university education, US: senio	159	12,2	12,2	34,2
	mbo (intermediate vocational education, US: junior college)	279	21,5	21,5	55,7
	hbo (higher vocational education, US: college)	310	23,9	23,9	79,6
	wo (university)	188	14,5	14,5	94,1
	other	34	2,6	2,6	96,7
	Not (yet) completed any education	32	2,5	2,5	99,2
	Not yet started any education*	11	,8	,8	100,0
	Total	1298	100,0	100,0	



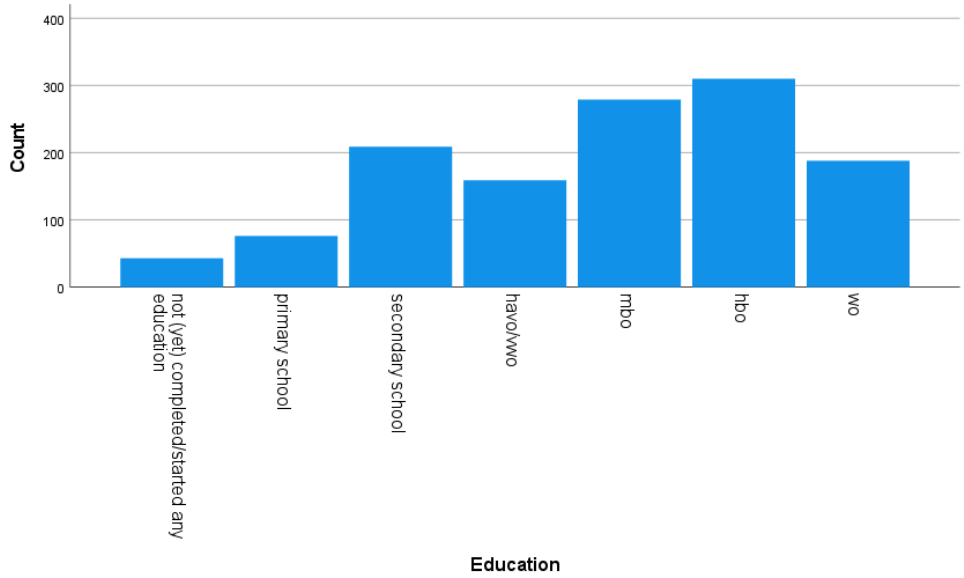
The figures above show the descriptives of the original question. The respondents are not equally distributed across the categories. Most respondents report having an MBO or HBO level diploma. Very few respondents have not obtained or started any education.

To control for the effect of educational background on the social integration of immigrants, a few changes have been made. The answer possibility 7 is vague and therefore cannot be placed on an ordinal scale between the other answer possibilities. Therefore, this answer is recoded into a missing value. Answer possibilities 8 and 9 formally indicate that a respondent has not obtained any degree. These situations are ranked below other educational backgrounds and were recoded into the value 0.

```
RECODE oplmet (9=0) (8=0) (7=SYSMIS) (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) INTO Education.
EXECUTE.

FREQUENCIES VARIABLES=Education
  /ORDER=ANALYSIS.

GRAPH
  /BAR(SIMPLE)=COUNT BY Education.
```

Education

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	43	3,3	3,4	3,4
	1,00	76	5,9	6,0	9,4
	2,00	209	16,1	16,5	25,9
	3,00	159	12,2	12,6	38,5
	4,00	279	21,5	22,1	60,6
	5,00	310	23,9	24,5	85,1
	6,00	188	14,5	14,9	100,0
	Total	1264	97,4	100,0	
Missing	System	34	2,6		
Total		1298	100,0		



The final variable *education* has an unequal distribution of respondents among the categories. Most respondents have obtained an (or an equivalent of an) MBO or HBO diploma, as seen in the high number of respondents in those categories. There are relatively few respondents who fit the first two categories that represent not having completed or started any education and having a primary school

diploma. Due to some categories having very few respondents, the outcomes of the analyses can have a lower statistical power and limited possibilities to generalise the effects of the educational background on the social integration of immigrants.

1.5 Group of origin

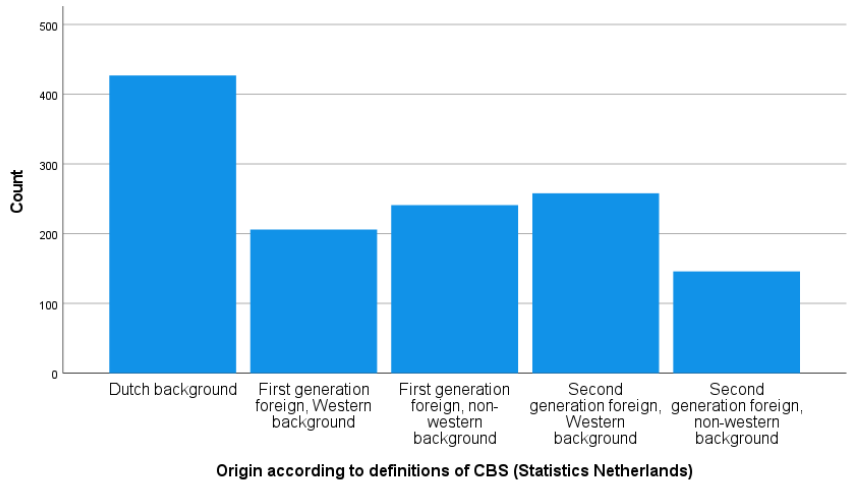
The original item from the dataset is shown below.

- Herkomstgroep: origin according to definitions of Statistics Netherlands
  - 0= Dutch background
  - 101= first generation foreign, Western background
  - 102= first generation foreign, non-western background
  - 201= second generation foreign, Western background
  - 202= second generation foreign, non-western background
  - 999= origin unknown or part of the information unknown (missing values)

```
FREQUENCIES VARIABLES=herkomstgroep
  /ORDER=ANALYSIS.

GRAPH
  /BAR(SIMPLE)=COUNT BY herkomstgroep.
```

Origin according to definitions of CBS (Statistics Netherlands)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Dutch background	427	32,9	33,4	33,4
	First generation foreign, Western background	206	15,9	16,1	49,5
	First generation foreign, non-western background	241	18,6	18,9	68,4
	Second generation foreign, Western background	258	19,9	20,2	88,6
	Second generation foreign, non-western background	146	11,2	11,4	100,0
	Total	1278	98,5	100,0	
Missing	System	20	1,5		
Total		1298	100,0		



Origin according to definitions of CBS (Statistics Netherlands)

The figures above show the descriptives of the original question. Most respondents report having a Dutch background according to the definitions of Statistics Netherlands (Statistics Netherlands, n.d.-b). The distribution of respondents across the categories is not equal, but the differences in group size are relatively small.

The values of all the answer possibilities are recoded into lower values, to facilitate an easier interpretation of the final variable. The last answer possibility is already recoded into a missing value, so all other answer possibilities can be used for the final variable.

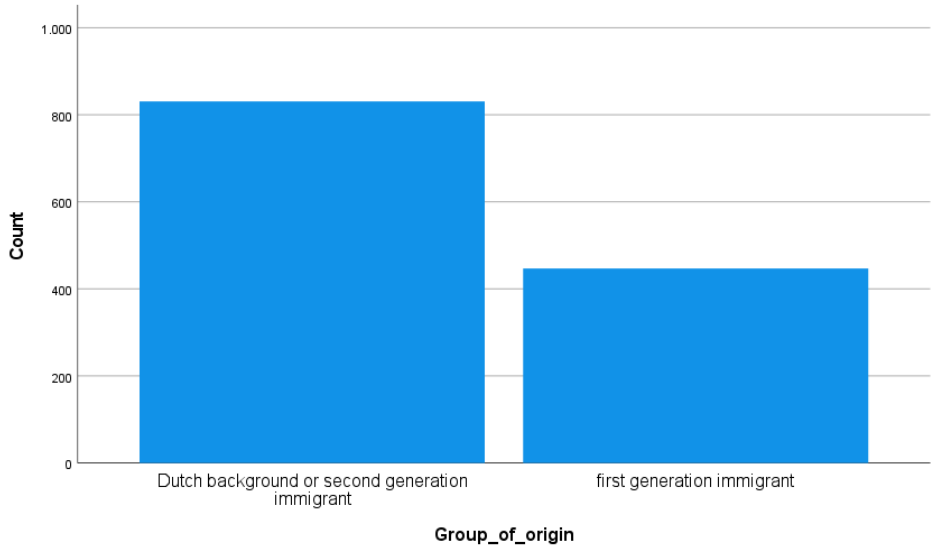
As the variable *group of origin* is a nominal variable, the answer possibilities are divided into two groups. According to the literature, second-generation immigrants have on average built up more resources, skills and knowledge that favours their process of becoming socially integrated in comparison to first-generation immigrants (Smith et al., 2018). Therefore, a division between

immigrants with a Dutch background/second-generation immigrants and first-generation immigrants is made.

```
RECODE herkomstgroep (0=0) (101=1) (102=1) (201=0) (202=0) (999=SYSMIS) INTO
Group_of_origin.
EXECUTE.

FREQUENCIES VARIABLES=Group_of_origin
  /ORDER=ANALYSIS.

GRAPH
  /BAR(SIMPLE)=COUNT BY Group_of_origin.
```

Group_of_origin

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	831	64,0	65,0	65,0
	1,00	447	34,4	35,0	100,0
	Total	1278	98,5	100,0	
Missing	System	20	1,5		
Total		1298	100,0		



The final variable *group of origin* has an unequal distribution of respondents among the categories. Most respondents fit into the first category, which indicates that many respondents have a Dutch background or are second generation immigrants.

1.6 Primary occupation

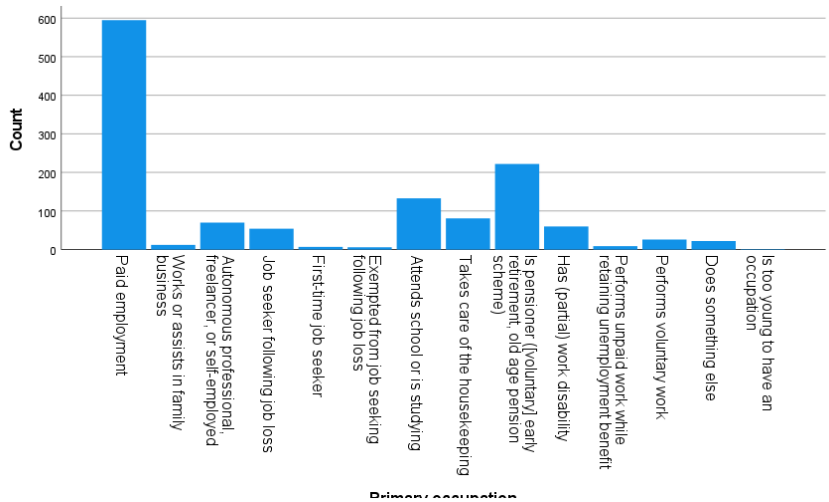
The original item from the dataset is shown below.

- Belbezig: primary occupation
  - 1= paid employment
  - 2= works or assists in family business
  - 3= autonomous professional, freelancer, or self-employed
  - 4= job seeker following job loss
  - 5= first-time job seeker
  - 6= exempted from job seeking following job loss
  - 7= attends school or is studying
  - 8= takes care of the housekeeping
  - 9= is pensioner ([voluntary] early retirement, old age pension scheme)
  - 10= has (partial) work disability
  - 11= performs unpaid work while retaining unemployment benefit
  - 12= performs voluntary work
  - 13= does something else
  - 14= is too young to have an occupation

```
FREQUENCIES VARIABLES=belbezig
  /ORDER=ANALYSIS.

GRAPH
  /BAR(SIMPLE)=COUNT BY belbezig.
```

Primary occupation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Paid employment	595	45,8	45,8	45,8
	Works or assists in family business	12	,9	,9	46,8
	Autonomous professional, freelancer, or self-employed	70	5,4	5,4	52,2
	Job seeker following job loss	54	4,2	4,2	56,3
	First-time job seeker	7	,5	,5	56,9
	Exempted from job seeking following job loss	6	,5	,5	57,3
	Attends school or is studying	133	10,2	10,2	67,6
	Takes care of the housekeeping	81	6,2	6,2	73,8
	Is pensioner ([voluntary] early retirement, old age pension scheme)	222	17,1	17,1	90,9
	Has (partial) work disability	60	4,6	4,6	95,5
	Performs unpaid work while retaining unemployment benefit	9	,7	,7	96,2
	Performs voluntary work	26	2,0	2,0	98,2
	Does something else	22	1,7	1,7	99,9
	Is too young to have an occupation	1	,1	,1	100,0
	Total	1298	100,0	100,0	



The figures above show the descriptives of the original question. Through comparing all the possible categories that indicate the primary occupation of a respondent, most respondents report being in paid employment. Apart from respondents who have a paid job, others often report being a pensioner or student. Many categories have very few respondents, so the distribution of data is not equal.

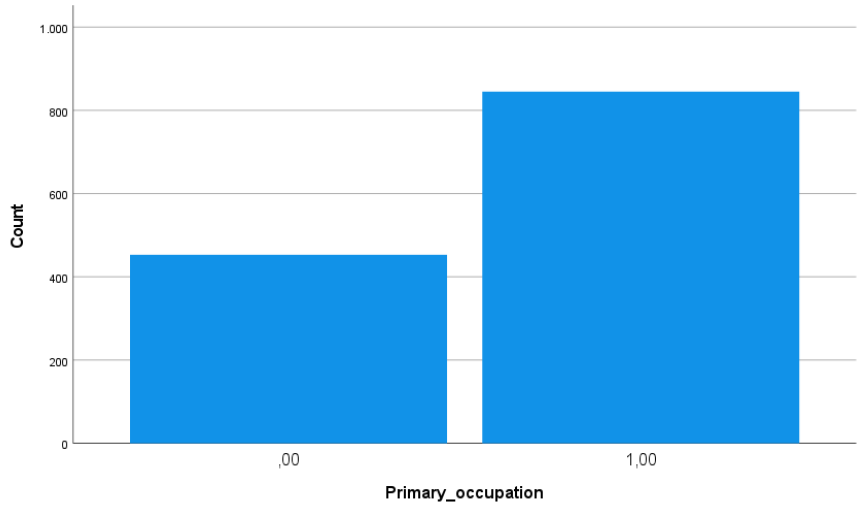
The original item has many answer possibilities that all indicate the different occupational statuses of respondents. Many independent categories have few respondents which could create difficulties with generalisation of the statistical outcomes. As stated in the literature, being employed or doing volunteer work creates opportunities for social interaction (Grimm et al., 2023; Sveen et al.,

2022). Having such an occupational status could therefore enhance the process of becoming socially integrated. For the final variable, a division is created between occupational statuses that facilitate socializing and those that do not facilitate socializing.

```
RECODE belbezig (1=1) (2=1) (3=1) (4=0) (5=0) (6=0) (7=1) (8=0) (9=0) (10=0) (11=1) (12=1) (13=0)
   (14=0) INTO Primary_occupation.
EXECUTE.

FREQUENCIES VARIABLES=Primary_occupation
 /ORDER=ANALYSIS.

GRAPH
 /BAR(SIMPLE)=COUNT BY Primary_occupation.
```

Primary_occupation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	453	34,9	34,9	34,9
	1,00	845	65,1	65,1	100,0
	Total	1298	100,0	100,0	



The final variable *primary occupation* has an unequal distribution of respondents among the two categories. Most respondents fit the latter category, indicating that many respondents have an occupation that facilitates socializing.

1.7 Moderator: Gender and Sporting

```
FREQUENCIES VARIABLES=Gender sporting
  /FORMAT=NOTABLE
  /STATISTICS=MEAN
  /ORDER=ANALYSIS.

COMPUTE sporting_c=sporting - 2.1784.
EXECUTE.

COMPUTE gender_c=Gender - 0.5385.
EXECUTE.

COMPUTE SportingxGender=sporting_c * gender_c.
EXECUTE.
```

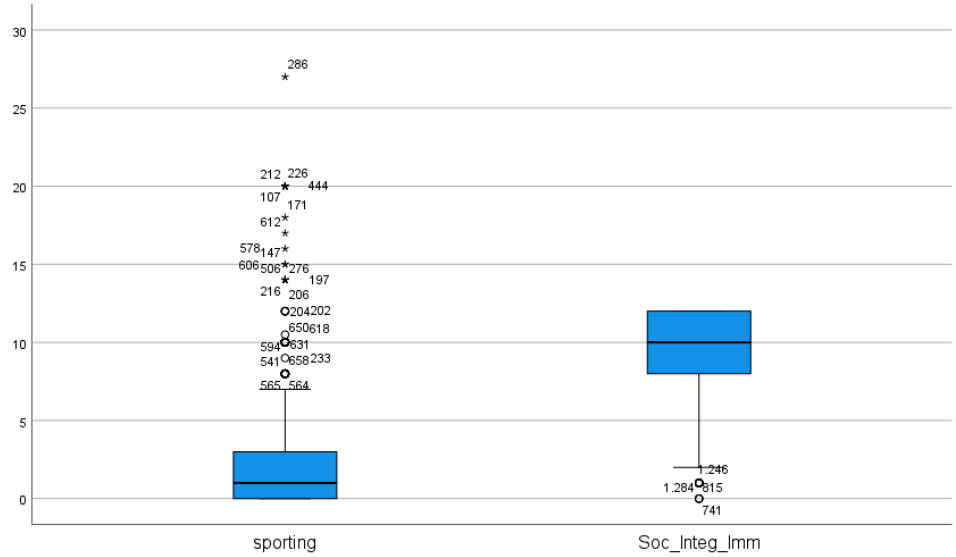
Statistics

		Gender	sporting
N	Valid	1298	1298
	Missing	0	0
Mean		,5385	2,1784

The interaction between *sporting* and *gender* is made by mean centering the variables before creating the interaction term.

1.8 Boxplots

```
EXAMINE VARIABLES=sporting Soc_Integ_Imm
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.
```



The boxplots for the continuous variables *sporting* and *social integration of immigrants* are given. Sporting has a long upper tail in comparison to the lower tail. This points out the right-skewed distribution of the data. There are quite a few (extreme) outliers which are also shown in the boxplot. Furthermore, the boxplot of social integration of immigrants does not have an upper tail. This indicates the left-skewed distribution of the data.

1.9 Missing values

The overview of missing data per variable is given here. The final sample size is smaller than the size of the dataset, as some respondents did not answer all the relevant questions or showed missing values on one or more questions after recoding the original variables. An indicator is made for the multivariate missing data and is used in analysing the descriptives of final variables and the final regressions.

Table 7 below shows the missing data statistics of all the final variables. Compared to the original dataset, the variable sporting has the highest number of missing data of all variables used in this research (70.2%). Other continuous variables also show some significant declines in the number of valid cases. In contrast, the final variables of the moderator Gender and the control variables do not have many missing values. The decline from 2227 respondents in the original dataset to 412 respondents in the sample size for this research is mainly influenced by the high number of missing data among the explanatory variables.

```
FREQUENCIES VARIABLES=Gender sporting Soc_Integ_Imm Group_of_origin Education
Primary_occupation
  SportingxGender
 /FORMAT=NOTABLE
 /STATISTICS=MEAN
 /ORDER=ANALYSIS.
```

Statistics

		Gender	sporting	Soc_Integ_Imm	Group_of_origin	Education	Primary_occupation	SportingxGender
N	Valid	1298	1298	1285	1278	1264	1298	1298
	Missing	0	0	13	20	34	0	0
Mean		,5385	2,1784	9,6226	,3498	3,7698	,6510	-,1066

As seen in the syntax below, a regression with saved residuals has been done. The respondents that have a valid answer on all the variables have been marked as complete cases. The respondents that had one or more missings on the variables were marked as incomplete cases and were filtered out of the final sample.

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT Soc_Integ_Imm
 /METHOD=ENTER Education Group_of_origin Primary_occupation Sporting Reading Going_out
   Gender MX1 MX2 MX3
 /SAVE RESID.


RECODE RES_1 (SYSMIS=0) (ELSE=1) INTO Observations.
EXECUTE.


USE ALL.
COMPUTE filter_$=(Observations = 1).
VARIABLE LABELS filter_$ 'Observations = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
```

```
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

USE ALL.
FILTER BY filter_$.
EXECUTE.
```

Appendix 2 – Bivariate descriptives and model analysis

This appendix contains information on the hierarchical linear regression analysis that has been done with the corresponding syntax and interpretation. Before that, the bivariate statistics are given to indicate correlations between the variables. The output from the regression analysis will be used to test the hypotheses.

2.1. Bivariate statistics

The correlations are given for the continuous and categorical variables below. For continuous variables, correlation matrixes are made. Furthermore, scatterplots are made using continuous independent variables and the dependent variable. For all categorical variables, crosstabs are used to indicate the correlations between them. Lastly, F-tests and T-tests have been done to indicate the correlation between the continuous and categorical variables.

2.1.1 Pearson's correlations

The correlations between continuous and continuous or dummy variables are estimated by using the Pearson correlation. The continuous variables *sporting* and *social integration of immigrants* and the dummy variables *gender*, *group of origin* and *primary occupation* are used.

```
CORRELATIONS
  /VARIABLES=sporting Soc_Integ_Imm Gender Group_of_origin Primary_occupation
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.
```

Correlations

		sporting	Soc_Integ_Imm	Gender	Group_of_origin	Primary_occupation
sporting	Pearson Correlation	1	,123**	-,061*	-,045	,090**
	Sig. (2-tailed)		<,001	,031	,116	,002
	N	1232	1232	1232	1232	1232
Soc_Integ_Imm	Pearson Correlation	,123**	1	,027	-,194**	,106**
	Sig. (2-tailed)	<,001		,353	<,001	<,001
	N	1232	1232	1232	1232	1232
Gender	Pearson Correlation	-,061*	,027	1	-,002	-,064*
	Sig. (2-tailed)	,031	,353		,935	,024
	N	1232	1232	1232	1232	1232
Group_of_origin	Pearson Correlation	-,045	-,194**	-,002	1	-,086**
	Sig. (2-tailed)	,116	<,001	,935		,002
	N	1232	1232	1232	1232	1232
Primary_occupation	Pearson Correlation	,090**	,106**	-,064*	-,086**	1
	Sig. (2-tailed)	,002	<,001	,024	,002	
	N	1232	1232	1232	1232	1232

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

2.1.2 Cramer's V

The correlations between two categorical variables or two dummy variables are estimated using the Cramer's V. The categorical variables and dummy variables that are used are *gender*, *education*, *group of origin* and *primary occupation*.

```
CROSSTABS
  /TABLES=Gender BY Education
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Gender * Education Crosstabulation

Count

		Education							
		,00	1,00	2,00	3,00	4,00	5,00	6,00	Total
Gender	,00	16	28	93	72	128	132	103	572
	1,00	25	41	112	87	139	174	82	660
Total		41	69	205	159	267	306	185	1232

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	9,968[a]	6	,126
Likelihood Ratio	9,971	6	,126
Linear-by-Linear Association	3,686	1	,055
N of Valid Cases	1232		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 19,04.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,090	,126
	Cramer's V	,090	,126
N of Valid Cases		1232	

```
CROSSTABS
  /TABLES=Gender BY Group_of_origin
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Gender * Group_of_origin Crosstabulation

Count

		Group_of_origin		
		,00	1,00	Total
Gender	,00	374	198	572
	1,00	433	227	660
Total		807	425	1232

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,007[a]	1	,935		
Continuity Correction[b]	,000	1	,983		
Likelihood Ratio	,007	1	,935		
Fisher's Exact Test				,952	,491
Linear-by-Linear Association	,007	1	,935		
N of Valid Cases	1232				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 197,32.

b. Computed only for a 2x2 table

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-,002	,935
	Cramer's V	,002	,935
N of Valid Cases		1232	

```
CROSSTABS
  /TABLES=Gender BY Primary_occupation
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Gender * Primary_occupation Crosstabulation

Count

		Primary_occupation		
		,00	1,00	Total
Gender	,00	179	393	572
	1,00	247	413	660
Total		426	806	1232

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5,091[a]	1	,024		
Continuity Correction[b]	4,824	1	,028		
Likelihood Ratio	5,106	1	,024		
Fisher's Exact Test				,026	,014
Linear-by-Linear Association	5,087	1	,024		
N of Valid Cases	1232				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 197,79.

b. Computed only for a 2x2 table

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-,064	,024
	Cramer's V	,064	,024
N of Valid Cases		1232	

```
CROSSTABS
  /TABLES=Education BY Group_of_origin
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Education * Group_of_origin Crosstabulation

Count

		Group_of_origin		
		,00	1,00	Total
Education	,00	24	17	41
	1,00	46	23	69
	2,00	135	70	205
	3,00	117	42	159
	4,00	185	82	267
	5,00	206	100	306
	6,00	94	91	185
Total		807	425	1232

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	25,342[a]	6	<,001
Likelihood Ratio	24,712	6	<,001
Linear-by-Linear Association	3,984	1	,046
N of Valid Cases	1232		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 14,14.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,143	<,001
	Cramer's V	,143	<,001
N of Valid Cases		1232	

```
CROSSTABS
  /TABLES=Education BY Primary_occupation
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Education * Primary_occupation Crosstabulation

Count

		Primary_occupation ,00	Primary_occupation 1,00	Total
Education	,00	16	25	41
	1,00	33	36	69
	2,00	113	92	205
	3,00	47	112	159
	4,00	80	187	267
	5,00	93	213	306
	6,00	44	141	185
Total		426	806	1232

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	60,142[a]	6	<,001
Likelihood Ratio	58,450	6	<,001
Linear-by-Linear Association	36,478	1	<,001
N of Valid Cases	1232		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 14,18.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,221	<,001
	Cramer's V	,221	<,001
N of Valid Cases		1232	

```
CROSSTABS
  /TABLES=Primary_occupation BY Group_of_origin
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Primary_occupation * Group_of_origin Crosstabulation

Count

		Group_of_origin		
		,00	1,00	Total
Primary_occupation	,00	255	171	426
	1,00	552	254	806
Total		807	425	1232

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9,180[a]	1	,002		
Continuity Correction[b]	8,802	1	,003		
Likelihood Ratio	9,090	1	,003		
Fisher's Exact Test				,003	,002
Linear-by-Linear Association	9,172	1	,002		
N of Valid Cases	1232				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 146,96.

b. Computed only for a 2x2 table

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-,086	,002
	Cramer's V	,086	,002
N of Valid Cases		1232	

2.1.3 UNIANOVA

The correlation between the categorical variable and the continuous variable is measured using the UNIANOVA method and the corresponding R^2. By The variables that are used for this method are *sporting*, *social integration of immigrants*, and *education*.

```
UNIANOVA Soc_Integ_Imm BY Education
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN=Education.
```

Between-Subjects Factors

		N
Education	,00	41
	1,00	69
	2,00	205
	3,00	159
	4,00	267
	5,00	306
	6,00	185

Tests of Between-Subjects Effects

Dependent Variable: Soc_Integ_Imm

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	116,213[a]	6	19,369	2,768	,011
Intercept	70752,076	1	70752,076	10112,240	,000
Education	116,213	6	19,369	2,768	,011
Error	8570,929	1225	6,997		
Total	123225,000	1232			
Corrected Total	8687,142	1231			

a. R Squared = ,013 (Adjusted R Squared = ,009)

```
UNIANOVA sporting BY Education
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN=Education.
```

Between-Subjects Factors

		N
Education	,00	41
	1,00	69
	2,00	205
	3,00	159
	4,00	267
	5,00	306
	6,00	185

Tests of Between-Subjects Effects

Dependent Variable: sporting

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	85,292[a]	6	14,215	1,438	,197
Intercept	3538,252	1	3538,252	357,966	<,001
Education	85,292	6	14,215	1,438	,197
Error	12108,301	1225	9,884		
Total	18008,680	1232			
Corrected Total	12193,593	1231			

a. R Squared = ,007 (Adjusted R Squared = ,002)

2.2. Linear regression analysis

The hierarchical linear regression analysis is done using four models. In the first model, social integration of immigrants is used as the dependent variable with control variables education, group of origin, and primary occupation. The independent variable sporting has been added to the second model. The moderator gender has been included in the third model. The interaction term and all the other variables are put in the fourth model.

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA CHANGE
  /CRITERIA=PIN(.05) POUT(.10)
```

```
/NOORIGIN
/DEPENDENT Soc_Integ_Imm
/METHOD=ENTER Education Group_of_origin Primary_occupation
/METHOD=ENTER sporting
/METHOD=ENTER Gender
/METHOD=ENTER SportingxGender.
```

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,235[a]	,055	,053	2,58538	,055	23,885	3	1228	<,001
2	,257[b]	,066	,063	2,57118	,011	14,606	1	1227	<,001
3	,261[c]	,068	,064	2,56978	,002	2,336	1	1226	,127
4	,263[d]	,069	,064	2,56957	,001	1,201	1	1225	,273

a. Predictors: (Constant), Primary_occupation, Group_of_origin, Education

b. Predictors: (Constant), Primary_occupation, Group_of_origin, Education, sporting

c. Predictors: (Constant), Primary_occupation, Group_of_origin, Education, sporting, Gender

d. Predictors: (Constant), Primary_occupation, Group_of_origin, Education, sporting, Gender, SportingxGender

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	478,954	3	159,651	23,885	<,001[b]
	Residual	8208,188	1228	6,684		
	Total	8687,142	1231			
2	Regression	575,516	4	143,879	21,764	<,001[c]
	Residual	8111,626	1227	6,611		
	Total	8687,142	1231			
3	Regression	590,941	5	118,188	17,897	<,001[d]
	Residual	8096,201	1226	6,604		
	Total	8687,142	1231			
4	Regression	598,873	6	99,812	15,117	<,001[e]
	Residual	8088,269	1225	6,603		
	Total	8687,142	1231			

a. Dependent Variable: Soc_Integ_Imm

b. Predictors: (Constant), Primary_occupation, Group_of_origin, Education

c. Predictors: (Constant), Primary_occupation, Group_of_origin, Education, sporting

d. Predictors: (Constant), Primary_occupation, Group_of_origin, Education, sporting, Gender

e. Predictors: (Constant), Primary_occupation, Group_of_origin, Education, sporting, Gender, SportingxGender

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	9,138	,207		44,141	<,001	8,732	9,545
	Education	,162	,046	,099	3,510	<,001	,072	,253
	Group_of_origin	-1,081	,156	-,194	-6,933	<,001	-1,387	-,775
	Primary_occupation	,402	,158	,072	2,543	,011	,092	,712
2	(Constant)	8,980	,210		42,760	<,001	8,568	9,392
	Education	,159	,046	,097	3,463	<,001	,069	,249
	Group_of_origin	-1,058	,155	-,189	-6,819	<,001	-1,363	-,754
	Primary_occupation	,352	,158	,063	2,234	,026	,043	,661
	sporting	,089	,023	,106	3,822	<,001	,044	,135
3	(Constant)	8,835	,230		38,347	<,001	8,383	9,287
	Education	,162	,046	,099	3,528	<,001	,072	,252
	Group_of_origin	-1,057	,155	-,189	-6,812	<,001	-1,361	-,752
	Primary_occupation	,364	,158	,065	2,310	,021	,055	,674
	sporting	,091	,023	,108	3,903	<,001	,045	,137
	Gender	,225	,147	,042	1,528	,127	-,064	,515
4	(Constant)	8,835	,230		38,350	<,001	8,383	9,287
	Education	,161	,046	,099	3,507	<,001	,071	,251
	Group_of_origin	-1,061	,155	-,190	-6,836	<,001	-1,365	-,756
	Primary_occupation	,369	,158	,066	2,337	,020	,059	,678
	sporting	,095	,024	,112	4,011	<,001	,048	,141
	Gender	,226	,147	,042	1,530	,126	-,064	,515
	SportingxGender	,051	,047	,031	1,096	,273	-,040	,143

a. Dependent Variable: Soc_Integ_Imm

Appendix 3 - Model assumptions, multicollinearity and outliers

This appendix displays the details behind the linear regression. First, the four assumptions for linear regression are examined. After that, the possibility of multicollinearity among variables is inspected. A check for influential points and outliers is done at the end of this appendix.

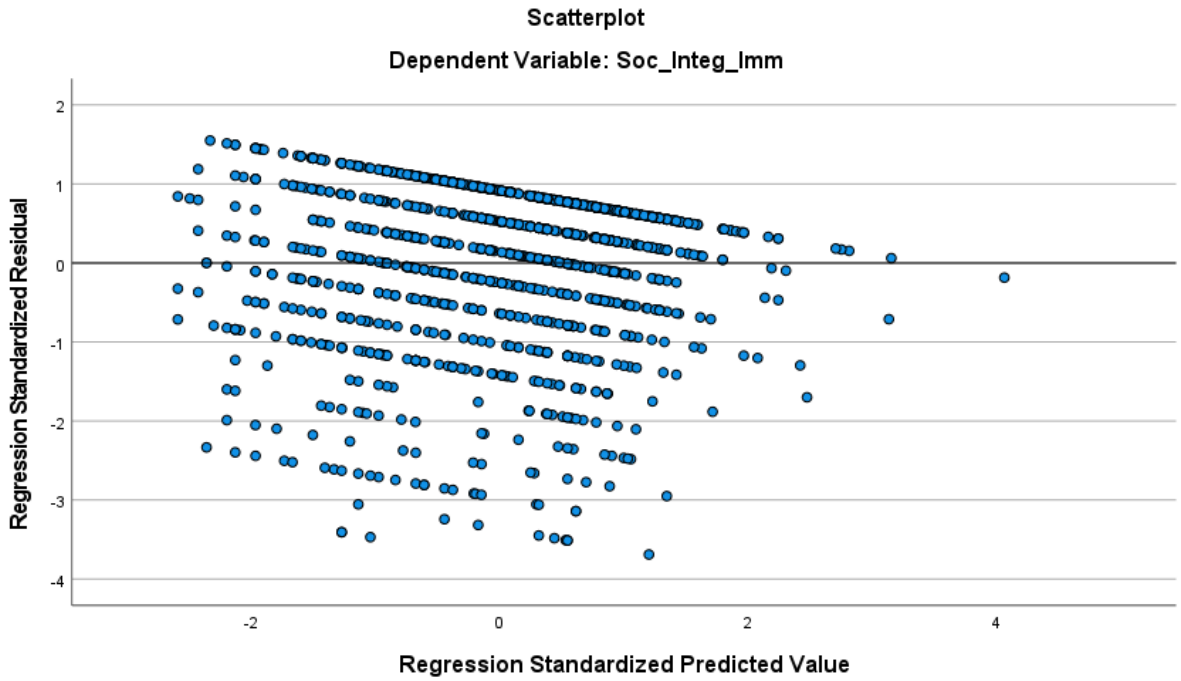
```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Soc_Integ_Imm
  /METHOD=ENTER Education Group_of_origin Primary_occupation
  /METHOD=ENTER sporting
  /METHOD=ENTER Gender
  /METHOD=ENTER SportingxGender
  /PARTIALPLOT ALL
  /SCATTERPLOT=(*ZRESID ,*ZPRED)
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /SAVE PRED ZPRED COOK LEVER RESID ZRESID DFFIT.
```

3.1 Model assumptions

The first assumption of linear regression analysis is that observations are independent of each other. The LISS Immigrant panel that is used for this research does not fully contain respondents who participate independently from each other. Households can have several people participating in the same research. Therefore, some respondents in the sample could show similarities with other respondents who are from the same household. However, the sample size of the entire LISS Immigrant panel is relatively large. Therefore, this issue is not too problematic for the statistical outcomes of this research.

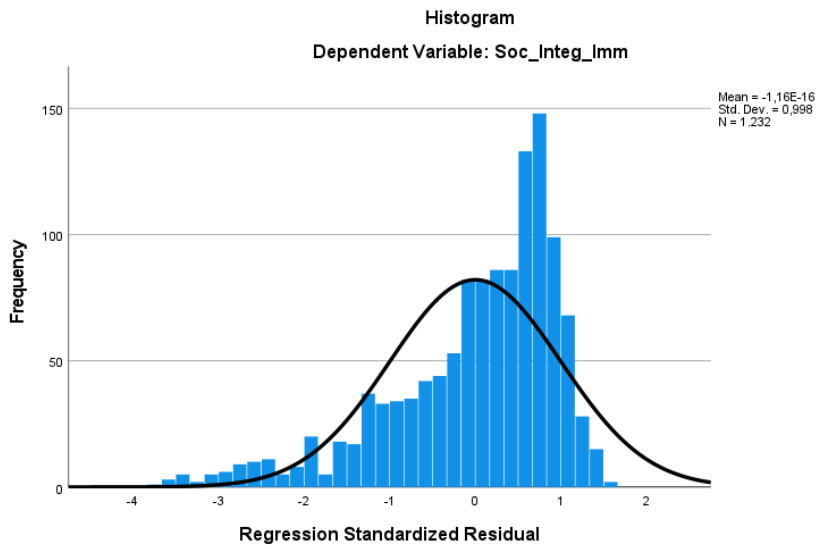
The second assumption requires a linear relationship between the explanatory variable and the dependent variable. A residual plot is made to examine whether the mean of the residuals is close to 0 for every set of values on the horizontal axis and whether there are systematic deviations from the line y=0. As seen in the figure below, a scatterplot of the standardized residuals and the

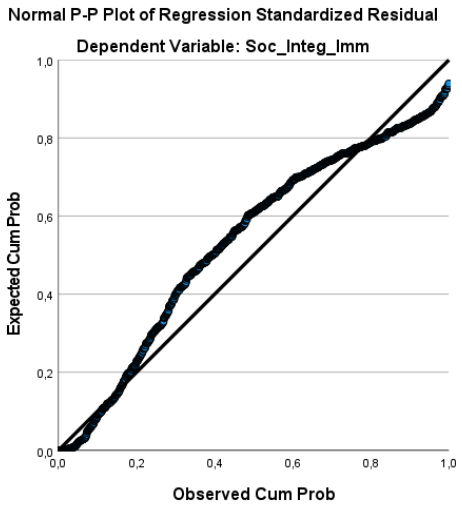
standardized predicted residuals is made. Not all the points in the graph are nicely divided across the horizontal axis. Besides that, not all standardized residuals are close to the line that goes through y=0. In conclusion, the assumption of linearity is slightly violated.



The third assumption requires homoscedasticity. In the scatterplot above, the standard deviation is not equal for all the values on the horizontal axis. There seems to be unequal variance around the y=0 line for different values of x. Therefore, the assumption of homoscedasticity is also slightly violated.

The fourth and last assumption of linear regression analysis is about normality. The standardized residuals should be divided according to a normal distribution of the data. A histogram of the standardized residuals is shown below. The data does not have a normal curve and the distribution is left-skewed. There are a few outliers visible on the left tail of the graph. Furthermore, the peak of the distribution is positioned too far on the right side of the distribution. A P-P plot is shown below too and deviates a bit from the diagnol. This too indicates that the data does not have a normal distribution. The dependent variable *social integration of immigrants* is not a perfect continuous variable. This will have caused the assumption of normality to be slightly violated.





3.2. Multicollinearity

It is possible that some independent variables correlate strongly. A VIF test for multicollinearity will examine whether this is the case for the set of variables used in this research. The VIF scores of the fourth and final model in the output below show VIF ranges from 1.009 to 1.051. As these values are all below the criteria of VIF < 4, multicollinearity is not present among the variables.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	9,138	,207		44,141	<,001		
	Education	,162	,046	,099	3,510	<,001	,965	1,036
	Group_of_origin	-1,081	,156	-,194	-6,933	<,001	,987	1,013
	Primary_occupation	,402	,158	,072	2,543	,011	,961	1,040
2	(Constant)	8,980	,210		42,760	<,001		
	Education	,159	,046	,097	3,463	<,001	,965	1,036
	Group_of_origin	-1,058	,155	-,189	-6,819	<,001	,986	1,014
	Primary_occupation	,352	,158	,063	2,234	,026	,955	1,048
	sporting	,089	,023	,106	3,822	<,001	,990	1,010
3	(Constant)	8,835	,230		38,347	<,001		
	Education	,162	,046	,099	3,528	<,001	,963	1,038
	Group_of_origin	-1,057	,155	-,189	-6,812	<,001	,986	1,014
	Primary_occupation	,364	,158	,065	2,310	,021	,952	1,050
	sporting	,091	,023	,108	3,903	<,001	,987	1,013
	Gender	,225	,147	,042	1,528	,127	,991	1,009
4	(Constant)	8,835	,230		38,350	<,001		
	Education	,161	,046	,099	3,507	<,001	,963	1,039
	Group_of_origin	-1,061	,155	-,190	-6,836	<,001	,985	1,015
	Primary_occupation	,369	,158	,066	2,337	,020	,951	1,051
	sporting	,095	,024	,112	4,011	<,001	,971	1,030
	Gender	,226	,147	,042	1,530	,126	,991	1,009
	SportingxGender	,051	,047	,031	1,096	,273	,981	1,019

a. Dependent Variable: Soc_Integ_Imm

3.3. Outliers

The standardized residuals, leverage, Cook's distance and the DFFIT have been examined in order to test for outliers and influential points in the selected sample. Some cases might be influential in the y-direction and/or x-direction. The standardized residuals have been saved when doing a regression in order to examine these in terms of their magnitude.

For outliers in the y-direction, the standardized residuals of all cases are examined. Cases that have standardized residuals larger than 3 or smaller than -3 are displayed in the table below. 16 influential cases have been found and are presented as the number of the household member.

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA CHANGE
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT Soc_Integ_Imm
```

```
/METHOD=ENTER Education Group_of_origin Primary_occupation sporting Gender
SportingxGender
/SCATTERPLOT=(*ZRESID ,*ZPRED)
/RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID)
/SAVE COOK LEVER RESID ZRESID SRESID DFFIT.
```

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	7,8340	12,4755	9,6420	,69749	1232
Std. Predicted Value	-2,592	4,062	,000	1,000	1232
Standard Error of Predicted Value	,126	,799	,186	,055	1232
Adjusted Predicted Value	7,8131	12,4994	9,6426	,69885	1232
Residual	-9,47952	3,98461	,00000	2,56330	1232
Std. Residual	-3,689	1,551	,000	,998	1232
Stud. Residual	-3,694	1,556	,000	1,000	1232
Deleted Residual	-9,50644	4,01433	-,00055	2,57738	1232
Stud. Deleted Residual	-3,714	1,557	-,001	1,002	1232
Mahal. Distance	1,939	118,000	5,995	6,698	1232
Cook's Distance	,000	,017	,001	,002	1232
Centered Leverage Value	,002	,096	,005	,005	1232

a. Dependent Variable: Soc_Integ_Imm

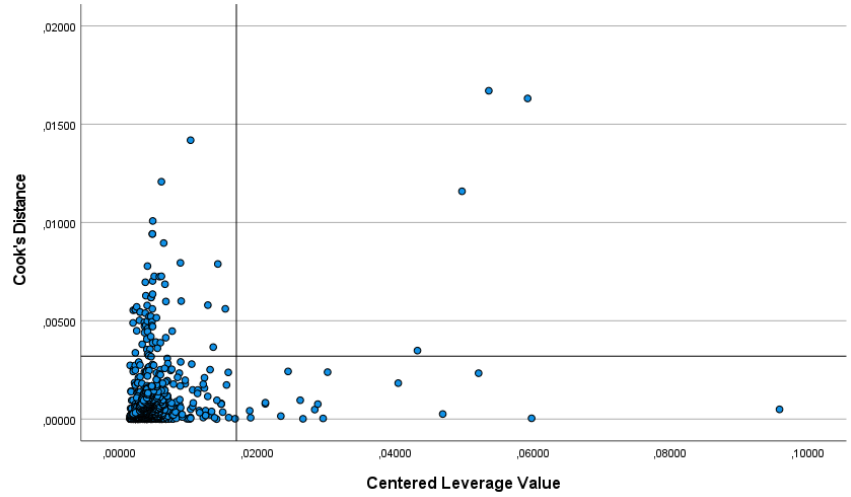
Case ID	Standardized residuals
822866	-3,689
805722	-3,511
840057	-3,511
835379	-3,507
857302	-3,482
825573	-3,469
891183	-3,448
811886	-3,406
896826	-3,406
848199	-3,316
866149	-3,242
829696	-3,140
845852	-3,140
835421	-3,059
856689	-3,054
846099	-3,054

The cases that have a significant influence on the regression in the x-direction can be examined through their leverage scores. The leverage shows how much a single case influences the fit of a regression model. A common rule of thumb identifies a case as an outlier when $h_c > \frac{3p}{n}$, with P representing the amount of parameters in the model and N the sample size. There are 7 parameters in the final model with a sample size of 1232 respondents, therefore $h_c > 0.017$. Cases that have a leverage score higher than 0.017 are considered an outlier.

Furthermore, the Cook's distance calculates the product of the standardized residuals and the leverage. By doing so, it presents influential cases in both the x- and y-direction. A common rule of thumb tells that a case is possibly problematic for the estimated regression coefficients when $CD_i > \frac{4}{n}$, with N representing the sample size. This model includes a sample size of 1232 respondents, therefore $CD_i > 0.0032$.

```
GRAPH
  /SCATTERPLOT(BIVAR)=LEV_1 WITH COO_1
```

```
/MISSING=LISTWISE.
```



To end with, the DFFIT value shows the change of the predicted standardized y when leaving out a single case from the regression. A common rule of thumb for this measure is $DFFIT > \left|3\sqrt{\frac{p}{n}}\right|$. P stands for the amount of parameters and n is the sample size. For this model, 7 parameters and a sample size of 1232 respondents results in $DFFIT > 0.226$. This procedure leads to no cases that cross the DFFIT value of 0.226.

A short list of cases has been found that might be of significant influence on the regression coefficients. These cases might change the outcomes of the analyses and could therefore change certain conclusions. A new regression is done to examine the influence of these cases by leaving them out of the regression. Doing so, 84 cases have been filtered.

The figures below show that leaving the outliers out of the sample does not make a difference in the conclusions that are drawn from the final regression. Therefore, the 84 outliers that were removed can still be included in the sample. Also, the outliers are not systematic errors. They are valid cases that do not need to be excluded from the sample. The fact that there are influential cases in the sample still needs to be addressed in the results.

```
IF  (filter_$) filter_outliers=LEV_1 < 0.017 & COO_1 < 0.0032.
EXECUTE.


REGRESSION
```

```
/SELECT=filter_outliers EQ 1
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA CHANGE
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT Soc_Integ_Imm
/METHOD=ENTER Education Group_of_origin Primary_occupation sporting Gender
SportingxGender
/SCATTERPLOT=(*ZRESID ,*ZPRED)
/RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID)
/SAVE PRED ZPRED COOK LEVER RESID ZRESID DFFIT.
```

filter_outliers

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	84	6,8	6,8	6,8
	1,00	1148	93,2	93,2	100,0
	Total	1232	100,0	100,0	

ANOVA[a,b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	389,786	6	64,964	15,475	<,001[c]
	Residual	4790,043	1141	4,198		
	Total	5179,829	1147			

a. Dependent Variable: Soc_Integ_Imm

b. Selecting only cases for which filter_outliers = 1,00

c. Predictors: (Constant), SportingxGender, Education, Group_of_origin, Gender, sporting, Primary_occupation

Model Summary[b,c]

	R					Change Statistics				
Model	filter_outliers = 1,00 (Selected)	filter_outliers ~= 1,00 (Unselected)	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	,274[a]	,798	,075	,070	2,04893	,075	15,475	6	1141	<,001

a. Predictors: (Constant), SportingxGender, Education, Group_of_origin, Gender, sporting, Primary_occupation

b. Unless noted otherwise, statistics are based only on cases for which filter_outliers = 1,00.

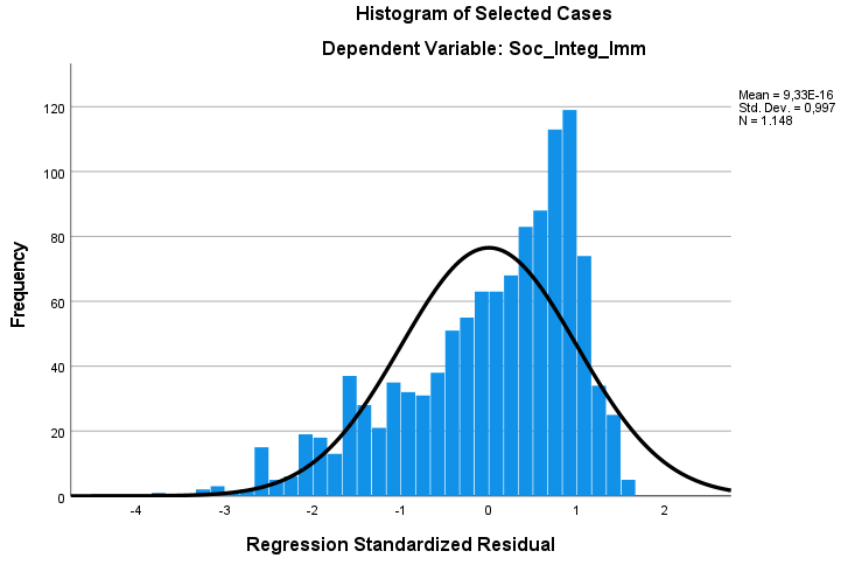
c. Dependent Variable: Soc_Integ_Imm

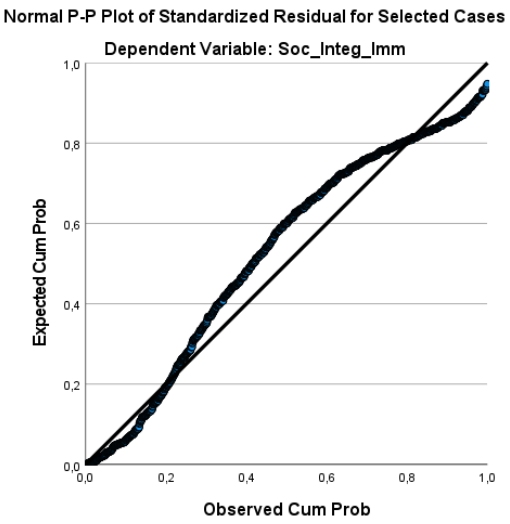
Coefficients[a,b]

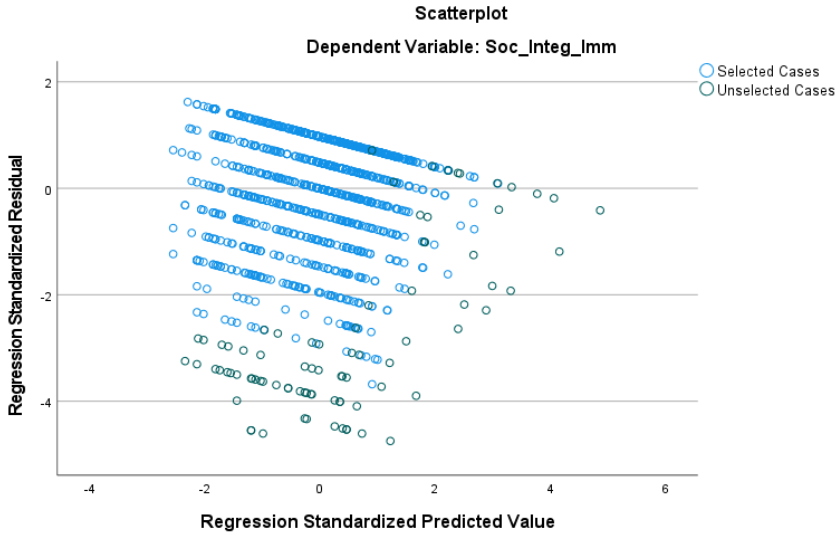
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	9,404	,194		48,472	<,001
	Education	,120	,038	,091	3,145	,002
	Group_of_origin	-,922	,129	-,206	-7,161	<,001
	Primary_occupation	,163	,131	,036	1,242	,214
	sporting	,111	,025	,130	4,450	<,001
	Gender	,275	,123	,065	2,240	,025
	SportingxGender	,040	,049	,024	,818	,414

a. Dependent Variable: Soc_Integ_Imm

b. Selecting only cases for which filter_outliers = 1,00




Normal P-P Plot of Standardized Residual for Selected Cases
Dependent Variable: Soc_Integ_Imm
Expected Cum Prob
Observed Cum Prob
0,0
0,2
0,4
0,6
0,8
1,0


Scatterplot
Dependent Variable: Soc_Integ_Imm
Selected Cases
Unselected Cases
Regression Standardized Residual
Regression Standardized Predicted Value
2
0
-2
-4
-4
-2
0
2
4
6

Appendix 4 - Use of AI software

During the revision of this thesis, ChatGPT and Grammarly have been used to increase the quality of this work. I will elaborate on my decisions for using this AI software and show some examples of how I have used the suggestions of ChatGPT and Grammarly.

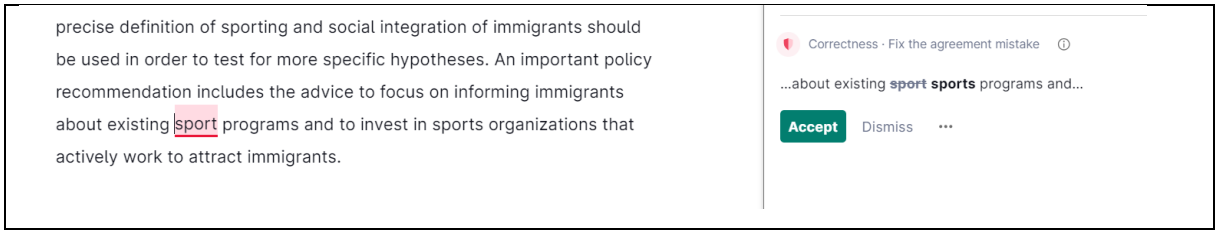
ChatGPT has been used to rewrite some sentences and to check for any writing mistakes. I believe it is important for an academic text to be clear and to have variety in its sentences. After I had written a paragraph I used ChatGPT to suggest how I could rewrite some sentences. Some of my sentences lacked a clear structure or needed to be split into two sentences. Sometimes ChatGPT gave suggestions that did not fit my criteria of being consistent with specific terminology. Therefore, I only used synonyms for words or got inspired to change the structure of my sentence after receiving suggestions from ChatGPT. Below I have presented an example of what suggestion ChatGPT gave when I asked to rewrite a sentence.

Example 1

Rewrite this sentence: Due to the presence of different migration backgrounds, efforts have been made in the research to include representative group sizes of respondents with different migration backgrounds.

ChatGPT: *Given the diversity of migration backgrounds, the research has made efforts to include representative group sizes for respondents from various migration backgrounds.*

Grammarly has been used to check for writing mistakes and sometimes to change a sentence. I received feedback that my texts contain spelling mistakes and that my writing style sometimes is grammatically incorrect. I therefore used Grammarly to check for such mistakes. Below I have given an example of what Grammarly has suggested to me.



I have made efforts to check for the reliability of the generated suggestions by ChatGPT or Grammarly. Usually, I check synonyms online on whether it fits the message that I want to write in my paragraph. Besides that, I tried to be critical of the terminology that both AI software programs

generated for the concepts that I covered in my thesis. Furthermore, I have searched for the grammatical writing rules for British English in moments when I was not sure if Grammarly or ChatGPT gave correct suggestions. By doing so, I have made efforts to check the reliability of the information that was generated by ChatGPT and Grammarly.

In order to secure my academic integrity I have paid attention to make clear references to the literature that I have used and to use my own words when citing work from other authors. I have made an effort to check my literature on whether I have cited all sources correctly.

Writing this thesis using AI software has helped me in my learning process. I got new ideas on how to make my texts more interesting to read and how to avoid having a passive writing style. Moreover, I learned about referencing different sources and especially writing English academic texts. I occasionally looked up the rules for citing literature that does not originate from English sources. For example, I used some Dutch sources that needed to be translated into English. On the whole, using AI software has given me more insight into my writing style and has helped me to develop a more critical view of how I could improve my academic texts.