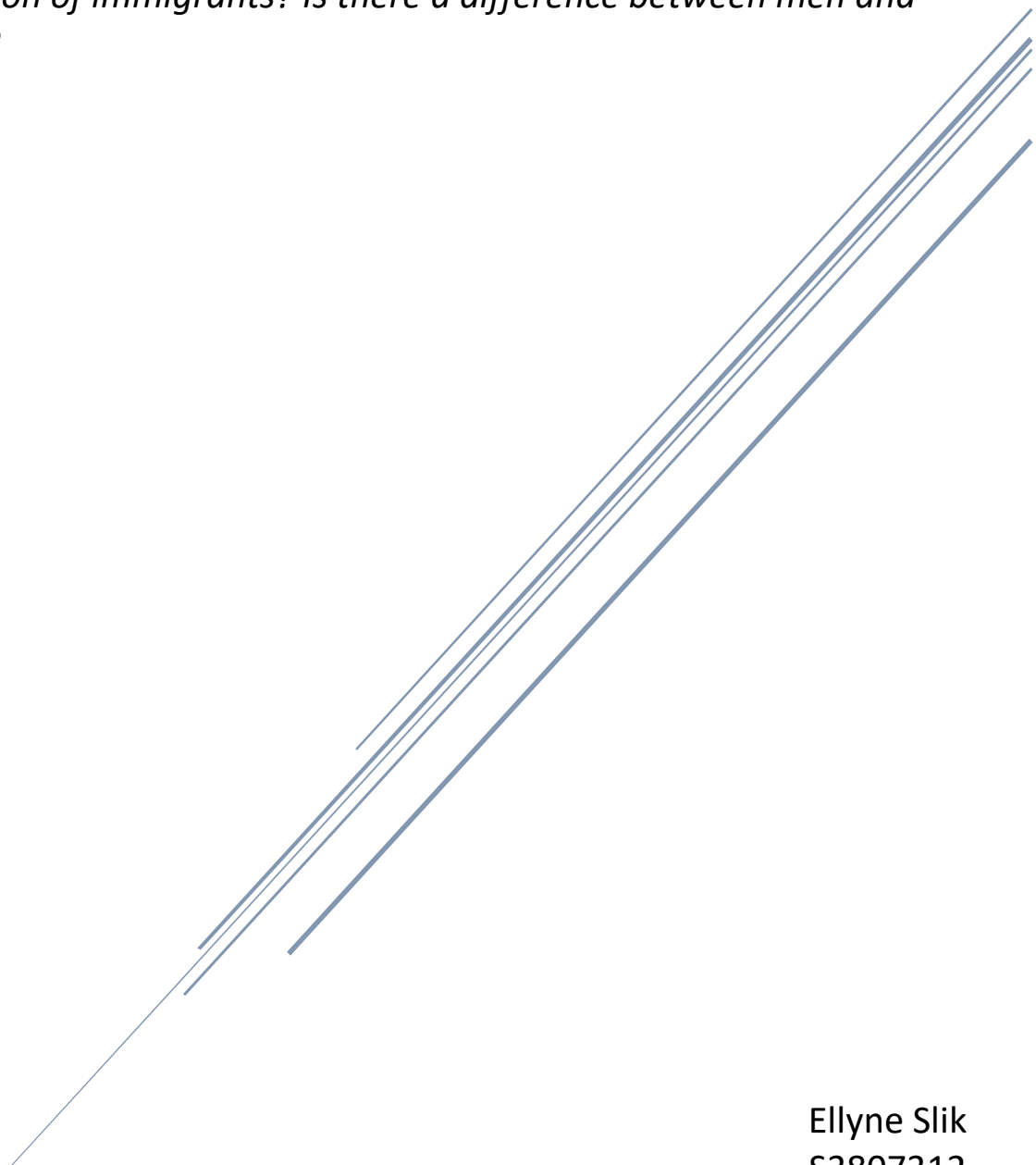
HOW THE DIVERSE EDUCATIONAL LEVELS OF SOCIAL CONTACTS HELP SOCIAL INTEGRATION

What is the effect of network diversity (of the five closest contacts mentioned in the survey) in terms of education on the social integration of immigrants? Is there a difference between men and women?



university of groningen
faculty of behavioural and social sciences

Ellyne Slik
S3807312
e.c.slik@student.rug.nl
Francesca Giardini & Gert Stulp
Resit
24-08-2022

Abstract

The social integration of immigrants is an important phenomenon, as it can cause quite some difficulties for both the immigrants and the host countries. Also, in the Netherlands, problems with cohesion, crime, and discrimination can be linked to the social integration of immigrants. A lot of solutions are being sought to make the process of social integration for immigrants more effective. In this research, we look if having diverse social contacts with diverse educational levels will help the social integration of immigrants. On top of that, we will investigate if this relationship is different for men or women. The dataset of the LISS immigrant panel was used, which is a longitudinal dataset of respondents living in the Netherlands. Ultimately 640 respondents were used in this research, all immigrants from Turkey, Morocco, Netherlands Antilles, Surinam, Indonesia, and other Western origins. The results of this research show evidence for the relationship between network diversity in terms of education on the social integration of immigrants, however, it does not show evidence of the effect of gender on this relationship.

Table of Contents

1. Introduction

In the last decades, migration left its mark on the citizen's daily concerns in the Netherlands (Huijnk & Andriessen, 2016). With the fear of unemployment and a disbalance in cohesion while more immigrants arrive every year. It is now become one of the key debates during the national elections. As of 2021, 2.5 million citizens were not born in the Netherlands, where Turkey, Surinam, and Morocco formed the most prominent groups of immigrants living in the Netherlands (Centraal Bureau Voor Statistiek, 2022). This is one of the reasons why the Netherlands is called a 'migration country' (Cuyvers, 2020). Although most immigrants seem to integrate well in terms of work and income, still some difficulties occur regarding crime and discrimination (Huijnk & Andriessen, 2016; Cuyvers, 2020). One step for immigrants to overcome these difficulties is to establish a social network around them. This is beneficial for the immigrants as social contacts are crucial for their happiness (Arpino & de Valk, 2018). As well as it is beneficial for the Dutch people. Some Dutch people may refuse to favor integration, but when they come in contact with immigrants, they might get used to their different standards (Schlueter & Scheepers, 2010). All in all, it is essential to work on the social integration of immigrants in the Netherlands, mainly because social integration brings social cohesion, a strong institutional foundation, and a culture of acceptance (Cruz-Saco, 2008).

Over the years, many studies showed concerns about the social integration of immigrants. Several studies agree that one of the most important aspects of social integration is the social networks surrounding immigrants. The literature mentions the importance of migrant networks, as they affect the local- and national-level economies and are crucial for immigrants to make migration possible (Poros, 2011). Moreover, differentiated social relations could lead to a better quality of social capital (Kindler, Ratcheva & Piechowska, 2015), which is essential for the social integration process. Another effective mean to obtain social integration is through education. The literature shows that more educated people tend to relate with the community more intensively than less educated

people, improving their position in society (Depalo, Fiani & Venturini, 2006). Moreover, education is also an essential tool for learning the native language of the host country (Nawyn, Gjokai, Agbényiga & Grace, 2012). On top of that, is the educational levels of the people in the social network of the immigrants. As they can also influence the interests, habits, and use of talents the immigrants have. The research on this subject is however limited. Another important aspect is the potential differences between men and women important when researching the relationship between network diversity in terms of education and social integration. Earlier research shows mixed feelings about this topic. As some studies argue that women are less socially integrated, because of traditional task distribution, work opportunities, and the way they cope with social contacts (Dalgard & Thapa, 2007; Bilecen & Seibel, 2021). While others show that men are less socially integrated, due to the higher adjusting abilities of women (Avenarius, 2012).

Although many studies show the importance of social networks on the social integration of immigrants, the literature lacks a view on the specific subject in social network analysis, which is network diversity. Where most studies assume having social contacts is essential, I argue that looking at the differences in social contacts is essential. These social contacts have very different traits, such as other educational levels, which may influence the immigrant's social integration process.

In this research, we shall focus on immigrants living in the Netherlands. It aims to study the feeling of social integration among immigrants. By doing this, we will look at the immigrants' social contacts and what educational level they carry, in other words, the network diversity in terms of education. Moreover, we hope to clarify the effect of gender on the relationship between network diversity in terms of education and the feeling of social integration. Therefore, the following research question was formed: *What is the effect of network diversity (of the five closest contacts mentioned in the survey) in terms of education on the social integration of immigrants? Is there a difference between men and women?*

2. Theory

In this chapter, we will first give further inside on the terms social integration and network diversity in terms of education by providing definitions and further explanations. Furthermore, we will discuss the relationship between network diversity in terms of education and social integration and the effect of gender on this relationship. After that, we will explain possible external factors we will use as control variables in the research. Lastly, we will visualize these relationships in a research model.

Social integration of immigrants

Earlier, we stated the importance of social integration for both the immigrants and the Dutch people. Social integration is also an essential aspect used in sociology and has many different definitions. It can refer to immigrants feeling a part of the society they are currently living in and need to accept and follow that country's social values and norms (Laurentsyeva & Venturini, 2017). On top of that, social integration refers to the quantity and quality of the social interactions of immigrants (Rubin, Watt & Ramelli, 2012). Although these definitions are fundamental, we will focus mainly on life satisfaction and loneliness in this research. According to Adams and Serpe (2000), social integration can indirectly positively affect life satisfaction. Moreover, as long as people have meaningful social contacts, they experience less loneliness (Stevens & Westerhof, 2006). This crosses over the first and second definitions given in this paragraph. Social integration is linked to social loneliness, and so it is essential to feel that a person can participate in a group and be accepted as a group member (Russell, Cutrona Rose & Yurko, 1984). This also shows the importance of social integration, as it is an essential part of the life satisfaction of immigrants. The social integration process can be different for everyone, as external factors also impact this process. Such as the difference between first and second-generation immigrants.

In this research, we focus on first and second-generation immigrants. First-generation immigrants moved from a foreign country to the Netherlands and were not born in the

Netherlands. They expected to temporarily stay in the country as they moved to the Netherlands for working purposes. Second-generation immigrants are children of first-generation immigrants and are indeed born in the Netherlands. So they spent their whole life here and followed Dutch education. In this case, second-generation immigrants are more familiar with the Dutch culture and language (Algan, Dustmann, Glitz & Manning, 2010). In this research, we do not distinguish between the two generations.

Network diversity in terms of education

An essential part for immigrants in the social integration process is to gain social contacts (Arpino & de Valk, 2018). So in this research, we will look at individuals' social networks and how this will benefit the social integration of immigrants. In social network analysis, we investigate the patterning of relations among social actors at different levels (Breiger, 2004). This is interesting because the position of an individual will show the constraints and opportunities, and therefore we can predict outcomes such as beliefs and behavior (Everett, Borgatti & Johnson, 2018). In social network analysis, we work with the terms 'ego' and 'alters'. Where the ego represents the individual whom the social network is built around, and the alters are the ego's relationships (Everett et al., 2018). We will use these terms in the following paragraphs when describing social networks.

We established that social contacts are important for the social integration process (Arpino & de Valk, 2018). On top of that, we will argue that not just having social contacts helps, but most importantly, what kind of social contacts. Specifically, whether having diverse social contact will help the social integration process. So in this research, we will focus on network diversity, generally defined as being socially connected with people of different backgrounds (Pachuki, 2020). The focus on these different backgrounds can change with each research. We will specifically look at network diversity in terms of education. This means that the ego is socially connected with people with different educational levels. We will look at how having contact with alters with different levels of education affects the ego's social integration. Researchers have not yet focused on whether having diverse social

contacts helps social integration. However, in the next paragraph, we argue that this is the case.

Relationship between network diversity in terms of education and the social integration of immigrants

The relationship between network diversity in terms of education and the social integration of immigrants can be mainly explained through the social capital theory and with that the bridging, bonding, and human capital theories (Kindler et al., 2015; Chiswick & Miller, 2009). These work closely together but do have their differences. Firstly we will elaborate on these theories and their definitions; after that, we will elaborate more on the relationship between network diversity in terms of education and the social integration of immigrants.

Many sociologists joined the discussion on the social capital theory; the three most known sociologists are Putnam, Coleman, and Bourdieu. Bourdieu (1986, page 249) explained social capital as ‘the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance or recognition’ (Kindler et al., 2015, page 5). Putnam (2007) adds the concept of social network and the norms of reciprocity and trustworthiness as essential for social capital (Kindler et al., 2015). In this research, we will mostly hold onto the definition of Coleman (1990, page 305); he describes social capital as ‘resources that can be used by the actor to realize their interests’ (Kindler et al., 2015, page 5). Social capital implies that attaining goals works better for people who are well equipped with social resources, like strong relationships. Additionally, people will invest in relationships with others because of the expected gained resources made by these relationships. The stronger these relationships, the more likely the sharing and exchanging of resources (Lancee, 2010). Thus, immigrants can use these resources to realize their interests (Kindler et al., 2015).

There are several important aspects of social capital when explaining the relationship between network diversity in terms of education and the social integration of immigrants. As

Dahinden (2013) describes, social capital shows the importance of resources present in networks. When looking at social networks, the concepts of bonding and bridging capital play a significant role (Kindler et al., 2015; Arpino & de Valk, 2018). Bonding capital corresponds to the idea of homophily, which refers to selection. People, therefore, tend to choose people in their close network who are similar to them (McPherson, Smith-Lovin & Cook, 2001; Kindler et al., 2015). Furthermore, having cohesive networks can have an advantage in maintaining one's resources (Kindler et al., 2015). However, in this research, we try to find the impact of having diverse social contacts and focus more on bridging capital. Having diverse networks can form bridges among social contacts and may help obtain new resources (Pachuki, 2020; Kindler et al., 2015). For example, through these bridges, we can contact different friends from different networks. When bringing them together, these will form another bridge, and will also come in contact with other resources, like ideas, habits, and even other social contacts.

The relationship between network diversity in terms of education and the social integration of immigrants can be explained through these concepts of social capital. We established in the earlier paragraphs that the more diverse social network an ego has, the better the quality of social capital (Dahinden, 2013). By having a diverse social network immigrants can use bridging capital to build bridges in the social network in various ways (Pachuki, 2020). These bridges can lead to more opportunities for attaining social resources, and the diversity of these social resources will also increase (Kanas, Chiswick, Van Der Lippe & Van Tubergen, 2012; Lin, 2008). Especially when having social contacts with different educational levels. The skills, knowledge, and experiences, described as human capital (Kucharčíková, 2011), attained by alters through their different educational levels can help the social integration process of the ego (Chiswick & Miller, 2009). As it gives them more opportunities in the host country, which is an essential part of the social integration process for immigrants. It can teach them skills they need, but most importantly, they can show and use their skills and knowledge and attain more social contacts.

Although bonding and bridging capital are two different concepts, it is important to distinguish between them. They do, however, follow up on each other in some situations. Bonding ties are strong ties with mutual trust like families or close friends, they can lead to opportunities to come in contact with bridging ties, which are weak ties and mostly people with different backgrounds (Kindler et al., 2015). By hanging out with siblings or close friends, you will most likely meet people who are not a strong tie to you but are a strong tie to your friend. When meeting these people, bridges are created with weak ties and people who will most likely have different background characteristics than you. In this research, we focus on the five closest contacts of the egos. These five closest contacts are not necessarily bridging ties, but as we argued above, bonding ties can also indirectly lead to meeting new weak ties. Both bonding and bridging capital are in this way important for the social integration process of immigrants, mainly because if immigrants fail to gain ties they are prone to loneliness and lower levels of life satisfaction (Arpino & de Valk, 2018; Pachuki, 2020). As loneliness and life satisfaction are important aspects of the social integration of an immigrant, it can be expected that social capital can influence the feeling of social integration (Kindler et al., 2015).

We also need to consider that five close contacts are not a lot and that the origin of the social contacts can impact social integration. Having more contacts of Dutch origin will most likely be more beneficial for the immigrant social integration process (Kanas et al., 2012). Also, more contacts can lead to more social integration as social contacts are an essential aspect of the social integration process of immigrants.

In brief, the relationship between network diversity in terms of education and the social integration of immigrants can be explained through mainly the social capital theory. It shows that having diverse contacts can lead to a diverse set of means, skills, and knowledge. All these different views of different educational levels, and the social contacts gained through bonding and bonding capital, lead to a higher feeling of life satisfaction and a lower sense of loneliness. All in all, network diversity in terms of education will positively

affect the social integration of immigrants. Based on the arguments given above, the following hypothesis was formed:

Hypothesis 1: Having a more diverse network in terms of education will positively impact the social integration of immigrants.

Gender

Besides the relationship between network diversity in terms of education and the social integration of immigrants, we are also interested in the effect of gender on this relationship.

Some researchers believe that women are better socially integrated. As Avenarius (2012) states that women become well-adjusted to life in another country while men have difficulty with the traditional male authority, specifically with their attitudes towards family, marriage, and gender roles like women's participation in the workplace. To add, women can often overcome social isolation and become more influential than men in society (Avenarius, 2012). It should be noted that this data comes from research focused on Taiwan, so it may not be applicable to all immigrants in the Netherlands but is nevertheless an interesting point to mention.

However, the literature also shows various explanations as to why it might be the other way around. Firstly, the traditional task distribution between men and women plays a significant role. Women are still designated to their household responsibilities and are less likely to spend their time working. This makes it difficult for women to participate in social life and cause social isolation (Dalgard & Thapa, 2007). As a workplace provides more opportunities for meeting others and creating these needed personal ties and so on coming in contact with the opportunities of social resources needed for social integration (Bilecen & Seibel, 2021). On top of that, women are still pushed into more traditional gender roles by society. They are often more strongly monitored by family members (Röder & Mühlau 2014). This makes it more difficult for immigrant women to meet other people than for immigrant men (Bilecen & Seibel, 2021). Thirdly, many immigrant families tend to invest more in the

human capital of a men than in that of a women (Bilecen & Seibel, 2021). For example by offering more help with finding a job and finding resources to improve the Dutch language (van Tubergen & Kalmijn 2008). Lastly, women tend to spend more time with strong ties than men (Inglehart & Norris, 2003). As they also prefer a smaller network characterized by high levels of trust (Burt, 1998). All in all, it is hard for women to break gender and work stereotypes when having immigrated (Riaño, Baghadi & Wastl-Walter, 2006). As the women are seen as '[…] poorly integrated, as uneducated, and as victims […]" (Riaño et al., 2006, page 6). Which will make the position of women go into a vicious circle. In this way, based on the arguments in this paragraph social integration is more complex for women because of their lack of social contacts.

With both these stances in mind, we believe that the stance of women having a harder time at social integration comes on top. That is why the following hypothesis is formed:

Hypothesis 2: Men show a stronger effect when looking at the relationship between network diversity in terms of education and the social integration of immigrants than women do.

Control variables

There could also be some other external factors that can influence the relationship between network diversity in terms of education and the social integration of immigrants. Based on scientific insights, three factors are included in the analysis as control variables.

The first control variable included in the analysis is income. Income can provide more means and opportunities to engage in social participation and educational leisure activities (Bult, Verschuren, Jongsmans, Lindemans & Ketelaar, 2011; Kanas et al., 2012). For example, get more classes to improve their Dutch language skills or participate in other formal activities. This will give immigrants with a higher income more chances to gain social contacts as they have the means to search for these contacts. On top of that, having a partner and frequent contact with family, friends, and neighbors increases subsequent

income (Kanas et al., 2012). In this way, income can play a significant role in the social integration process of immigrants.

Secondly, primary occupation gives immigrants opportunities to have more contact with Dutch speakers, as they have more contacts when working than immigrants who are not working (Valenta, 2008; Kanas et al., 2012). On top of that, it will be easier to gain social connections and so have other influences, as you don't have to look for them. These contacts will be beneficial for the social integration of immigrants and can significantly differentiate between immigrants, especially between men and women (Riaño et al., 2006).

Thirdly, we will control for age since younger immigrants are more likely to have more education because they are obliged to do so in the Netherlands. In this sense, they will pick up the Dutch language faster than the older immigrants, which means they have more opportunities for interaction with natives (Martinovic et al., 2009). On top of that, research shows that elderly immigrants show less social integration as they have fewer social contacts (Tselios, Nobavk, van Dijk & McCann, 2015).

Research model

In this chapter, the theory showed a possible positive effect of network diversity in terms of education (of the five closest contacts mentioned in the survey) on the social integration of immigrants. Furthermore, that men most likely have a higher effect on the relationship between network diversity (of the five closest contacts mentioned in the survey) in terms of education and the social integration of immigrants when gender is added as a moderation effect. Lastly, we explained how alternative solutions would be excluded from the analysis. In the research model below, these relationships are visualized:

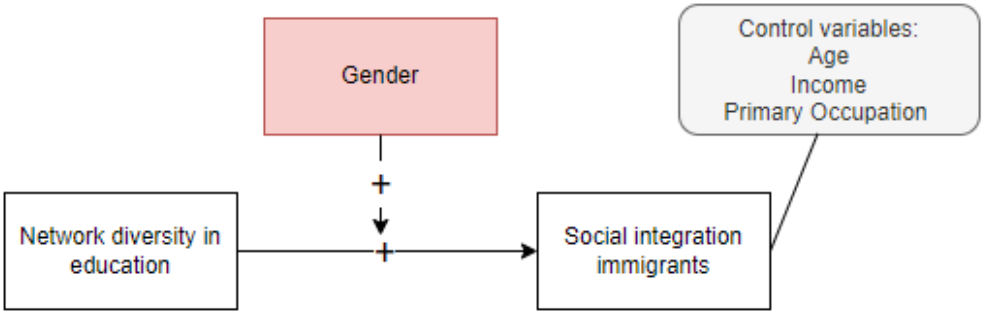


Figure 1: Research Model

3. Methods

3.1 Description of the participants

In this thesis, we use the LISS migrant panel dataset. This research was carried out in February 2014 about social integration and spending of leisure time among immigrants of the Netherlands. A variety of information was collected about the respondents who participated. All to track down their social integration process, leisure time activities, and well-being.

The primary sampling of the respondents was based on individuals, but when a selected person agreed to participate, the household was included. The respondents could participate via an internet survey where a reminder was sent twice to non-respondents (Mulder, 2014). When respondents did not have access to a computer and Internet, they received simPCs and broadband Internet. When they participated, respondents received incentives of 15 euros per hour (Centerdata, 2014).

The LISS migrant panel includes two waves. The first wave was carried out in 2011; in this research, we use the questionnaire part of the second wave of the longitudinal survey. Overall, 1748 households were selected as panel member; of these household members, there was a nonresponse of 444 and a response of 1304. Out of these responses, 1270 were complete, and 34 were incomplete. The respondents of the dataset originate from different nationalities, which are The Netherlands (32,3%), Turkey (5,6%), Morocco (5,8%), Netherlands Antilles (4,3%), Suriname (5,0%), Indonesian-Dutch (8,6%), other Western origins (25,4%), and other non-western origins (12,8%). All non-Dutch respondents are from the first or second generation. Since this research focuses on immigrants, all Dutch respondents were removed from the dataset through the variable ‘country of origin’. This led to a population of 1080 respondents originating from Turkey, Morocco, Netherlands Antilles, Suriname, Indonesia, other Western origins, and other non-western origins. When constructing the variables used in this research, certain variables seemed to have many missing values as participants did not respond to this question. Especially the newly made

variable 'network diversity in terms of education' and the variables 'income' and 'primary occupation'. We removed these missing values from the dataset to complete the analysis correctly. This resulted in a population of 640 respondents.

In addition to removing the missing values, a missing values analysis was done beforehand. This analysis illustrates that the extracted missing values were respondents with an average lower age and income. This means that when working with the 640 respondents, we must consider that the removed respondents made the average age and income in the dataset lower. The average age without missing data is 47,36 years old, and with the missing is 38,26. The average income without the missing data is 1548,52, and with the missing data 1246,67. Further information on the analysis of the missing values is shown in appendix 1.

3.2 Research Design

Respondents were asked how they spent their leisure time and how satisfied they were with this time spent. Following up, there are many leisure activities mentioned in the survey in which respondents could participate. These activities could also include social participation activities, like voluntary work. On top of that, respondents were asked about their time spent with social contacts. One of the aspects necessary for this research are the questions if the respondents are satisfied with their social contacts and if they feel a sense of life satisfaction or loneliness. Lastly, personal questions about respondents' life satisfaction, primary occupation, and domestic situation were asked, among other things.

3.3 Operationalisation

In this research, six variables are used in the analyses, all operationalized differently. The following paragraphs describe how these variables are operationalized and which variables will ultimately be used in the analysis. The dependent variable 'social integration of immigrants' and the independent variable 'network diversity in terms of education' are made out of a set of items used in the dataset. The moderation variable 'gender' and the control

variable ‘primary occupation’ were recoded. The variables ‘income’ and ‘age’ were already set to use in the analysis.

3.3.1 Network diversity in terms of education

In our theory, we stated that network diversity in terms of education looks at the diversity of educational levels of the alters of the ego. So to construct the variable for the independent variable network diversity in terms of education, multiple questions from the questionnaire needed to be combined. The respondents were asked: “What is the highest level of education person 1 completed?”. This question was asked five times to gain five social contacts of every respondent. The respondents could answer these questions with scores (1) not (yet) completed any educational program; (2) primary school; (3) VMBO, LBO, MULO, ULO, MAVO; (4) HAVO/VWO; (5) MBO; (6) HBO; (7) university; (8) I don’t know. Firstly we excluded the last category (8) I don’t know. From there on, we looked at the possibilities to construct the variable network diversity in terms of education to show the diverse networks of the egos. Consequently, we decided to construct a maximum and a minimum variable first. In that way, we could subtract the maximum from the minimum, which shows us how diverse our ego’s network is when considering the educational levels of the alters. When the ego has a higher score on this new variable, it indicates a higher diversity in their network in terms of education. The score of the variable network diversity in terms of education varies from 0 to 6, where 0 means a low score of network diversity, and 6 means a highly diverse network of the ego in terms of education.

We should also note that network diversity in terms of education is measured with only the five closest contacts the respondent named in the survey. On top of that, most respondents did not include all five closest contacts. As most only named two or three of their closest contacts. This did not create problems with the making of the variable but can give another view as the contacts used are limited.

3.3.2 Social integration of immigrants

In our dataset, no original variable for social integration was constructed. So with our theory in mind and the other variables available in the LISS migrant panel, we needed to find the ones who would describe the social integration of immigrants the best. As our theory paragraph stated, the aspects of social contact, life satisfaction, and loneliness are significant in the definition of social integration.

When looking at the variables representing these concepts, we found several life satisfaction and loneliness variables. These are the following six variables: 'I have a sense of emptiness around me', 'I miss having people around me', 'I often feel deserted', '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' and 'there are enough people to whom I feel closely connected'. Respondents could answer these questions with (1) yes; (2) more or less; or (3) no. To have all questions in the same direction, the last three questions were mirrored, where (1) no; (2) more or less; (3) yes. In the LISS migrant panel dataset, there is also a question describing the important aspect of social contact in the definition of social integration: ‘How satisfied are you with your social contacts?’ The scale of this question goes from 0 to 10, with 0 meaning not at all satisfied and 10 meaning completely satisfied. When looking at the scale of the earlier used questions, we can conclude that combining these will not work. So we decided to only work with the earlier mentioned six variables. Subsequently, we formed a mean variable by these six questions. The score of this variable varies from 1 to 3. This score indicates, that the higher the score, the higher the feeling of being socially integrated the immigrant has.

3.3.3 Gender

For the moderation effect gender, respondents were asked in the variable ‘geslacht’ about their gender. They could use the following answer categories (1) male and (2) female. This variable was recoded to (0) female and (1) male and renamed ‘gender’.

3.3.4 Age and income

The respondents' age in this questionnaire varies from 16 to 88 years old. It is a continuous variable and is used as a control variable in the analysis of this research. Income is a continuous variable and is also used as a control variable in the analysis of this research. This variable shows a wide range in incomes from 0 to 68388 euro's a month. We recoded this variable by dividing the original score by 1000. In this way, we can have a more effortless look, which makes the interpretation of the analysis easier.

3.3.5 Primary Occupation

The last control variable used in the analysis of this research is the variable primary occupation. This is the categorical variable 'belbezig', with many answer categories: (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; and (13) does something else. Due to many answer options for 'belbezig', a recoding occurred, resulting in a new dummy variable with the categories: (0) employed and (1) unemployed, which was also renamed to Primary_occupation. Under employed, we took the answer categories 1,2,3,11, and 12. We included 11 and 12 because you will have more access to social contacts even when doing unpaid work. The answer categories 4 to 10 are unemployed, as they have no occupation. This recoding was made because we are merely interested in if the respondents have an occupation or not, and not really in what kind of occupation the respondent has. The answer category 13 was excluded since it did not have enough information to know where it needed to be fitted in.

3.3.6 Interaction variables

The models have added interaction variables to test the possible moderation effects. These interaction variables consist of the product between network diversity in terms of education

and the moderator gender. As the moderator gender is dichotomized, there was no need to center the moderator. We did, however, center, the independent variable network diversity in terms of education.

3.4 Analysis Plan

A linear regression will be performed to answer the research question and test the given hypotheses. The following models will be estimated in the analysis. We start with an empty model with only the dependent variable 'social integration of immigrants' and the control variables 'age', 'income' and 'primary occupation'.

In model 2, the independent variable 'network diversity in terms of education' will be added. In this model, a linear regression will be used to determine whether having a diverse network in terms of the educational level of an immigrant's social contacts leads to a higher sense of social integration among immigrants.

In model 3, the moderation effect 'gender' will be added to the linear regression analysis described in model 2. Here we will examine whether gender affects the feeling of being socially integrated of the immigrants. It is discussed whether this effect applies more to men or women.

In model 4, we will examine whether gender affects the relationship between network diversity in terms of education level and the sense of social integration in immigrants. We do this by adding the interaction effect of the independent variable network diversity in terms of education level and the moderation effect of gender. If this interaction is significant, the moderator will indeed have an effect.

4. Results

4.1 Descriptive statistics

4.1.1 Univariate statistics

Firstly we will explore the data by looking at the descriptive statistics of the variables used in the analysis. For the continuous variables, the mean, standard deviation, minimum and maximum, median, and percentiles are shown. For the categorical and dummy variables, the proportions are shown. These values are shown in table 1.

Table 1: Descriptive statistics of the named variables in the analysis.

Variables		N	Mean (SD)	Minimum	Maximum	1st Quartile	Median	3rd Quartile
Age		637	47,76 (16,615)	16,000	88,000	36,000	48,000	61,000
Income		637	1,468 (1,116)	0,000	10,00	0,750	1,350	2,000
Network Diversity in terms of education		637	1,955 (1,504)	0,000	6,000	1,000	2,000	3,000
Social Integration		637	2,601 (0,442)	1,000	3,000	2,333	2,667	3,000
Gender								
Female = 0	Female	376 (59,0%)						
Male = 1	Male	261 (41,0%)						
Primary Occupation								
Employed = 0	Employed	341 (53,5%)						
Unemployed = 1	Unemployed	296 (46,5%)						

Table 1 shows us some interesting aspects of the descriptive statistics of the variables used in the analysis. The mean of network diversity in terms of education is 1,955, this is relatively low when considering that the scores have a scale from 0 to 6. This means that most respondents do not have a very diverse network. This is also shown in the median with a score of 2. Table 1 shows that the mean of the variable social integration has a score

of 2,601. This is a high score when looking at the minimum and maximum scores of this variable. This suggests that in this dataset, most respondents do feel socially integrated. Moreover, table 1 shows more females (59,0%) than males (41,0%) in this dataset.

Furthermore, the average age of the respondents is 47,76 years old. There is quite some scatter in the variable age, shown through the standard deviation (SD= 16,615) and the minimum (Min= 16,000) and maximum (Max = 88,000). Income has a low average of 1,468 with a minimum of 0,00 and a maximum of 10,00. The standard deviation (1,116) also indicates a wide range of incomes. Lastly, primary occupation shows that most respondents are employed (53,5%) in comparison to the 46,5% who are unemployed.

4.1.2 Bivariate statistics

The bivariate analysis looks at the interrelationship between the variables used in the study. We used Pearson's correlation for the correlation between two continuous variables. For the relationship between two nominal variables, we measured both the chi-square test and Cramer's V. Table 2 shows these bivariate statistics.

Table 2: Bivariate statistics of the named variables in the analysis.

	Gender	**Age**	**Income**	**Primary Occupation**	**Network diversity**	**Social integration**
Gender	-					
Age	0,043[a]	-				
Income	0,277**[a]	0,267**[a]	-			
Primary Occupation	1,029[b] 0,040[c]	0,153**[a]	-,393**[a]	-		
Network Diversity	-0,085*[a]	-0,020[a]	-0,006[a]	-0,030[a]	-	
Social integration	0,031[a]	-0,025[a]	0,123**[a]	-0,074[a]	0,105*[a]	-

[a]Pearson's correlation, [b]Chi-square test, [c]Cramer's V.
*significant when p<0,05, **significant when p<0,01; two-sided test; N = 637.

Table 2 shows the correlation between the used variables in this analysis. There are a few continue variables that are significantly related. Age and income show a significant and positive correlation, which means that older respondents have a higher income. Also, social integration shows a significant positive correlation with income. This indicates that someone with a higher feeling of social integration has a higher income than someone with a lower feeling of social integration. And most importantly for this research, social integration and network diversity in terms of education show a significant positive correlation. So this indicates that someone with a higher feeling of social integration also has a higher score on network diversity in terms of education. Other continuous variables show no correlation between one another.

Table 2 also shows a correlation between continuous and nominal variables. Primary occupation shows a significant positive correlation with age and a significant negative correlation with income. Furthermore, gender and network diversity in terms of education show a significant negative correlation. This indicates that women might have a higher score on network diversity in terms of education than men. And lastly, gender and income show a significant positive correlation. So the older you get, the more income you generate.

4.2 Model evaluation

4.2.1 Quality of the models and assumptions

To estimate the quality of the linear regression models, we will look at the adjusted R square and the F change. The adjusted R square is used to correct for the addition of one or more variables. The F Changes shows whether there is a significant improvement in the quality of the model. These values are shown in table 3, and we will discuss them shortly in the next paragraphs.

Table 3 shows that the adjusted R square scores are varied. So shows model 1 an adjusted R square of 0,014, however, in model 2, the adjusted R square increased to a score of 0,024. This means that when network diversity in terms of education was added, more

variance was explained. However, both model 3 (R_a^2=0,022) and 4 (R_a^2=0,21) have not changed much compared to the score on the adjusted R square in model 2. So both the moderator gender and the interaction effect of network diversity in terms of education and gender did not increase the explained variance in the models.

There is a significant improvement in the quality of the model when the F change has increased in comparison to the earlier model. In table 3 we can see that model 2 (F change = 7,116; p = 0,008) shows a significant improvement compared to model 1 (F change = 4,082; p = 0,007). This means that when including network diversity in terms of education there is an increasement in the fit of the model on the data. The F change in model 3 (F change = 0,015; p = 0,903) and model 4 (F change = 0,030; p = 0,863) decrease and show no more significant p-values. In summary, while the independent variable network diversity in terms of education level significantly improves the model, gender and the interaction-effect do not.

Other than the adjusted R square and the F change to check on the quality of the models, there are also some assumptions for a linear analysis. Firstly, is the use of independent observations. This means that the answers given by the respondents are independent of each other. This dataset works with households, so in theory, it is possible that several respondents from a household were asked. As a result, the answers given may not be independent of each other. Unfortunately, we cannot completely rule this out; as a result of which, this assumption has been violated.

Other assumptions are those of linearity, homoscedasticity, and normality. All these are violated, as further details are shown in Appendix 3. This can affect the analysis and tests negatively. So conclusions about the population need to be made more carefully. To reduce the negative aspect of this effect, we decided to use a smaller alpha of 0.01 instead of 0.05, which makes it harder to find significant results. We will also carry out a binary logistic analysis to see if the possible effect is there found as well. These results can be found in Appendix 2.

4.2.2 Outliers and multicollinearity

Besides the model fit and the assumptions, we will also look at possible influential outliers and multicollinearity. Multicollinearity shows the interrelationship between the variables. This can be calculated using the ViF values. These values are shown in Table 3. All ViF-values scored below the value of 4, so there is no evidence of interrelationship between any of the variables.

To gain insight into the influence of outliers, the standardized residuals, the leverage, and the Cook's Distance were measured. The standardized residuals show seven outliers. On top of that, the leverage shows four influential points, and the Cook's distance shows 31 influential points. However, almost all of these outliers and influential points have a relatively low score on social integration. As with the standardized residuals, these scores can also look like outliers, but there are no analysis errors. After a thorough analysis of the outliers and influential points, which is shown in more detail in Appendix 3, we decided to delete 3 outliers. Which resulted in a population of 637 that we used in this Results paragraph and the analyses.

4.3 Hypothesis testing

In this paragraph, we will look at our hypotheses and the outcomes of the linear analysis shown in table 3. We will do this by following our earlier established analysis plan described in the methods paragraph.

In model 1 we look at the effect of the control variables on social integration. Only the variable income significantly affects the dependent variable social integration (b=0,053; p=0,004). The slope indicates that when income increases by 1 euros per month, the score on social integration will increase by as little as 0,053. This means when someone has a higher feeling of social integration, the chance of having a higher income is significant compared to having a lower income. This significant effect is shown in all models. The other control variables age and primary occupation show no significant effect on social integration.

In model 2, the independent variable network diversity in terms of education was added to the model. This was done to test the first hypothesis: “Having a more diverse network in terms of education will positively impact the social integration of immigrants”. Table 3 shows a significant effect of the relationship between network diversity in terms of education and the social integration of immigrants; this is shown in model 2 when looking at the slope and p-value of network diversity in terms of education (b= 0,031; p=0,008). The slope indicates that the when the variable network diversity in terms of education increases by 1, the score on social integration will increase by 0,031. This effect is significant at a significance level of 0.01, however, on a scale from 1-3, it is quite a small effect. This means that the chance of having a high feeling of social integration is a little greater among respondents who scored high on network diversity in terms of education. This effect is in line with our first hypothesis and the theory, so the first hypothesis may be accepted. When adding the independent variable network diversity in terms of education, other effects of the control variables did not change dramatically. As income still showed a significant effect with a similar slope. On top of that, age and primary occupation still showed no significant effects.

The moderator gender was added in model 3. Table 3 also shows no significant effect for the moderator (B = 0,004; p=0,903). Scoring higher on gender, and so being a male, leads to an increase of 0,019 in social integration. This indicates that men do feel a little more socially integrated than women. When looking at other variables used in model 3, all other variables almost stayed the same.

In model 4, the interaction variable was added. Here we will test our second hypothesis: “Men show a stronger effect when looking at the relationship between network diversity in terms of education and the social integration of immigrants than women do”. Table 3 again shows no significant effect for the added interaction variable (B = -0,004; p=0,863), so we cannot accept the second hypothesis of this research.

As explained earlier, we also decided on doing a binary logistic analysis next to the linear analysis carried out in this Results paragraph. The main reason for this was the

skewed distribution of the dependent variable social integration, which is shown in more detail in appendix 3. The binary logistic analysis showed no significant effects for both hypotheses. We have to mention that due to dichotomizing the social integration variable, we can argue that quite some information was lost. This can arise questions about the power of the models used in the binary logistic analysis. More detailed information on the logistic analysis can be found in appendix 2.

Table 3: Model estimates of all independent variables, including slopes, standard deviation, p-value of the slope, VIF-value, adjusted R^2 value, F-change value, and p-value of F-change.

	Model 1		**Model 2**		**Model 3**		**Model 4**		**ViF**
	b (SE)	**p**	**b (SE)**	**p**	**b (SE)**	**p**	**b (SE)**	**p**	
Constant	2,603 (0,056)	<0,001	2,599 (0,059)	<0,001	2,598 (0,056)	<0,001	2,598 (0,056)	<0,001	
Age	-0,002 (0,001)	0,165	-0,002 (0,001)	0,170	-0,002 (0,001)	0,174	-0,002 (0,001)	0,173	1,182
Income	0,053 (0,018)	0,004	0,054 (0,018)	0,003	0,053 (0,019)	0,005	0,053 (0,019)	0,005	1,482
Primary Occupation	-0,011 (0,040)	0,786	-0,007 (0,039)	0,850	-0,008 (0,040)	0,842	-0,008 (0,040)	0,838	1,306
Network Diversity in terms of education			0,031 (0,012)	0,008	0,031 (0,012)	0,008	0,032 (0,015)	0,030	1,686
Moderator: Gender					0,004 (0,037)	0,903	0,004 (0,037)	0,906	1,102
Interaction: Network Diversity in terms of education (centered) x Gender							-0,004 (0,024)	0,863	1,686
R_a^2		0,014		0,024		0,022		0,021	
F Change	4,082	0,007	7,116	0,008	0,015	0,903	0,030	0,863	
N		637		637		637		637	

In this analysis, we use a significance level of 0,01

5. Conclusion and Discussion

The social integration of immigrants is a significant phenomenon, as it can cause quite some difficulties for both the immigrants and the host countries. As in the Netherlands problems with cohesion, crime, and discrimination can be linked to the social integration of immigrants. This research looked at the social integration of immigrants in the Netherlands and if having a diverse network in terms of education will help the immigrant feel more socially integrated. On top of that, this research looked at the possible differences between men and women regarding the relationship between network diversity in terms of education and the social integration of immigrants. Because we were interested in these aspects, the following research question was formulated: *What is the effect of network diversity (of the five closest contacts mentioned in the survey) in terms of education on the social integration of immigrants? Is there a difference between men and women?*

Based on this research question, two hypotheses were formulated. The first hypothesis was the following: '*Having a more diverse network in terms of education will positively impact the social integration of immigrants'*. In the theory paragraph, we suggested that the relationship between network diversity in terms of education and the social integration of immigrants can be explained through the social capital theory. Having diverse contacts can lead to a diverse set of means, skills, and knowledge. All these different views of different educational levels, and the social contacts gained through bonding and bonding capital, lead to a higher feeling of life satisfaction and a lower sense of loneliness. The positive effect between network diversity in terms of education and the social integration of immigrants was also found in the analysis of this research. Even though there were some limitations, which we will discuss later in this chapter, we still believe there is evidence for this hypothesis. As it might be an interesting relationship to keep in mind while helping immigrants with their social integration in the Netherlands and for follow-up research.

The second hypothesis was: '*Men show a stronger effect when looking at the relationship between network diversity in terms of education and the social integration of immigrants than women do'*. In the theory paragraph, we suggested that men show a stronger relationship between network diversity in terms of education and the social integration because women still have to cope with traditional task distributions and fewer opportunities to work. As they are stronger monitored by their families, and have a lesser chance because of this to improve their human capital. This results in having more difficulty to find social contacts than men. This effect was, however, not found in the analysis in this research. This is interesting as, we also indicated in the theory paragraph, some research found women to be more socially integrated than men. As they are better at adjusting and are able to improve their life satisfaction more quickly as in some countries women have fewer rights than here (Avenarius, 2012; Bilecen & Seibel, 2021).

An effect that we did not necessarily want to test in this research, but shows an interesting result in the analysis, is the influence of income on social integration. This effect was already found in the literature as income can provide more means and opportunities to engage in activities that benefit the social integration process (Bult et al., 2011; Kanas et al., 2012).

Even though we got some interesting results, some limitations need to be mentioned. The first is the violation of all four assumptions of the linear analysis. The first assumption, independent observations, was violated because some respondents might not have answered independently, due to the research questioning more than one person in a household. The other assumptions, linearity, homoscedasticity, and normality are all violated as well. Therefore, a smaller alpha of 0.01 has been chosen, which makes it harder to find significant results.

Other aspects worth mentioning in sight of how this research was carried out, are the missing data, the reliability, and the validity. While removing missing data, we did a missing value analysis. This analysis illustrates that the extracted missing values were respondents

with an average lower age and income. This means that when working with the 640 respondents, we must consider that the removed respondents made the average age and income in the dataset lower. This might have changed the outcome of the analysis, as we look at an average higher age and income.

Some other points are the reliability and validity of this research. Sometimes, questions about loneliness and life satisfaction may be difficult to answer. Moreover, the respondents may have deliberately entered incorrect answers, for example, because of social desirability. This means that the answers given may not exactly match reality. On top of that uses this research secondary data. Because the data was not collected with this research in mind, some concepts have not been measured in the way they were meant for in this research. For example, social integration has in this research been measured by using questions about life satisfaction and loneliness. Although, as discussed in the theory, social contacts are also important when it comes to the social integration of immigrants. Moreover, these questions can be answered quite subjectively, as there can be different perceptions about life satisfaction and loneliness. In further research, this can be solved by using more objective measurement methods for social integration.

We should also note that network diversity in terms of education is measured with only the five closest contacts the respondent named in the survey. On top of that, most respondents did not include all five closest contacts. As most only named two or three of their closest contacts. This did not create problems with the making of the variable but can give another view as the social contacts used are limited. The fact that not everyone named all five close contacts can also be a consequence of a long questionnaire. As people tend to pay less attention or show less effort at the end of long questionaries. The questions used for the variable network diversity in terms of education were almost at the end.

Despite the mentioned limitations of this research, the results are still interesting for follow-up research. The researched relationships in this study are especially interesting because of the social networks. In further research, we recommend using statistical tools as

they can show the quality of the ties. Unfortunately, we could not show this with SPSS, but this is possible with UCINET. The difference between the first and second generation and the origin of the social contacts can also be shown in further research, as these might influence the social integration of immigrants (Algan, Dustmann, Glitz & Manning, 2010; Kanas et al., 2012). Another recommendation is to use more objectively measured questions to measure social integration. This may, at least partly, solve the problems of reliability and validity. All in all is it an interesting relationship to keep researching about. As it can give a better view on the life of immigrants and their social integration process. And maybe in that way, we can slowly start to solve the problems occurring with the social integration of immigrants.

References

Adams, R. E., & Serpe, R. T. (2000). Social integration, fear of crime, and life satisfaction. *Sociological perspectives*, *43*(4), 605-629.

Algan, Y., Dustmann, C., Glitz, A., & Manning, A. (2010). The economic situation of first and second-generation immigrants in France, Germany and the United Kingdom. *The Economic Journal, 120* (542), 4-30.

Arpino, B., & de Valk, H. (2018). Comparing life satisfaction of immigrants and natives across Europe: The role of social contacts. *Social indicators research*, *137*(3), 1163-1184.

Aslund, O., Bohlmark, A., & Nordstrom Skans, O. (2009). Age at migration and social integration. *IZA Discussion Paper, 4263*, 1-37.

Avenarius, C. B. (2012). Immigrant networks in new urban spaces: Gender and social integration. *International Migration*, *50*(5), 25-55.

Bilecen, B., & Seibel, V. (2021). Network explanations of the gender gap in migrants' employment patterns: Use of online and offline networks in the Netherlands. *Journal of Family Research*, *33*(2), 541–565. https://doi.org/10.20377/jfr-484

Bourdieu, P. (1986). The forms of capital. In J.G. Richardson (Ed.) *Handbook of Theory & Research for the Sociology of Education.* Greenwood Press.

Breiger, R. L. (2004). The analysis of social networks. In M. Hardy & A. Bryman (Ed.), *Handbook of Data Analysis* (pp. 505-526). SAGE Publications Ltd.

Bult, M. K., Verschuren, O., Jolngmans, M. J., Lindeman, E., & Ketelaar, M. (2011). What influences participation in leisure activities of children and youth with physical disabilities? A systematic review. *Research in developmental disabilities*, *32*(5), 1521-1529.

Burt, R.S. (1998). The gender of social capital. Rationality and Society, 10, 1, 5-47. https://doi.org/10.1177/104346398010001001.

Centerdata. (2014, March2 – February 25). Social integration and leisure LISS core study (wave 2). [Dataset and codebook]. Retrieved 19 May 2022, from https://www.lissdata.nl/sites/default/files/bestanden/Sample%20and%20recruitment%20of%20the%20MESS%20Immigrant%20panel.pdf

Centraal Bureau voor de Statistiek. (2022, 16 februari). *Hoeveel inwoners zijn in het buitenland geboren?* Retrieved from, https://www.cbs.nl/nl-nl/dossier/dossier-asiel-migratie-en-integratie/hoeveel-inwoners-zijn-in-het-buitenland-geboren-

Chiswick, B. R., & Miller, P. W. (2009). The international transferability of immigrants' human capital. *Economics of Education Review*, *28*(2), 162-169.

Coleman, J.S. (1990). *Foundations of Social Theory*. Cambridge: Harvard University Press.

Cruz-Saco, M. A. (2008). Promoting social integration: Economic, social and political dimensions with a focus on Latin America. *Journal Of Philosophical Economics*, 1-10.

Cuyvers, P. (2020, 2 juli). Vergeet de vlekjes niet bij het grote integratiesucces. *Trouw.* Retrieved from, https://www.volkskrant.nl/columns-opinie/vergeet-de-vlekjes-niet-bij-het-grote-integratiesucces~bff8c1cf/?referrer=https%3A%2F%2Fwww.google.com%2F

Dahinden, J. (2013). Cities, migrant incorporation, and ethnicity: A network perspective on boundary work. *Journal of International Migration and Integration*, *14*(1), 39-60.

Depalo, D., Faini, R., & Venturini, A. (2006). The social assimilation of immigrants. *IZA Discussion Paper, 2439*, 1-32.

Everett, M. G., Borgatti, S. P., & Johnson, J. C. (2018). *Analyzing social networks*. (2 ed.) SAGE Publications Ltd.

Huijnk, W., Andriessen, I. (2016). *Integratie in zicht?* (Nr. 740). Sociaal Cultureel Planbureau. Retrieved from, https://www.scp.nl/publicaties/publicaties/2016/12/15/integratie-in-zicht

Inglehart, R. & Norris, P. (2003). Rising tide: Gender equality and cultural change around the world. Cambridge: Cambridge University Press.

Kanas, A., Chiswick, B. R., Van Der Lippe, T., & Van Tubergen, F. (2012). Social contacts and the economic performance of immigrants: A panel study of immigrants in Germany. *International Migration Review*, *46*(3), 680-709.

Kindler, M., Ratcheva, V., & Piechowska, M. (2015). Social networks, social capital and migrant integration at local level. European literature review. *Institute For Research Into Superdiversity*, *6*, 1-22.

Kucharčíková, A. (2011). Human capital–definitions and approaches. *Human Resources Management & Ergonomics*, *5*(2), 60-70.

Lancee, B. (2010). The economic returns of immigrants' bonding and bridging social capital: The case of the Netherlands. *International Migration Review*, *44*(1), 202-226.

Laurentsyeva, N., & Venturini, A. (2017). The social integration of immigrants and the role of policy–A literature review. *Intereconomics*, *52*(5), 285-292.

Lin, N. (2008). A network theory of social capital. *The handbook of social capital*, *50*(1), 69.

Martinovic, B., Van Tubergen, F., & Maas, I. (2009). Changes in immigrants' social integration during the stay in the host country: The case of non-western immigrants in the Netherlands. *Social Science Research*, *38*(4), 870-882.

McPherson, M., Smith-Lovin, L., Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology, 27*(1), 415-444.

Mulder, J. (2014). *Social Integration and Leisure* (Wave 2). Tilburg University. Geraadpleegd op 4 april 2022, van https://www.dataarchive.lissdata.nl/study_units/view/571

Nafukho, F. M., Hairston, N., & Brooks, K. (2004). Human capital theory: Implications for human resource development. *Human Resource Development International*, *7*(4), 545-551.

Nawyn, S. J., Gjokaj, L., Agbényiga, D. L., & Grace, B. (2012). Linguistic isolation, social capital, and immigrant belonging. *Journal of Contemporary Ethnography*, *41*(3), 255-282.

Pachucki, M. C., & Leal, D. F. (2020). Is having an educationally diverse social network good for health?. *Network Science*, *8*(3), 418-444.

Poros, M. (2011). Migrant social networks: Vehicles for migration, integration, and development. *Migration Policy Institute*, *30*.

Putnam, R.D. (2007). E pluribus unum: civic engagement in a diverse and changing society, *Scandinavian Political Studies*, *30*(2), 137-174.

Riaño, Y., Baghdadi, N., & Wastl-Walter, D. (2006). *Social integration and social exclusion of skilled immigrant women from countries outside the European Union* (Nr. 51).Final research report. Bern, Universität Bern.

Röder, A., & Mühlau, P. (2014). Are they acculturating? Europe's immigrants and gender egalitarianism. Social Forces, 92(3), 899-928

Rubin, M., Watt, S.E., & Ramelli, M. (2012). Immigrants' social integration as a function of approach-avoidance orientation and problem-solving style. *International Journal of Intercultural Relations, 36*(4), 498-505.

Russell, D., Cutrona, C. E., Rose, J., & Yurko, K. (1984). Social and emotional loneliness: an examination of Weiss's typology of loneliness. *Journal of personality and social psychology*, *46*(6), 1313.

Schlueter, E., & Scheepers, P. (2010). The relationship between outgroup size and anti-outgroup attitudes: A theoretical synthesis and empirical test of group threat-and intergroup contact theory. *Social Science Research*, *39*(2), 285-295.

Stevens, N., & Westerhof, G. J. (2006). Marriage, social integration, and loneliness in the second half of life: A comparison of Dutch and German men and women. *Research on Aging*, *28*(6), 713-729.

Tselios, V., Noback, I., van Dijk, J., & McCann, P. (2015). Integration of immigrants, bridging social capital, ethnicity, and locality. *Journal of Regional Science*, *55*(3), 416-441.

Valenta, M. (2008). The workplace as an arena for identity affirmation and social integration of immigrants. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, *9*(2).

Van de Werfhorst, H. G., & Heath, A. (2019). Selectivity of migration and the educational disadvantages of second-generation immigrants in ten host societies. *European Journal of Population*, *35*(2), 347-378.

Van Tubergen, F. & Kalmijn, M. (2008). Language proficiency and usage among migrants in the Netherlands: Incentives or opportunities? European Sociological Review, 25, 2, 169-182. https://doi.org/10.1093/esr/jcn043

Appendix 1: Operationalisations

This appendix shows how the variables included in the analysis are coded. For each variable, it is shown what it initially looked like and what recodings were executed. The syntax is given for each variable, a frequency distribution and descriptive statistics of both the original and the new variable are included in the analysis.

Independent variable: Network diversity in terms of education

Original variables

The independent variable network diversity in terms of education contains five questions, which all asked about the level of education of the five closest social contacts of the respondent. The following questions are asked: “What is the highest level of education person 1 completed?”.

```
FREQUENCIES VARIABLES=fb14b429 fb14b440 fb14b451 fb14b462 fb14b473
  /ORDER=ANALYSIS.
```

What is the highest level of education that person 1 completed?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	not (yet) completed any educational program	12	1,9	1,9	1,9
	primary school	29	4,5	4,6	6,5
	VMBO, LBO, MULO, ULO, MAVO (lower/intermediate secondary education, US: junior h	111	17,3	17,5	23,9
	HAVO / VWO (higher/pre-university secondary education, US: senior high school)	62	9,7	9,8	33,7
	MBO (intermediate professional education, US: junior college)	106	16,6	16,7	50,4
	HBO (higher professional education, US: college)	151	23,6	23,8	74,2
	university	133	20,8	20,9	95,1
	I don't know	31	4,8	4,9	100,0
	Total	635	99,2	100,0	
Missing	System	5	,8		
Total		640	100,0		

What is the highest level of education that person 2 completed?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	not (yet) completed any educational program	11	1,7	1,9	1,9
	primary school	22	3,4	3,8	5,7
	VMBO, LBO, MULO, ULO, MAVO (lower/intermediate secondary education, US: junior h	97	15,2	16,8	22,5
	HAVO / VWO (higher/pre-university secondary education, US: senior high school)	74	11,6	12,8	35,3
	MBO (intermediate professional education, US: junior college)	105	16,4	18,2	53,5
	HBO (higher professional education, US: college)	137	21,4	23,7	77,2
	university	95	14,8	16,4	93,6
	I don't know	37	5,8	6,4	100,0
	Total	578	90,3	100,0	
Missing	System	62	9,7		
Total		640	100,0		

What is the highest level of education that person 3 completed?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	not (yet) completed any educational program	7	1,1	1,4	1,4
	primary school	19	3,0	3,8	5,2
	VMBO, LBO, MULO, ULO, MAVO (lower/intermediate secondary education, US: junior h	90	14,1	18,0	23,2
	HAVO / VWO (higher/pre-university secondary education, US: senior high school)	69	10,8	13,8	37,0
	MBO (intermediate professional education, US: junior college)	82	12,8	16,4	53,4
	HBO (higher professional education, US: college)	110	17,2	22,0	75,4
	university	83	13,0	16,6	92,0
	I don't know	40	6,3	8,0	100,0
	Total	500	78,1	100,0	
Missing	System	140	21,9		
Total		640	100,0		

What is the highest level of education that person 4 completed?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	not (yet) completed any educational program	6	,9	1,6	1,6
	primary school	13	2,0	3,5	5,1
	VMBO, LBO, MULO, ULO, MAVO (lower/intermediate secondary education, US: junior h	60	9,4	16,3	21,4
	HAVO / VWO (higher/pre-university secondary education, US: senior high school)	49	7,7	13,3	34,7
	MBO (intermediate professional education, US: junior college)	42	6,6	11,4	46,1
	HBO (higher professional education, US: college)	89	13,9	24,1	70,2
	university	86	13,4	23,3	93,5
	I don't know	24	3,8	6,5	100,0
	Total	369	57,7	100,0	
Missing	System	271	42,3		
Total		640	100,0		

What is the highest level of education that person 5 completed?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	not (yet) completed any educational program	3	,5	1,2	1,2
	primary school	6	,9	2,3	3,5
	VMBO, LBO, MULO, ULO, MAVO (lower/intermediate secondary education, US: junior h	39	6,1	15,2	18,8
	HAVO / VWO (higher/pre-university secondary education, US: senior high school)	25	3,9	9,8	28,5
	MBO (intermediate professional education, US: junior college)	35	5,5	13,7	42,2
	HBO (higher professional education, US: college)	61	9,5	23,8	66,0
	university	58	9,1	22,7	88,7
	I don't know	29	4,5	11,3	100,0
	Total	256	40,0	100,0	
Missing	System	384	60,0		
Total		640	100,0		

Recodings

We first extracted the system missing values and made new items for all five social contacts without missing values.

```
RECODE fb14b429 fb14b440 fb14b451 fb14b462 fb14b473 (1=1) (2=2)
(3=3) (4=4) (5=5) (6=6) (7=7) (8=SYSMIS) INTO Person_1 Person_2
Person_3 Person_4 Person_5. VARIABLE LABELS  Person_1 'What is the
highest level of education that person 1 completed?' /Person_2
'What is the highest level of education that person 2 completed?'
/Person_3 'What is '+ 'the highest level of education that person
3 completed?' /Person_4 'What is the highest level '+ 'of
education that person 4 completed?' /Person_5 'What is the highest
level of education that '+ 'person 5 completed?'. EXECUTE.
```

What is the highest level of education that person 1 completed?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	12	1,9	2,0	2,0
	2,00	29	4,5	4,8	6,8
	3,00	111	17,3	18,4	25,2
	4,00	62	9,7	10,3	35,4
	5,00	106	16,6	17,5	53,0
	6,00	151	23,6	25,0	78,0
	7,00	133	20,8	22,0	100,0
	Total	604	94,4	100,0	
Missing	System	36	5,6		
Total		640	100,0		

What is the highest level of education that person 2 completed?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	11	1,7	2,0	2,0
	2,00	22	3,4	4,1	6,1
	3,00	97	15,2	17,9	24,0
	4,00	74	11,6	13,7	37,7
	5,00	105	16,4	19,4	57,1
	6,00	137	21,4	25,3	82,4
	7,00	95	14,8	17,6	100,0
	Total	541	84,5	100,0	
Missing	System	99	15,5		
Total		640	100,0		

What is the highest level of education that person 3 completed?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	7	1,1	1,5	1,5
	2,00	19	3,0	4,1	5,7
	3,00	90	14,1	19,6	25,2
	4,00	69	10,8	15,0	40,2
	5,00	82	12,8	17,8	58,0
	6,00	110	17,2	23,9	82,0
	7,00	83	13,0	18,0	100,0
	Total	460	71,9	100,0	
Missing	System	180	28,1		
Total		640	100,0		

What is the highest level of education that person 4 completed?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	6	,9	1,7	1,7
	2,00	13	2,0	3,8	5,5
	3,00	60	9,4	17,4	22,9
	4,00	49	7,7	14,2	37,1
	5,00	42	6,6	12,2	49,3
	6,00	89	13,9	25,8	75,1
	7,00	86	13,4	24,9	100,0
	Total	345	53,9	100,0	
Missing	System	295	46,1		
Total		640	100,0		

What is the highest level of education that person 5 completed?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	3	,5	1,3	1,3
	2,00	6	,9	2,6	4,0
	3,00	39	6,1	17,2	21,1
	4,00	25	3,9	11,0	32,2
	5,00	35	5,5	15,4	47,6
	6,00	61	9,5	26,9	74,4
	7,00	58	9,1	25,6	100,0
	Total	227	35,5	100,0	
Missing	System	413	64,5		
Total		640	100,0		

As seen in the tables above, not all respondents filled in all five closest contacts. To not lose any information, we took the minimum and maximum of each case and subtracted these from each other. The higher the score, the greater the network diversity in terms of education. There was no need for everyone to fill in all five contacts, as we only looked at network diversity, which was still measured using this method. How this was precisely done is shown in the syntax below.

```
COMPUTE ND_MAX=MAX(Person_1,Person_2,Person_3,Person_4,Person_5).
EXECUTE.
COMPUTE ND_MIN=MIN(Person_1,Person_2,Person_3,Person_4,Person_5).
EXECUTE.
COMPUTE Network_Diversity=ND_MAX - ND_MIN.
EXECUTE.
```

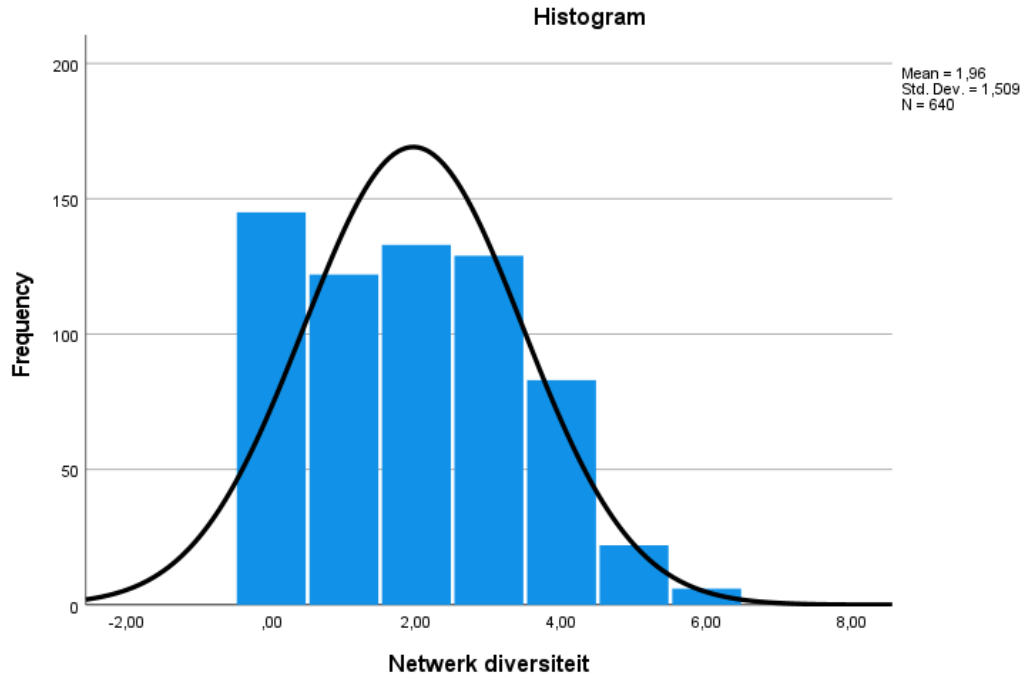
Variable used in the analysis

After recoding, our final independent variable network diversity in terms of education was constructed. The new variable Network_Diversity_Mean is normally divided, and the distributions are shown in the table below. Not many respondents show a high score of network diversity in terms of education. Most respondents show no network diversity at all.

```
FREQUENCIES VARIABLES=Network_Diversity
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.
```

Netwerk diversiteit

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	145	22,7	22,7	22,7
	1,00	122	19,1	19,1	41,7
	2,00	133	20,8	20,8	62,5
	3,00	129	20,2	20,2	82,7
	4,00	83	13,0	13,0	95,6
	5,00	22	3,4	3,4	99,1
	6,00	6	,9	,9	100,0
	Total	640	100,0	100,0	



Dependent variable: Social integration of immigrants

Original variables

The dependent variable social integration of immigrants contains six questions, which all asked about the level of life satisfaction and loneliness. These are the following six variables: 'I have a sense of emptiness around me', 'I miss having people around me', 'I often feel deserted', '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' and 'there are enough people to whom I feel closely connected'. In the syntax and the following tables, the distribution of all items are shown.

```
FREQUENCIES VARIABLES=fb14b307 fb14b308 fb14b309 fb14b310
fb14b311 fb14b312
  /ORDER=ANALYSIS.
```

I have a sense of emptiness around me

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	30	4,7	4,7	4,7
	more or less	143	22,3	22,3	27,0
	no	467	73,0	73,0	100,0
	Total	640	100,0	100,0	

there are enough people I can count on in case of a misfortune

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	436	68,1	68,1	68,1
	more or less	168	26,3	26,3	94,4
	no	36	5,6	5,6	100,0
	Total	640	100,0	100,0	

I know a lot of people that I can fully rely on

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	345	53,9	53,9	53,9
	more or less	214	33,4	33,4	87,3
	no	81	12,7	12,7	100,0
	Total	640	100,0	100,0	

there are enough people to whom I feel closely connected

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	405	63,3	63,3	63,3
	more or less	178	27,8	27,8	91,1
	no	57	8,9	8,9	100,0
	Total	640	100,0	100,0	

I miss having people around me

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	69	10,8	10,8	10,8
	more or less	146	22,8	22,8	33,6
	no	425	66,4	66,4	100,0
	Total	640	100,0	100,0	

I often feel deserted

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	30	4,7	4,7	4,7
	more or less	73	11,4	11,4	16,1
	no	537	83,9	83,9	100,0
	Total	640	100,0	100,0	

We also looked at the opportunity to include an item about social contacts, an important aspect of social integration. The item would be: ‘How satisfied are you with your social contacts?’ However, we decided not to implement this item when looking at Cronbach's alpha. The table of Cronbach’s alpha is shown below.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
How satisfied are you with your social contacts?	15,6125	7,011	,014	,806
I have a sense of emptiness around me	43,1750	22575,325	,029	,001
there are enough people I can count on in case of a misfortune	43,2328	22588,141	-,045	,001
I know a lot of people that I can fully rely on	43,4453	22571,631	,039	,000
there are enough people to whom I feel closely connected	43,3141	22574,385	,029	,001
I miss having people around me	43,3016	22572,230	,038	,000
I often feel deserted	43,0656	22577,980	,014	,001

Recodings

Respondents could answer these questions with (1) yes, (2) more or less, and (3) no. To have all questions in the same direction, three questions (fb14308, fb14308=9, fb14310) were mirrored, where (1) no, (2) more or less, and (3) yes. Subsequently, the six questions were combined by making a mean variable. The new variable Sociale_integratie_Mean ranges from 1 to 3, with a higher score indicating a higher feeling of social integration.

```
RECODE fb14b308 fb14b309 fb14b310 (1=3) (2=2) (3=1) INTO
fb14b308_Gespiegeld fb14b309_Gespiegeld fb14b310_Gespiegeld.

VARIABLE LABELS  fb14b308_Gespiegeld 'there are enough people I
can count on in case of a '+ 'misfortune' /fb14b309_Gespiegeld 'I
know a lot of people that I can fully rely on'
/fb14b310_Gespiegeld 'there are enough people to whom I feel
closely connected'. EXECUTE.
```

```
COMPUTE
Sociale_integratie_Mean=MEAN(fb14b307,fb14b308_Gespiegeld,fb14b309
_Gespiegeld, fb14b310_Gespiegeld,fb14b311,fb14b312).

EXECUTE.
```

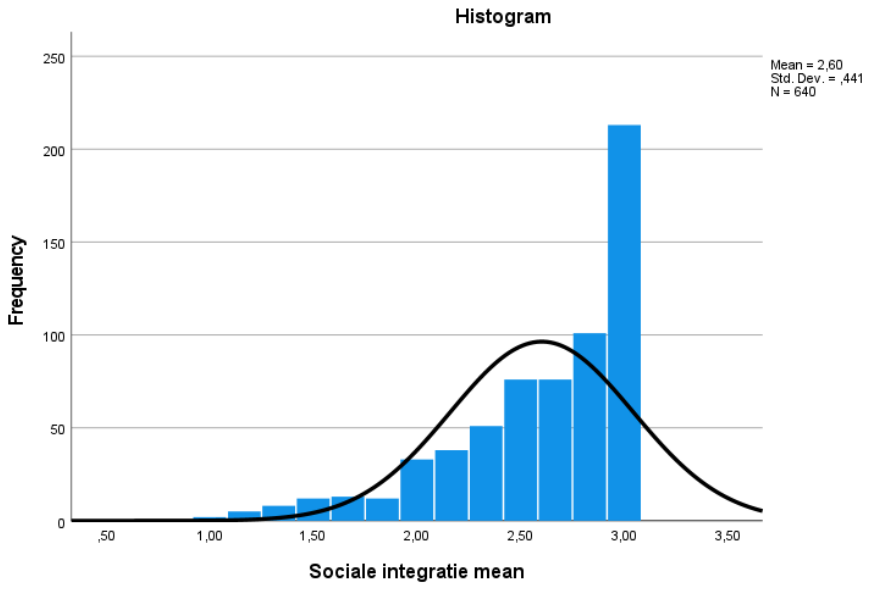
Variable used in the analysis

We used the mean variable in our analysis. As seen in the histogram this variable is not normally divided. This will be further elaborated on in the section model assumption. We do see that a lot immigrants feel an extremely sense of social integration.

```
FREQUENCIES VARIABLES=Sociale_integratie_Mean
  /STATISTICS=STDDEV VARIANCE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /ORDER=ANALYSIS.
```

Sociale integratie mean

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	2	,3	,3	,3
	1,17	5	,8	,8	1,1
	1,33	8	1,3	1,3	2,3
	1,50	12	1,9	1,9	4,2
	1,67	13	2,0	2,0	6,3
	1,83	12	1,9	1,9	8,1
	2,00	33	5,2	5,2	13,3
	2,17	38	5,9	5,9	19,2
	2,33	51	8,0	8,0	27,2
	2,50	76	11,9	11,9	39,1
	2,67	76	11,9	11,9	50,9
	2,83	101	15,8	15,8	66,7
	3,00	213	33,3	33,3	100,0
	Total	640	100,0	100,0	



Gender

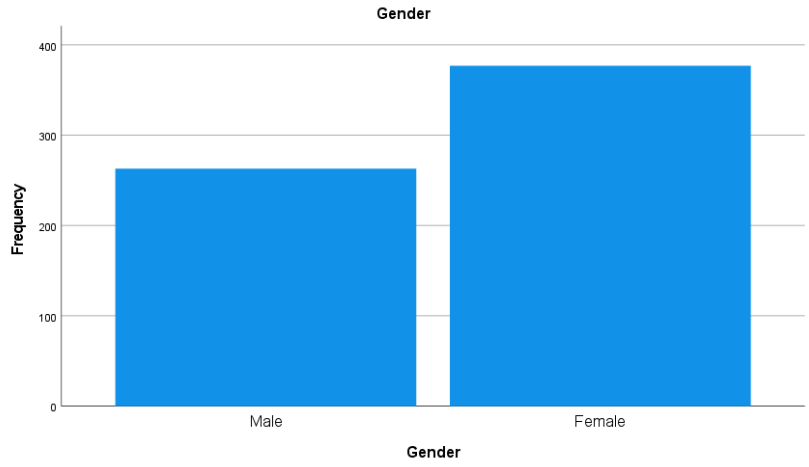
Original variables

For the moderation effect gender, respondents were asked in the variable ‘geslacht’ about their gender. They could use the following answer categories (1) male and (2) female.

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

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	263	41,1	41,1	41,1
	Female	377	58,9	58,9	100,0
	Total	640	100,0	100,0	



Recodings

This variable was recoded to (0) female and (1) male and renamed ‘gender’.

```
RECODE geslacht (2=0) (1=1) INTO Gender.
VARIABLE LABELS  Gender 'Gender, 0:female, 1:male'.
EXECUTE.
```

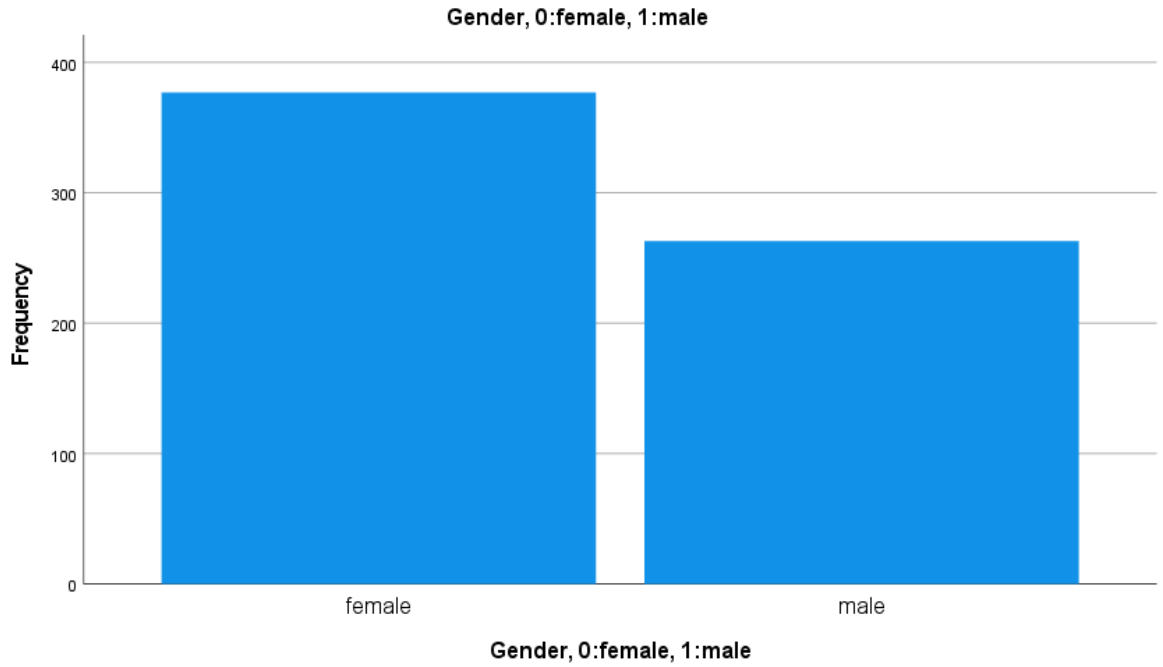
Variable used in the analysis

The descriptives of the new variable are as follows.

```
FREQUENCIES VARIABLES=Gender
  /BARCHART FREQ
  /ORDER=ANALYSIS.
```

Gender, 0:female, 1:male

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	female	377	58,9	58,9	58,9
	male	263	41,1	41,1	100,0
	Total	640	100,0	100,0	



Age

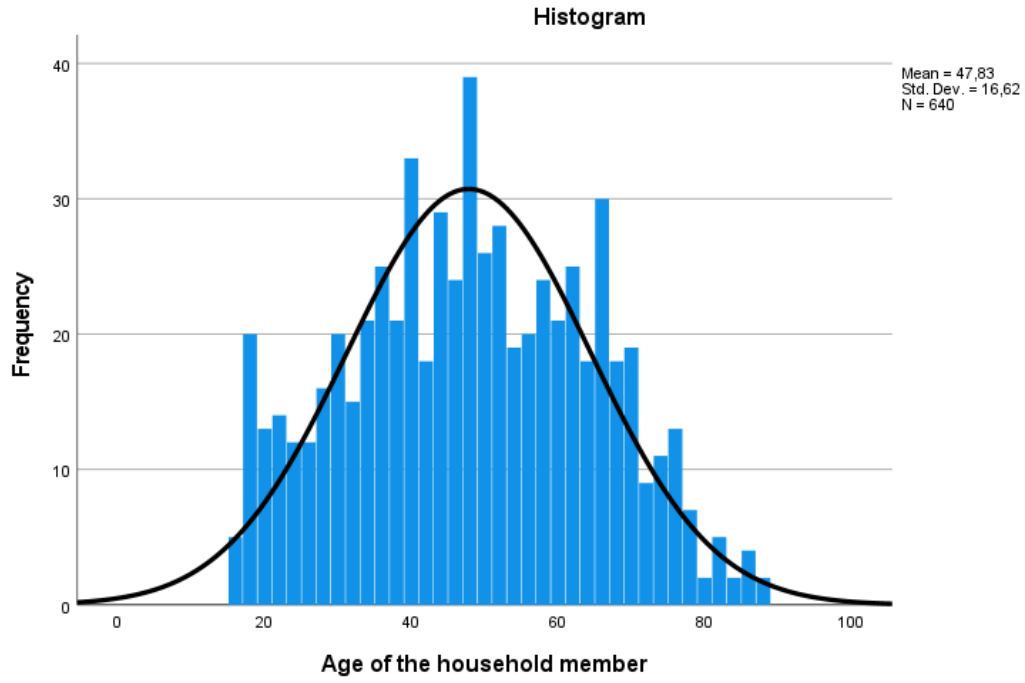
The respondents' age in this questionnaire varies from the age of 16 to 88 years old. It is a continuous variable and is used as a control variable in the analysis of this research.

```
FREQUENCIES VARIABLES=leeftijd
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.
```

Statistics

Age of the household member

N	Valid	640
	Missing	0
Mean		47,83
Median		48,00
Mode		48
Std. Deviation		16,620
Minimum		16
Maximum		88
Sum		30613



Income

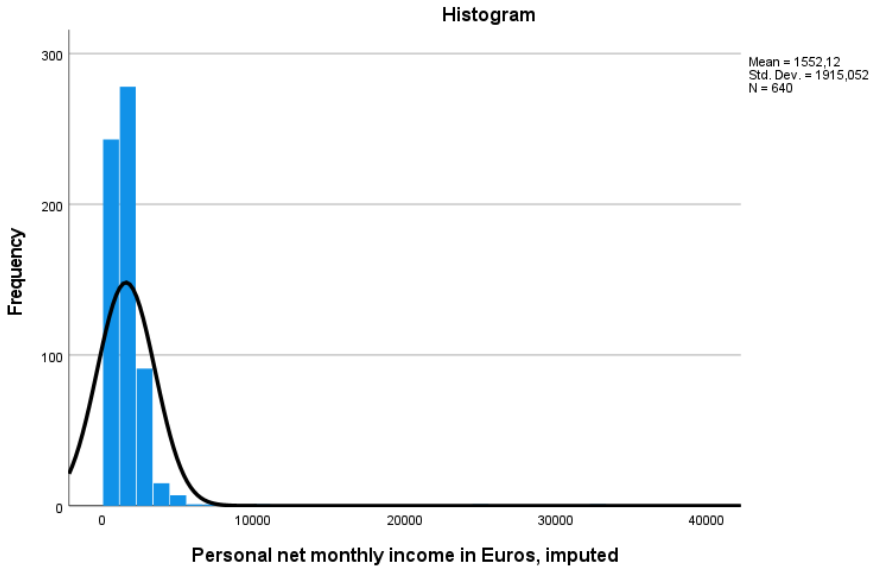
Income is a continuous variable and is also used as a control variable in the analysis of this research. This variable shows a wide range in incomes from 0 to 68388 euro's a month.

```
FREQUENCIES VARIABLES=nettoink_f
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.
```

Statistics

Personal net monthly income in Euros, imputed

N	Valid	640
	Missing	0
Mean		1552,12
Median		1350,00
Mode		0
Std. Deviation		1915,052
Minimum		0
Maximum		33274
Sum		993357



Recodings

We divided this variable by thousand, so it is easier in our regression to make assumptions and understands what happens. This new variable was called 'income'.

```
COMPUTE income=nettoink_f /
1000.
```

New variable

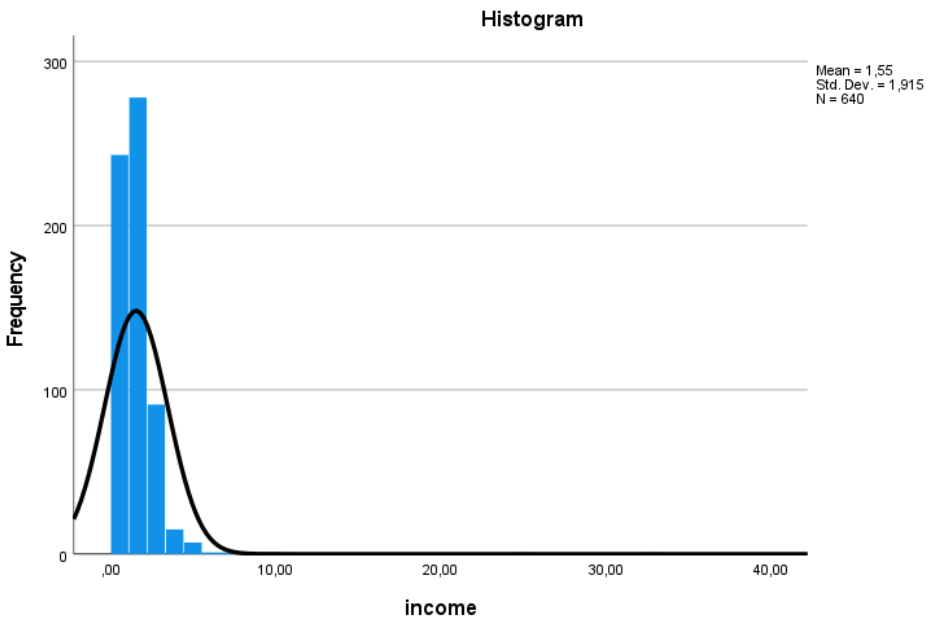
The distribution of the variable looks the same as the old one but is divided by 1000.

```
FREQUENCIES VARIABLES=income
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.
```

Statistics

income

N	Valid	640
	Missing	0
Mean		1,5521
Median		1,3500
Mode		,00
Std. Deviation		1,91505
Minimum		,00
Maximum		33,27
Sum		993,36



Primary Occupation

Original variables

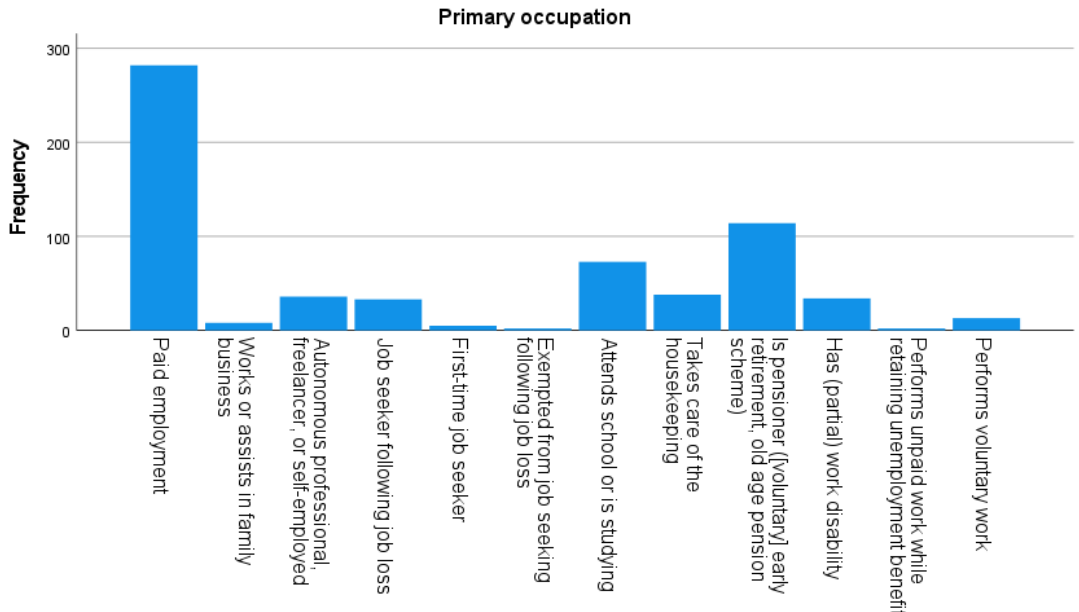
The last control variable used in the analysis of this research is the variable primary occupation. This is the categorical variable ‘belbezig’, with many answer categories (1) paid employment; (2) works or assists in family business; (3) autonomous professional, freelancer, or self-employed; (4) job 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

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

work while retaining unemployment benefit; (12) performs voluntary work; (13) does something else; and (14) is too young to have an occupation.

Primary occupation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Paid employment	282	44,1	44,1	44,1
	Works or assists in family business	8	1,3	1,3	45,3
	Autonomous professional, freelancer, or self-employed	36	5,6	5,6	50,9
	Job seeker following job loss	33	5,2	5,2	56,1
	First-time job seeker	5	,8	,8	56,9
	Exempted from job seeking following job loss	2	,3	,3	57,2
	Attends school or is studying	73	11,4	11,4	68,6
	Takes care of the housekeeping	38	5,9	5,9	74,5
	Is pensioner ([voluntary] early retirement, old age pension scheme)	114	17,8	17,8	92,3
	Has (partial) work disability	34	5,3	5,3	97,7
	Performs unpaid work while retaining unemployment benefit	2	,3	,3	98,0
	Performs voluntary work	13	2,0	2,0	100,0
	Total	640	100,0	100,0	



Recodings

Due to many answer options for ‘belbezig’, a recoding has taken place, resulting in two answer categories: 0= employed, and 1= unemployed. We only needed to know if the respondents were working or not, which was shown by this distribution.

```
RECODE belbezig (1 thru 3=0) (11 thru 12=0) (4 thru 10=1) INTO
Primary_occupation.
```

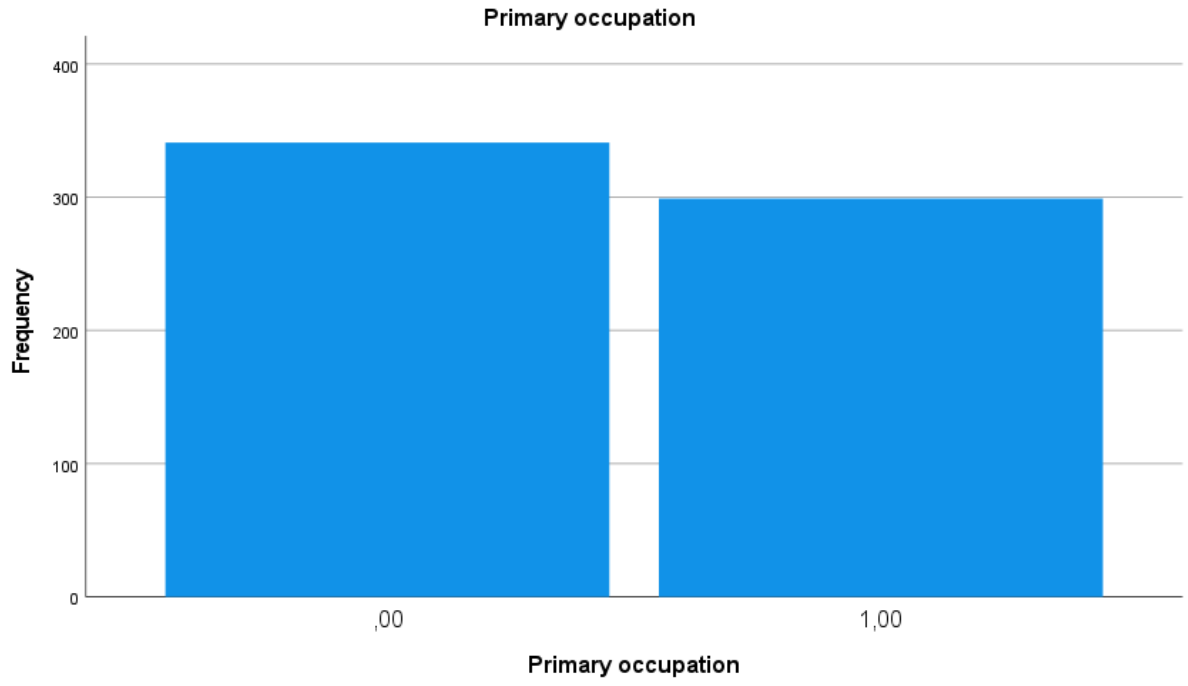
Variable used in the analysis

The descriptives of the new variable are as follows.

```
FREQUENCIES VARIABLES=Primary_occupation
  /STATISTICS=STDDEV VARIANCE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /BARCHART FREQ
  /ORDER=ANALYSIS.
```

Primary occupation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	341	53,3	53,3	53,3
	1,00	299	46,7	46,7	100,0
	Total	640	100,0	100,0	



Interaction variables

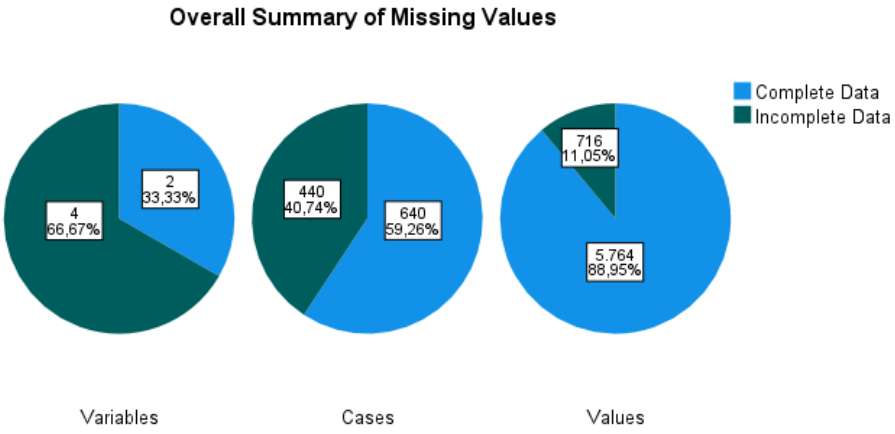
The interaction variables consist of the product between the moderator Gender and the independent variable Network Diversity in terms of education.

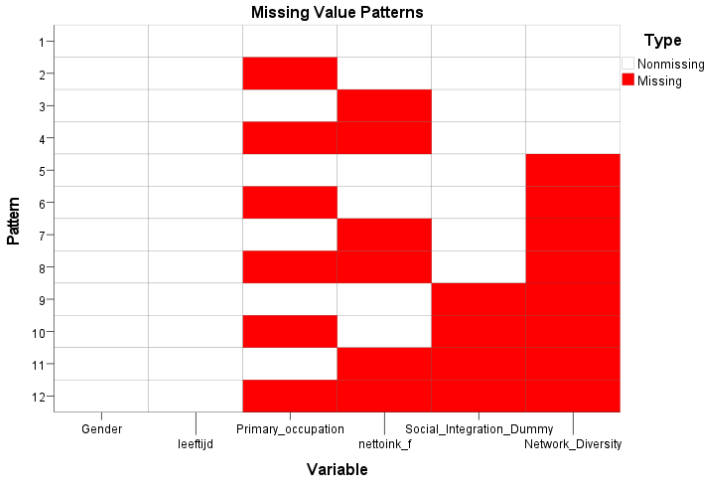
```
COMPUTE Network_Diversity_centr=Network_Diversity - 1.9545.
EXECUTE.
COMPUTE Interaction=Network_Diversity_centr * Gender.
EXECUTE.
```

Missing value analysis

When learning that our variables had quite some missing values, we decided to do a missing value analysis. In this way, we have a clear view of who the missing values are and their consequences for our data.

```
MULTIPLE IMPUTATION  Gender leeftijd Primary_occupation
nettoink_f Network_Diversity
    Social_Integration_Dummy
   /IMPUTE METHOD=NONE
   /MISSINGSUMMARIES  OVERALL VARIABLES (MAXVARS=25
MINPCTMISSING=10) PATTERNS.
MVA VARIABLES=leeftijd nettoink_f Network_Diversity
Social_Integration_Dummy Gender
    Primary_occupation
  /MAXCAT=25
  /CATEGORICAL=Gender Primary_occupation
  /TTEST NOPROB PERCENT=5
  /TPATTERN PERCENT=1.
```





Separate Variance t Tests[a]

		leeftijd	nettoink_f	Network_Diversity	Social_Integration_Dummy
Network_Diversity	t	6,8	,9	.	5,2
	df	919,1	517,4	.	239,5
	# Present	673	643	673	673
	# Missing	407	384	0	169
	Mean(Present)	47,87	1547,02	1,9108	,7236
	Mean(Missing)	41,20	1374,95	.	,5030
Social_Integration_Dummy	t	8,6	2,5	.	.
	df	443,7	1003,6	.	.
	# Present	842	803	673	842
	# Missing	238	224	0	0
	Mean(Present)	47,36	1548,52	1,9108	,6793
	Mean(Missing)	38,26	1246,67	.	.

For each quantitative variable, pairs of groups are formed by indicator variables (present, missing).

Appendix 2: Statistical Analyses

Univariate statistics

For the continuous variables, the mean, standard deviation, minimum and maximum, median, and percentiles are shown. For the categorical and dummy variables, the proportions are shown. The distributions between the score on the categorical and dummy variables are shown in appendix 1.

```
FREQUENCIES VARIABLES=Social_integratie_Mean Network_Diversity
Gender leeftijd income Primary_occupation
  /NTILES=4
  /STATISTICS=STDDEV VARIANCE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /ORDER=ANALYSIS.
```

Statistics

		Sociale integratie mean	Netwerk diversiteit	Gender, 0: female, 1:male	Age of the household member	income	Primary occupation
N	Valid	637	637	637	637	637	637
	Missing	0	0	0	0	0	0
Mean		2,6010	1,9545	,4097	47,76	1,4677	,4647
Median		2,6667	2,0000	,0000	48,00	1,3500	,0000
Mode		3,00	,00	,00	48	,00	,00
Std. Deviation		,44176	1,50350	,49217	16,615	1,11557	,49914
Variance		,195	2,261	,242	276,058	1,245	,249
Minimum		1,00	,00	,00	16	,00	,00
Maximum		3,00	6,00	1,00	88	10,00	1,00
Sum		1656,83	1245,00	261,00	30422	934,93	296,00
Percentiles	25	2,3333	1,0000	,0000	36,00	,7500	,0000
	50	2,6667	2,0000	,0000	48,00	1,3500	,0000
	75	3,0000	3,0000	1,0000	61,00	2,0000	1,0000

Bivariate statistics

The correlation between **Gender and Age** is measured by the Pearson's Correlation and T-Test for means methods.

```
CORRELATIONS
  /VARIABLES=Gender leeftijd
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.
T-TEST GROUPS=Gender (0 1)
  /MISSING=ANALYSIS
  /VARIABLES=leeftijd
  /ES DISPLAY(TRUE)
  /CRITERIA=CI(.95).
```

Correlations

		Gender, 0: female, 1:male	Age of the household member
Gender, 0:female, 1:male	Pearson Correlation	1	,043
	Sig. (2-tailed)		,275
	N	637	637
Age of the household member	Pearson Correlation	,043	1
	Sig. (2-tailed)	,275	
	N	637	637

The correlation between **Gender and Income** is measured by the Pearson's Correlation and T-Test for means methods.

```
CORRELATIONS
  /VARIABLES=Gender income
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.

T-TEST GROUPS=Gender (0 1)
  /MISSING=ANALYSIS
  /VARIABLES=income
  /ES DISPLAY(TRUE)
  /CRITERIA=CI(.95).
```

Correlations

		Gender, 0: female, 1:male	income
Gender, 0:female, 1:male	Pearson Correlation	1	,277**
	Sig. (2-tailed)		<,001
	N	637	637
income	Pearson Correlation	,277**	1
	Sig. (2-tailed)	<,001	
	N	637	637

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

The correlation between **Gender and Primary Occupation** is measured with the Chi squares and Cramers V.

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

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1,029[a]	1	,310		
Continuity Correction[b]	,872	1	,350		
Likelihood Ratio	1,030	1	,310		
Fisher's Exact Test				,333	,175
Linear-by-Linear Association	1,028	1	,311		
N of Valid Cases	637				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 121,28.

b. Computed only for a 2x2 table

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-,040	,310
	Cramer's V	,040	,310
N of Valid Cases		637	

The correlation between **Gender and Network Diversity in terms of education** is measured by the Pearson's Correlation and T-Test for means methods.

```
CORRELATIONS
  /VARIABLES=Gender
Network_Diversity
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.


T-TEST GROUPS=Gender (0 1)
  /MISSING=ANALYSIS

/VARIABLES=Network_Diversity
  /ES DISPLAY(TRUE)
  /CRITERIA=CI(.95).
```

Correlations

		Gender, 0: female, 1:male	Netwerk diversiteit
Gender, 0:female, 1:male	Pearson Correlation	1	-,085*
	Sig. (2-tailed)		,031
	N	637	637
Netwerk diversiteit	Pearson Correlation	-,085*	1
	Sig. (2-tailed)	,031	
	N	637	637

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

The correlation between **Gender and Social Integration** by the Pearson's Correlation and T-Test for means methods.

```
CORRELATIONS
  /VARIABLES=Gender Sociale_integratie_Mean
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.


T-TEST GROUPS=Gender (0 1)
  /MISSING=ANALYSIS
  /VARIABLES=Sociale_integratie_Mean
  /ES DISPLAY(TRUE)
  /CRITERIA=CI(.95).
```

Correlations

		Gender, 0: female, 1:male	Sociale integratie mean
Gender, 0:female, 1:male	Pearson Correlation	1	,031
	Sig. (2-tailed)		,433
	N	637	637
Sociale integratie mean	Pearson Correlation	,031	1
	Sig. (2-tailed)	,433	
	N	637	637

The correlation between **Age and Income** is measured by the Pearson's Correlation method.

```
CORRELATIONS
  /VARIABLES=leeftijd income
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.
```

Correlations

		Age of the household member	income
Age of the household member	Pearson Correlation	1	,267**
	Sig. (2-tailed)		<,001
	N	637	637
income	Pearson Correlation	,267**	1
	Sig. (2-tailed)	<,001	
	N	637	637

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

The ANOVA test and the R-squared measure the correlation between **Age and Primary Occupation.**

```
CORRELATIONS
  /VARIABLES=Primary_occupation leeftijd
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT leeftijd
  /METHOD=ENTER Primary_occupation.
```

Correlations

		Primary occupation	Age of the household member
Primary occupation	Pearson Correlation	1	,153**
	Sig. (2-tailed)		<,001
	N	637	637
Age of the household member	Pearson Correlation	,153**	1
	Sig. (2-tailed)	<,001	
	N	637	637

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

The correlation between **Age and Network Diversity in terms of education** is measured by the Pearson's Correlation method.

```
CORRELATIONS
  /VARIABLES=leeftijd
Network_Diversity
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.
```

Correlations

		Age of the household member	Netwerk diversiteit
Age of the household member	Pearson Correlation	1	-,020
	Sig. (2-tailed)		,614
	N	637	637
Netwerk diversiteit	Pearson Correlation	-,020	1
	Sig. (2-tailed)	,614	
	N	637	637

The correlation between **Age and Social Integration** is measured by the Pearson's Correlation method.

```
CORRELATIONS
  /VARIABLES=leeftijd Social_integratie_Mean
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.
```

Correlations

		Age of the household member	Sociale integratie mean
Age of the household member	Pearson Correlation	1	-,025
	Sig. (2-tailed)		,524
	N	637	637
Sociale integratie mean	Pearson Correlation	-,025	1
	Sig. (2-tailed)	,524	
	N	637	637

The correlation between **Income and Primary Occupation** is measured by the Pearson's Correlation and T-Test for means methods.

```
CORRELATIONS
  /VARIABLES=Primary_occupation income
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.

T-TEST GROUPS=Primary_occupation (0 1)
  /MISSING=ANALYSIS
  /VARIABLES=income
  /ES DISPLAY(TRUE)
  /CRITERIA=CI(.95).
```

Correlations

		Primary occupation	income
Primary occupation	Pearson Correlation	1	-,393**
	Sig. (2-tailed)		<,001
	N	637	637
income	Pearson Correlation	-,393**	1
	Sig. (2-tailed)	<,001	
	N	637	637

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

The correlation between **Income and Network Diversity in terms of education** is measured by the Pearson's Correlation method.

```
CORRELATIONS
  /VARIABLES=income Network_Diversity
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.
```

Correlations

		income	Netwerk diversiteit
income	Pearson Correlation	1	-,006
	Sig. (2-tailed)		,878
	N	637	637
Netwerk diversiteit	Pearson Correlation	-,006	1
	Sig. (2-tailed)	,878	
	N	637	637

The correlation between **Income and Social Integration** is measured by the Pearson's Correlation method.

```
CORRELATIONS
  /VARIABLES=income Social_integratie_Mean
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.
```

Correlations

		income	Sociale integratie mean
income	Pearson Correlation	1	,123**
	Sig. (2-tailed)		,002
	N	637	637
Sociale integratie mean	Pearson Correlation	,123**	1
	Sig. (2-tailed)	,002	
	N	637	637

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

The correlation between **Primary Occupation and Network Diversity in terms of education** is measured by the Pearson's Correlation and T-Test for means methods.

```
CORRELATIONS
  /VARIABLES=Primary_occupation
Network_Diversity
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.


T-TEST GROUPS=Primary_occupation (0 1)
  /MISSING=ANALYSIS
  /VARIABLES=Network_Diversity
  /ES DISPLAY(TRUE)
  /CRITERIA=CI(.95).
```

Correlations

		Primary occupation	Netwerk diversiteit
Primary occupation	Pearson Correlation	1	-,030
	Sig. (2-tailed)		,443
	N	637	637
Netwerk diversiteit	Pearson Correlation	-,030	1
	Sig. (2-tailed)	,443	
	N	637	637

The correlation between **Primary Occupation and Social Integration** is measured by the Pearson's Correlation and T-Test for means methods.

```
CORRELATIONS
  /VARIABLES=Primary_occupation
Social_integratie_Mean
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.


T-TEST GROUPS=Primary_occupation (0 1)
  /MISSING=ANALYSIS
  /VARIABLES= Social_integratie_Mean
  /ES DISPLAY(TRUE)
  /CRITERIA=CI(.95).
```

Correlations

		Primary occupation	Sociale integratie mean
Primary occupation	Pearson Correlation	1	-,074
	Sig. (2-tailed)		,062
	N	637	637
Sociale integratie mean	Pearson Correlation	-,074	1
	Sig. (2-tailed)	,062	
	N	637	637

The correlation between **Network Diversity in terms of education and Social Integration** is measured by the Pearson's Correlation method.

```
CORRELATIONS
  /VARIABLES=Network_Diversity
Social_integratie_Mean
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.
```

Correlations

		Netwerk diversiteit	Sociale integratie mean
Netwerk diversiteit	Pearson Correlation	1	,105**
	Sig. (2-tailed)		,008
	N	637	637
Sociale integratie mean	Pearson Correlation	,105**	1
	Sig. (2-tailed)	,008	
	N	637	637

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

Hypotheses testing

To test the hypotheses, a linear regression was done. This analysis was done as described in the analysis plan in the Methods paragraph.

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Sociale_integratie_Mean
  /METHOD=ENTER income leeftijd Primary_occupation
  /METHOD=ENTER Network_Diversity_centr
  /METHOD=ENTER Gender
  /METHOD=ENTER NDGender
  /SCATTERPLOT=(*ZRESID ,*ZPRED)
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /SAVE PRED ZPRED ADJPRED COOK LEVER RESID ZRESID DFBETA DFFIT.
```

Model Summary[e]

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	,138[a]	,019	,014	,43858	,019	4,082	3	633	,007
2	,173[b]	,030	,024	,43648	,011	7,116	1	632	,008
3	,173[c]	,030	,022	,43682	,000	,015	1	631	,903
4	,173[d]	,030	,021	,43715	,000	,030	1	630	,863

a. Predictors: (Constant), Primary occupation, Age of the household member, income

b. Predictors: (Constant), Primary occupation, Age of the household member, income, Network_Diversity_centr

c. Predictors: (Constant), Primary occupation, Age of the household member, income, Network_Diversity_centr, Gender, 0:female, 1: male

d. Predictors: (Constant), Primary occupation, Age of the household member, income, Network_Diversity_centr, Gender, 0:female, 1: male, NDGender

e. Dependent Variable: Sociale integratie mean

The image above shows that the adjusted R square scores are varied. So shows model 1 an adjusted R square of 0,014, however in model 2 the adjusted R square increased to a score of 0,024. This means that when network diversity in terms of education was added, more variance was explained. However, in both model 3 (R_a^2=0,022) and 4

(R_a^2=0,21) has not changed much compared the score on the adjusted R square in model 2. So both the moderator gender and the interaction effect of network diversity in terms of education and gender, did not increase the explained variance in the models.

In the image above, we can see that model 2 (F change = 7,116; p = 0,008) shows a significant improvement compared to model 1 (F change = 4,082; p = 0,007) when looking at the F change. This means that when including network diversity in terms of education there is an increasement in the fit of the model on the data. The F change in model 3 (F change = 0,015; p = 0,903) and model 4 (F change = 0,030; p = 0,863) decrease and show no more significant p-values. This means that when gender and the interaction-effect there is no increase in the fit of the model on the data.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	2,603	,056		46,637	<,001					
	income	,053	,018	,134	2,928	,004	,123	,116	,115	,735	1,360
	Age of the household member	-,002	,001	-,059	-1,390	,165	-,025	-,055	-,055	,849	1,177
	Primary occupation	-,011	,040	-,012	-,272	,786	-,074	-,011	-,011	,773	1,293
2	(Constant)	2,599	,056		46,771	<,001					
	income	,054	,018	,136	2,982	,003	,123	,118	,117	,735	1,360
	Age of the household member	-,002	,001	-,058	-1,373	,170	-,025	-,055	-,054	,849	1,177
	Primary occupation	-,007	,039	-,008	-,189	,850	-,074	-,008	-,007	,773	1,294
	Network_Diversity_centr	,031	,012	,105	2,668	,008	,105	,106	,105	,999	1,001
3	(Constant)	2,598	,056		46,141	<,001					
	income	,053	,019	,135	2,821	,005	,123	,112	,111	,675	1,481
	Age of the household member	-,002	,001	-,058	-1,362	,174	-,025	-,054	-,053	,846	1,181
	Primary occupation	-,008	,040	-,009	-,199	,842	-,074	-,008	-,008	,767	1,305
	Network_Diversity_centr	,031	,012	,105	2,666	,008	,105	,106	,105	,991	1,009
	Gender, 0:female, 1:male	,004	,037	,005	,121	,903	,031	,005	,005	,908	1,101
4	(Constant)	2,598	,056		46,100	<,001					
	income	,053	,019	,134	2,812	,005	,123	,111	,110	,675	1,482
	Age of the household member	-,002	,001	-,058	-1,363	,173	-,025	-,054	-,053	,846	1,182
	Primary occupation	-,008	,040	-,009	-,204	,838	-,074	-,008	-,008	,766	1,306
	Network_Diversity_centr	,032	,015	,111	2,170	,030	,105	,086	,085	,593	1,686
	Gender, 0:female, 1:male	,004	,037	,005	,118	,906	,031	,005	,005	,908	1,102
	NDGender	-,004	,024	-,009	-,173	,863	,060	-,007	-,007	,593	1,686

a. Dependent Variable: Sociale integratie mean

As explained in the analysis plan, only the control variables were included in the **first model**. Here only the variable income showed a significant effect. The rest of the variables show no significant effect.

The **second model** included the independent variable network diversity in terms of education. When including the independent variable, the effect of the control variables almost

all stay exactly. The effect of network diversity in terms of education is small and nearly significant.

The **third model** includes the moderator gender, which shows no significant effect. The other variables also do not change drastically when adding gender. Further conclusion on this can be read in the Result paragraph.

The **fourth model** includes the interaction effect, which is not significant. Network diversity in terms of education and gender shows different scores compared to model three. This is a logical reaction because the interaction is made out of these two variables. However, the control variables show again almost no changes.

Binary Logistic Analysis

Dependent variable: Social integration of immigrants
Original variables

The dependent variable social integration of immigrants contains six questions, which all asked about the level of life satisfaction and loneliness. These are the following six variables: 'I have a sense of emptiness around me', 'I miss having people around me', 'I often feel deserted', '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' and 'there are enough people to whom I feel closely connected'. In the syntax and the following tables, the distribution of all items are shown.

```
FREQUENCIES VARIABLES=fb14b307 fb14b308 fb14b309 fb14b310
fb14b311 fb14b312
  /ORDER=ANALYSIS.
```

I have a sense of emptiness around me

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	30	4,7	4,7	4,7
	more or less	143	22,3	22,3	27,0
	no	467	73,0	73,0	100,0
	Total	640	100,0	100,0	

there are enough people I can count on in case of a misfortune

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	436	68,1	68,1	68,1
	more or less	168	26,3	26,3	94,4
	no	36	5,6	5,6	100,0
	Total	640	100,0	100,0	

I know a lot of people that I can fully rely on

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	345	53,9	53,9	53,9
	more or less	214	33,4	33,4	87,3
	no	81	12,7	12,7	100,0
	Total	640	100,0	100,0	

there are enough people to whom I feel closely connected

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	405	63,3	63,3	63,3
	more or less	178	27,8	27,8	91,1
	no	57	8,9	8,9	100,0
	Total	640	100,0	100,0	

I miss having people around me

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	69	10,8	10,8	10,8
	more or less	146	22,8	22,8	33,6
	no	425	66,4	66,4	100,0
	Total	640	100,0	100,0	

I often feel deserted

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	30	4,7	4,7	4,7
	more or less	73	11,4	11,4	16,1
	no	537	83,9	83,9	100,0
	Total	640	100,0	100,0	

We also looked at the opportunity to include an item about social contacts, an important aspect of social integration. The item would be: ‘How satisfied are you with your social contacts?’ However, we decided not to implement this item when looking at Cronbach's alpha. The table of Cronbach’s alpha is shown below.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
How satisfied are you with your social contacts?	15,6125	7,011	,014	,806
I have a sense of emptiness around me	43,1750	22575,325	,029	,001
there are enough people I can count on in case of a misfortune	43,2328	22588,141	-,045	,001
I know a lot of people that I can fully rely on	43,4453	22571,631	,039	,000
there are enough people to whom I feel closely connected	43,3141	22574,385	,029	,001
I miss having people around me	43,3016	22572,230	,038	,000
I often feel deserted	43,0656	22577,980	,014	,001

Recodings

Respondents could answer these questions with (1) yes, (2) more or less, and (3) no. To have all questions in the same direction, three questions (fb14308, fb14308=9, fb14310) were mirrored, where (1) no, (2) more or less, and (3) yes. Subsequently, the six questions were combined by making a mean variable. The new variable Sociale_integratie_Mean ranges from 1 to 3, with a higher score indicating a higher feeling of social integration.

```
RECODE fb14b308 fb14b309 fb14b310 (1=3) (2=2) (3=1) INTO
fb14b308_Gespiegeld fb14b309_Gespiegeld fb14b310_Gespiegeld.

VARIABLE LABELS  fb14b308_Gespiegeld 'there are enough people I
can count on in case of a '+ 'misfortune' /fb14b309_Gespiegeld 'I
know a lot of people that I can fully rely on'
/fb14b310_Gespiegeld 'there are enough people to whom I feel
closely connected'. EXECUTE.
```

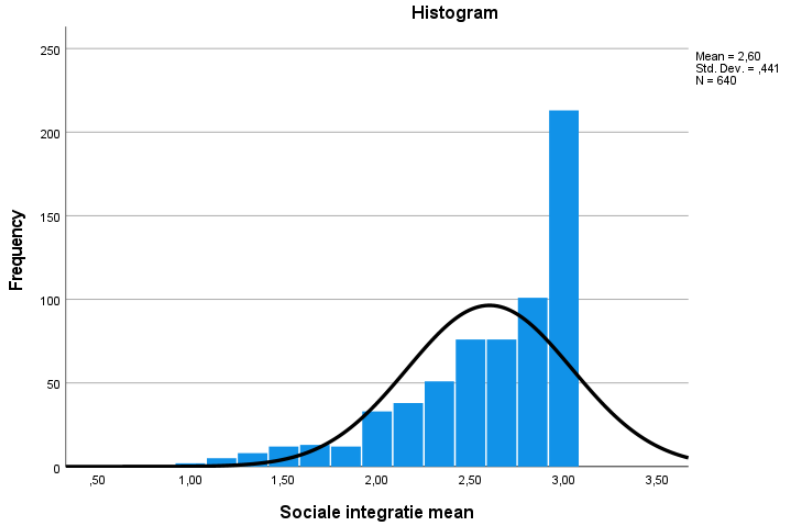
```
COMPUTE
Sociale_integratie_Mean=MEAN(fb14b307,fb14b308_Gespiegeld,fb14b309
_Gespiegeld, fb14b310_Gespiegeld,fb14b311,fb14b312).

EXECUTE.
```

```
RECODE Sociale_integratie_Mean (1.00 thru 2.49=0) (2.50 thru
3.00=1) INTO Social_Integration_Dummy. VARIABLE LABELS
Social_Integration_Dummy 'Dummy Social integration'. EXECUTE.
```

```
FREQUENCIES VARIABLES=Sociale_integratie_Mean
  /STATISTICS=STDDEV VARIANCE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /ORDER=ANALYSIS.
```

Sociale integratie mean

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	2	,3	,3	,3
	1,17	5	,8	,8	1,1
	1,33	8	1,3	1,3	2,3
	1,50	12	1,9	1,9	4,2
	1,67	13	2,0	2,0	6,3
	1,83	12	1,9	1,9	8,1
	2,00	33	5,2	5,2	13,3
	2,17	38	5,9	5,9	19,2
	2,33	51	8,0	8,0	27,2
	2,50	76	11,9	11,9	39,1
	2,67	76	11,9	11,9	50,9
	2,83	101	15,8	15,8	66,7
	3,00	213	33,3	33,3	100,0
	Total	640	100,0	100,0	



As this variable was not normally divided at all, we decided to make a dummy of the mean variable and use a logistic regression. To make this distribution meaningful we agreed that the scores of 1.00 to 2.49 on the mean variable, were coded as (0) no or little feeling of social integration. The scores of 2.50 to 3.00 were coded as (1) high feeling of social integration. Due to dichotomizing the social integration variable, we can argue that quite some information was lost.

Hypotheses testing binary logistic analysis

Here we show the results of the binary logistics analysis.

```
LOGISTIC REGRESSION VARIABLES Social_Integration_Dummy
  /METHOD=ENTER Primary_occupation income leeftijd
  /METHOD=ENTER Network_Diversity
  /METHOD=ENTER Gender
  /METHOD=ENTER NDGender
  /SAVE=DEV
  /PRINT=GOODFIT
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	637	100,0
	Missing Cases	0	,0
	Total	637	100,0
Unselected Cases		0	,0
Total		637	100,0

a. If weight is in effect, see classification table for the total number of cases.

As explained in the analysis plan, only the control variables were included in the **first model**. No significant effects are found in this model.

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3,857	3	,277
	Block	3,857	3	,277
	Model	3,857	3	,277

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	356,352[a]	,006	,014

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	14,096	8	,079

Classification Table[a]

			Predicted		
			Dummy Social integration		Percentage Correct
	Observed		,00	1,00	
Step 1	Dummy Social integration	,00	0	52	,0
		1,00	0	585	100,0
	Overall Percentage				91,8

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Primary occupation	,043	,048	,821	1	,365	1,044
	income	,343	,190	3,260	1	,071	1,410
	Age of the household member	-,012	,010	1,410	1	,235	,988
	Constant	2,353	,467	25,347	1	<,001	10,518

a. Variable(s) entered on step 1: Primary occupation, income, Age of the household member.

The **second model** included the independent variable network diversity in terms of education. It shows no significant effect of the relationship between network diversity in terms of education and the social integration of immigrants; this is shown in model 2 when looking at the slope and the odds of network diversity in terms of education (B= 0,057; odds ratio= 1,058; p=0,561). The odds ratio indicates that the odds increase by 1,058 times for having higher network diversity in terms of education than having a low score on network diversity in terms of education. This means that the chance of having a high feeling of social integration

is significantly greater among respondents who scored high on network diversity in terms of education. This effect is not significant at a significance level of 0,05. When adding the independent variable network diversity in terms of education, other effects of the control variables did not change dramatically. As income, age and primary occupation still showed no significant effects, with similar odds ratios.

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,340	1	,560
	Block	,340	1	,560
	Model	4,198	4	,380

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	356,012[a]	,007	,015

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	2,908	8	,940

Classification Table[a]

			Predicted		
			Dummy Social integration		Percentage Correct
	Observed		,00	1,00	
Step 1	Dummy Social integration	,00	0	52	,0
		1,00	0	585	100,0
	Overall Percentage				91,8

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Primary occupation	,045	,048	,894	1	,344	1,046
	income	,345	,190	3,296	1	,069	1,412
	Age of the household member	-,013	,010	1,429	1	,232	,988
	Netwerk diversiteit	,057	,098	,337	1	,561	1,058
	Constant	2,237	,506	19,554	1	<,001	9,363

a. Variable(s) entered on step 1: Netwerk diversiteit .

The **third model** includes the moderator gender, which shows no significant effect. It also shows for the moderator no significant effect. Table 5 shows that the slope and odds-ratio of gender in model 3 are not significant (B = 0,373; odds ratio= 1,453; p=0,253). The odds ratio indicates that men feel a stronger sense of social integration than women. When looking at other variables used in model 3, the effect on social integration are still not significant.

Block 3: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	1,350	1	,245
	Block	1,350	1	,245
	Model	5,547	5	,353

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	354,662[a]	,009	,020

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	4,523	8	,807

Classification Table[a]

			Predicted		
			Dummy Social integration		Percentage Correct
	Observed		,00	1,00	
Step 1	Dummy Social integration	,00	0	52	,0
		1,00	0	585	100,0
	Overall Percentage				91,8

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Primary occupation	,041	,048	,749	1	,387	1,042
	income	,283	,195	2,090	1	,148	1,326
	Age of the household member	-,012	,011	1,205	1	,272	,989
	Netwerk diversiteit	,067	,098	,462	1	,497	1,069
	Gender, 0:female, 1:male	,373	,327	1,308	1	,253	1,453
	Constant	2,138	,512	17,457	1	<,001	8,484

a. Variable(s) entered on step 1: Gender, 0:female, 1:male.

The **fourth model** includes the interaction effect, which is not significant. It again shows no significant effect for the added interaction variable (B = -0,263; odds ratio= 0,753; p=0,177). As the interaction variable is not significant, we cannot make conclusions about the second hypothesis.

The binary logistic analysis shows no significant effects for our main effects. So, in this case, the two hypotheses cannot be accepted. We have to mention that due to dichotomizing the social integration variable, we can argue that quite some information was lost. This can arise question about the power of the models.

Block 4: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	1,810	1	,178
	Block	1,810	1	,178
	Model	7,357	6	,289

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	352,852[a]	,011	,027

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	12,329	8	,137

Classification Table[a]

			Predicted		
			Dummy Social integration		Percentage Correct
	Observed		,00	1,00	
Step 1	Dummy Social integration	,00	0	52	,0
		1,00	0	585	100,0
	Overall Percentage				91,8

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Primary occupation	,041	,048	,730	1	,393	1,042
	income	,286	,197	2,098	1	,148	1,331
	Age of the household member	-,012	,011	1,292	1	,256	,988
	Netwerk diversiteit	,157	,121	1,688	1	,194	1,170
	Gender, 0:female, 1:male	,346	,327	1,119	1	,290	1,414
	NDGender	-,283	,210	1,821	1	,177	,753
	Constant	1,991	,517	14,807	1	<,001	7,322

a. Variable(s) entered on step 1: NDGender.

We used the Hosmer-Lemeshow test, Deviance, and classification tables for the model fit. The **Hosmer-Lemeshow test** show if the models fit with the data used in the analysis. If the p-value is below the 5% (0,05), it indicates a poor fit of the model. The X^2 values for all models, model 1 (X^2= 14,096; p= 0,079), model 2 (X^2= 2,908; p=0,940), model 3 (X^2 = 4,523; p=0,807) and model 4 (X^2= 12,329; p= 0,137) are above the 0,05. These values show that the observed values of social integration do not significantly deviate from the predicted values of social integration. This indicates no evidence that the models poorly fit with the used data in the analysis.

The **Deviance** decreases with each model as more variables are added to the models. This decrease indicates a reduction of errors, which means an increase in the quality of the models.

The **classification table** shows the percentage of well-predicted observations in the analysis; for all models applied that the models predict 91,8% of the observations correctly. However, because the score is the same in all models, it does not show a significant improvement in the models.

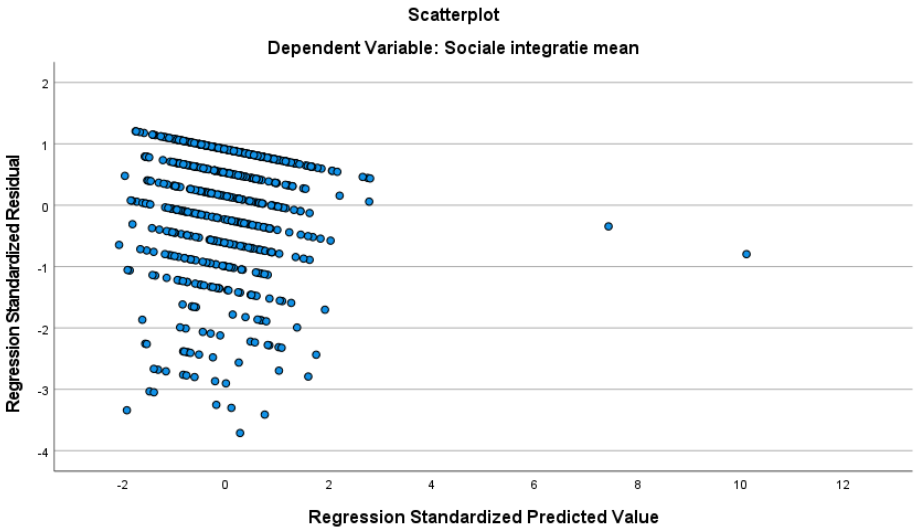
Appendix 3: Assumption testing and outlier analysis

Assumptions

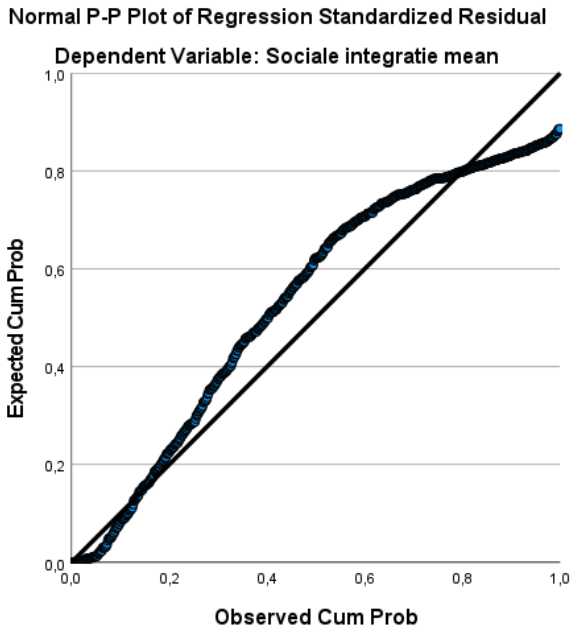
When doing a linear analysis, there are assumptions that need to be checked. The first one, is the use of independent observations. This means that the answers given by the respondents are independent of each other. This dataset works with households, so in theory, it is possible that several respondents from a household were asked. As a result, the answers given may not be independent of each other. Unfortunately, we cannot completely rule this out; as a result of which, this assumption has been violated.

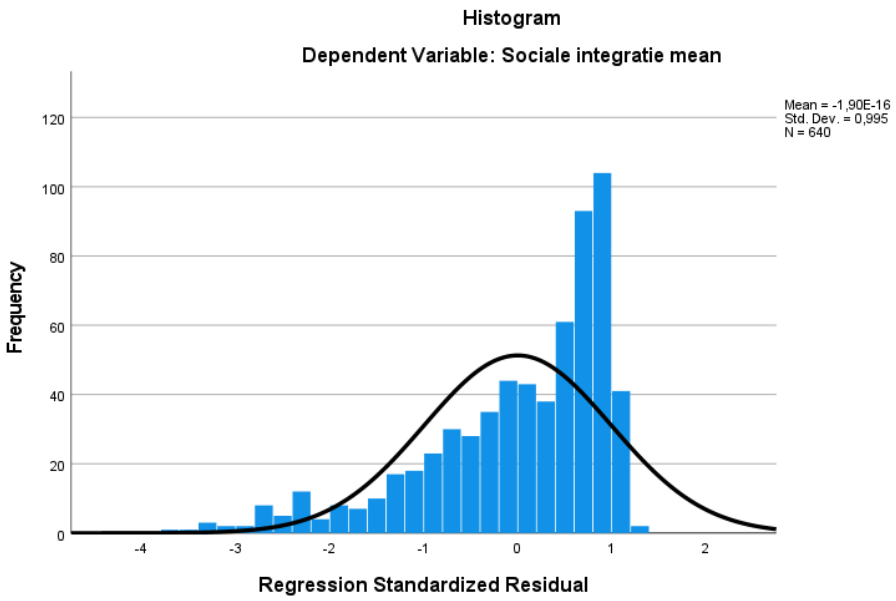
A second assumption is that of linearity. Where there must be a linear relationship between the independent and the dependent variable. This can be tested by creating a scatter plot of the residuals and the expected values of social integration of immigrants, shown in the figure down below. To establish a linear relationship, the residuals need to be situated around the zero line. In the figure we can see that this is not the case. That is why this assumption is violated.

The third assumption concerns homoscedasticity. This assumption is satisfied when the dispersion of the dependent variable is the same for all data. This also can be tested by creating a scatter plot of the residuals and the expected values of social integration of immigrants. In the figure below the dispersion of the residuals is not even. It can therefore be argued that this assumption has not been violated.



The last assumption is normality. This assumption is met if the residuals are normally distributed. We can test this with a histogram of the standardized residuals of the dependent variable social integration and a PP-plot of the standardized residuals of the dependent variable social integration. In these figures below it is shown that the residuals of the dependent variable social integration of immigrants are not normally distributed. There seems to be a left skewed distribution. This assumption has therefore been violated.





Multicollinearity

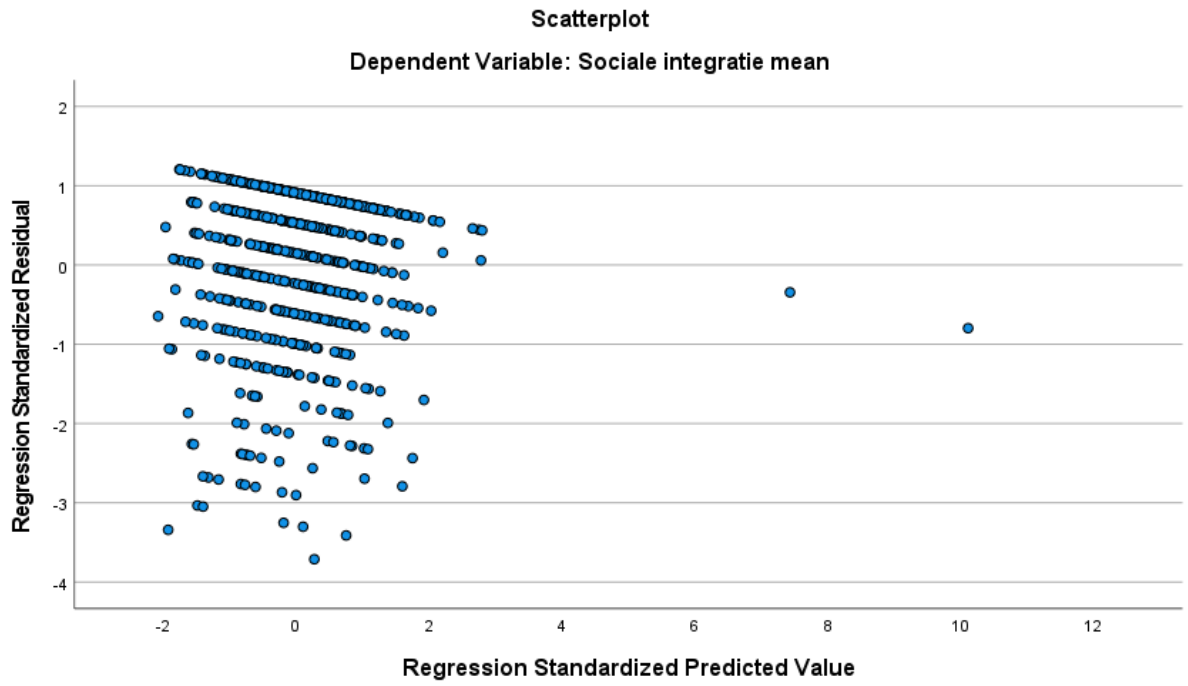
Multicollinearity shows the interrelationship between the variables. This can be calculated using the ViF values. These values are shown in Table 3. If the ViF score is below the score of 4, it can be assumed that there is no significant correlation between the variables. In this analysis, no ViF values scored above the value of 4, so there is no evidence of an interrelationship between any of the variables.

Outliers

To gain insight into the influence of outliers, several measures were examined, namely the standardized residuals, the leverage, and the Cook's Distance.

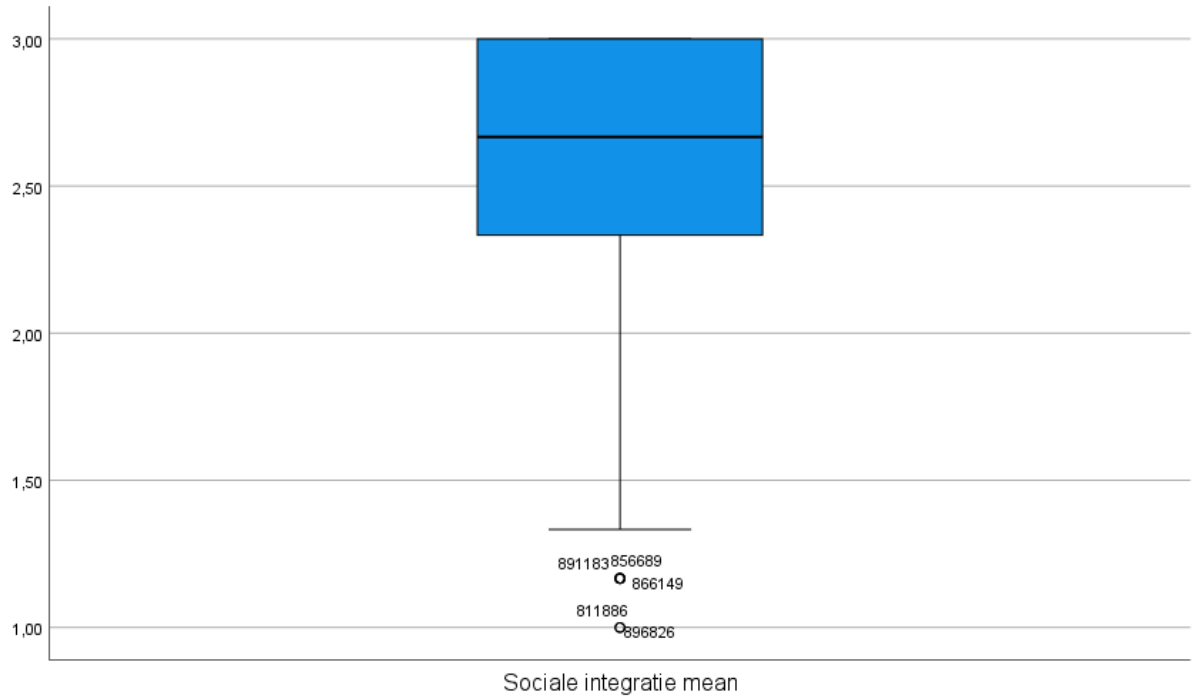
First of all, the standardized residuals are looked at. If a point is not between the values -3 and 3, it is considered an outlier. The figure below shows seven outliers on the below -3 in the figure. The leverage indicates the extent to which a point is influential on the direction of the model. This is calculated with the formula (3 x p)/n= (3x7)/640 = 0,033. According to the leverage, there are four outliers. The Cook's Distance is a composite measure of the standardized residuals and leverage. The Cook's Distance can be calculated by 4/n=4/640=0.00625. According to Cook's Distance, there are 31 outliers. The maximum tolerance for DFFIT is (2x(7/640)= 0,021875. The DFFIT shows how influential a point is in a

regression. There are 3 points that exceed this score of the DFFIT.



The case with the highest Cook's distance is 872341, with a Cook's distance of 0,17509. This case also has high values on both the leverage (0,49266) and DFFIT (-0,34022), which shows that case 872341 is an influential point. However, the residual is within the boundaries of 3 and -3, so it is not an outlier. This is also applies to case number 826263. Where the Cook's distance is 0,00784, the leverage is 0,25477 and the DFFIT is -0,05185. However, the residual is within the boundaries of 3 and -3, so it is not an outlier. So both case number 872341 and 826263 are influential points, but not outliers.

The cases 891183, 840057, 856689, 896826, 811886, 866149 and 857302 all have a high Cook's distance and do not fit in between the boundaries of 3 and -3 when looking at the residuals. Both the leverage and DFFIT scores of these points are not exceeded. We included an boxplot down below to shown the effect of these outliers. These outliers all score extremely low scores on social integration, which can be an explanation on why these values are outliers. Deleting these cases will not cause a poorer representation of the population. Therefore, these cases will not be deleted.



Case number 895016 shows a high leverage score of 0,03488, however it does not show high scores on the Cook's distance and DFFIT. It does found out of the boundaries of the leverage of 3 and -3. This case is an outlier, but is not very influential.

Lastly, case number 879602 also has a high Cook's distance of 0,01959 and a high score on the DFFIT of -0,02406. It, however, does not show exceptional score on the leverage or residuals methods. This case might be influential and have no outliers. However, when a predicted value would change with -0,02406 if this case was not in the analysis. This score is still low when looking at the scale of 1-3 of the social integration variable. So this point does not seem to be a problem.

In summary, cases 895016, 891183, 840057, 856689, 896826, 811886, 866149, and 857302 are outliers and the cases 872341 and 826263 are seen as influential cases. We decided on only deleting case numbers 890516, 872341 and 826263.

```
USE ALL.

COMPUTE filter_$=(nomem_encr ~= 895016 & nomem_encr ~= 872341 &
nomem_encr ~= 826263).

VARIABLE LABELS filter_$ 'nomem_encr ~= 895016 & nomem_encr ~=
872341 & nomem_encr ~= 826263 '+

 '(FILTER)'.

VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.

FORMATS filter_$ (f1.0).

FILTER BY filter_$.

EXECUTE.
```

```
REGRESSION

  /DESCRIPTIVES MEAN STDDEV CORR SIG N

  /MISSING LISTWISE

  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE ZPP

  /CRITERIA=PIN(.05) POUT(.10)

  /NOORIGIN

  /DEPENDENT Sociale_integratie_Mean

  /METHOD=ENTER income leeftijd Primary_occupation

  /METHOD=ENTER Network_Diversity

  /METHOD=ENTER Gender

  /METHOD=ENTER NDGender

  /SCATTERPLOT=(*ZRESID ,*ZPRED)

  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID)

  /SAVE PRED ZPRED ADJPRED COOK LEVER RESID ZRESID DFBETA DFFIT.
```