# THE INTEGRATION OF IMMIGRANT CHILDREN INTO THE ITALIAN SOCIETY: DOES PARENTAL EDUCATION MATTER? 

Bachelor Thesis<br>Cedric Chambrillon<br>429064<br>Erasmus University Rotterdam, Erasmus School of Economics

July $12^{\text {th }}, 2018$


#### Abstract

The degree of integration of immigrants into the majority group's culture is the result of various personal and societal characteristics. Not only are the social and psychological implications of high interest, the economic consequences of the integration of immigrants become more and more meaningful. Due to our current political context, immigrants are a constantly growing share of the European population. In this research I investigate if parental education affects the integration of immigrant children. I use the amount of friends and the language spoken at home of the children as measures of integration. Through analyzing historical backgrounds and antagonistic findings of the literature, I predict that a higher parental education engenders a higher amount of friends and a higher probability of speaking Italian at home for the child. This paper provides a summary of previous inquiries by exploring a broad literature, discussing the "Immigrant paradox" and showcasing the obtained predictions in the case of Italian $5^{\text {th }}$ grade primary education students.


## 1. Introduction

The start of the European Refugee crisis in 2015 marked the beginning of countless debates and controversy amongst politicians and policy makers on how to solve the many problems this crisis brought up. Among these problems, the successful integration of migrants into the majority group's culture is key in the process towards a viable society. At times integration is even a prerequisite to immigration in certain countries. This paper aims to analyze the effect of parental education on integration of immigrant children in Italy.

The recent increase in immigration across the globe has been a tough challenge for economists trying to understand the mechanisms behind this population transfer and politicians looking for adequate ways to handle it. Similarly, the notion of "Economics of Immigration", which appeared around the forties (Cohen, 1939), has rapidly grown in popularity and became a separate economic discipline in such. Recently, Borjas (1996) renamed this as the "New Economics of Immigration". The effects of these new immigration flows have been studied by a broad literature (Chiswick, 2005; Dustmann et al, 2005; Mahuteau et al, 2008; Bodvarsson et al, 2009). It has been found that in and outflow of employees have a substantial impact on the concerned economies and societies: immigration increases demand, affects the labor market and increases native employment (Ottaviano et al, 2013), increases the human capital stock (Akcigit et al, 2017), increases wage inequality (Card, 2009), reduces wages (Borjas, 2003) and increases productivity (Borjas 1995). While immigration has a cost (an increase in social program expenditures), Borjas (1995) describes this cost as an "investment that has a very high rate of return".

Theodore Roosevelt once said: "Every immigrant who comes [to the United States] should be required within five years to learn English or leave the country" displaying the value societies put on the integration of their immigrants into the majority group. Not only is the integration of the minority crucial for a cohesive society (Allport, 1954), it also is essential for their contribution to the labor market and the facilitation of inter-ethnic relations and decreasing cultural conflicts. The literature also showed that integrated immigrants have access to better employment, more career opportunities (Lin, 1999), and higher incomes (Chiswick, 1991; Dustmann, 1994; Chiswick and Miller, 1992).

Out of all the possible personal, cultural, and structural characteristics that impact the extent and speed of integration of migrants into a society, education has received substantial and
controversial attention from researchers. The term "immigrant paradox" has been widely used to describe the antagonistic findings encountered in the literature (see Sam et al., 2008; van Doorn, 2013; Tolsma et al., 2012). This paradox illustrates the fact that higher educated migrants integrate better and thus have more contact with the native culture. But as a consequence of being more exposed to national politics, participating in associations they percieve more personal discrimination.

While the "immigrant paradox" showcases the ambiguous effect education has on the integration of the immigrants, the "status paradox" describes the effects of education on the attitude of the minority towards the majority. The status paradox allows us to introduce the term of "moralism". Black's theory of moralism (1993) proposes that the social status of immigrants, which is here only defined by their education and job occupation, "influences the amount of conflicts [they] experience as well as how they respond to it" (Cooney, 2009). This theory advocates that higher status migrants respond to hostility, while lower status migrants isolate themselves and avoid conflicts. Both paradoxes discussed are thus complementary and can be synthetized as follows: better educated individuals integrate better and are more active in social life while less educated individuals isolate. As a consequence of being more active, the better educated immigrants are more exposed to discrimination. Through this research we investigate the attitude of immigrants towards the majority group through analyzing their habit to speak Italian at home and amount of peers. We believe that both variables are influenced by the attitude of immigrants towards natives.

Previous research found that education affects immigrants' attitude towards natives in two opposite ways. Better educated immigrants have more frequently contact with the majority group (e.g., Martinovic et al., 2009) while being more perceptive of hostility (ten Teije et al., 2013). The latter study was performed in the Netherlands on adult migrants, with poor measures of integration such as education level or professional activity. In this research we use Cooney's definition of social integration as the participation in social life (2009). Data from the Italian INVALSI allows us to have two measures that estimate the integration of immigrant children: the language spoken at home and the quality of the relationships with other students in the classroom. This dataset gives us the opportunity to look at social integration. This paper aims to answer the following question: does parental education affect the integration of their 5th grade immigrated children in Italy? This
question tries to answer the first part of the paradox previously mentioned. The second part of the paradox about the perception of discrimination will not be researched because of data limitations. Furthermore, perceived discrimination is hard to measure because of the young age of the children. Our first hypothesis states that higher education of parents increases the amount of friends an immigrated child has. Our second hypothesis is that having higher educated parents increases the probability of speaking Italian at home.

Section 2 of this paper describes the immigration particularities of Italy. Section 3 presents the Italian education system. The fourth section justifies the use of the variables to measure integration. Section 5 introduces the dataset used for the research while section 6 showcases the methodology used to obtain our results. Section 7 displays our findings and section 8 provides robustness checks. We discuss the implications of our findings on policy-making and future research in section 9 . Section 10 concludes.

## 2. Immigration in Italy

Italy is a very popular destination among migrants for its proximity to the Balkans, the African continent and lying on the Mediterranean Sea. With over five million immigrants within its borders, Italy is the 3rd European country with most immigrants behind Germany and the United Kingdom (OECD, 2018). The majority of immigrants in Italy come from Albania, Romania and Morocco. Together these three countries represent over $40 \%$ over the total migrant population (ISTAT, 2017). These numbers do not include refugees, mainly coming from Libya, who are illegally on the territory but still account for close to a million individuals.

This mass immigration is a recent phenomenon mainly caused by Italy opening its borders while other European countries restricted entries into their borders around the eighties. Before this Italy was mainly a country of mass emigration and internal migration. Between 1876 and 1976, 24 million Italians decided to emigrate. The notion of Italian diaspora has become universally accepted among historians and economists to describe this episode. The Italian economic boom also known as "economic miracle" from 1960 until 1973 marked the beginning of a period of positive migration balances. Not only foreigners entered the country, but also internal migration from the poorer South to the more developed Northern regions was initiated.

Overall, Italy is a very interesting case study for immigration, not only because of its past, but also because of the diversity of the immigrants, consisting of both economic refugees and asylum-seekers flowing in from different continents.

## 3. The Italian education system

Scholastic institutions provide formal education, and offer equal opportunities to acquire basic and identical knowledge for all students, this is not their only purpose. They also offer nonformal and informal education as described by a broad literature (see Ainsworth and Eaton, 2010). Schools and teachers allow students to develop fundamental skills and capacities through different projects, a process which is often referred to as non-formal education. In addition to knowledge, skills and capacities, schools also teach children about explicit and implicit social norms that are neither rewarded by a diploma nor purposely guided by a curriculum. This teaching is labelled unformal education and particularly important for immigrants. Schools provide a favorable environment and are an influential tool for integration of the immigrants into a new culture and therefore extremely important and interesting to study.

The Italian education system is comparable to any other European education system. Schooling is compulsory from age 6 until 16. Our study group falling in this category allows us to assume that all Italian and registered immigrant children are enrolled in a school. Education is free and all residents can access education regardless of their nationality. Assuming that not all children of the age of compulsory education are enrolled in a school, my research will focus on enrolled immigrants only, thus excluding the illegal immigrants and asylum-seekers discussed previously. Furthermore, both private and public education exist in Italy, but we control for these differences using school fixed effects which we will discuss later on. These school fixed effects also allow to take into account funding differences, since public schools are funded by the government. The Italian public education is considered to be better than the private education (Bertola, 2007).

## 4. Measures of integration

In previous research integration has been defined only by economic integration, i.e. the professional activity of immigrants. Since professional activity is highly correlated with education and because we want to analyze the effect of education on integration we cannot use type of job of the parents as a measure for economic integration because of multicollinearity concerns. Integration of immigrants is hard to measure. Previous research made the mistake to only consider economic integration, and fails to assert that integration of immigrants, as an important feature for society's cohesion, is also represented by social integration, participation in social life and language mastering. This research adds to the literature by providing analysis of precise and relevant measures revealing habits and social life characteristics of immigrants. The dataset provides us with two good measures of integration: the language spoken at home and the quality of the relationships of the child in the classroom. From our definition of integration as participation in social life, we derive that speaking the majority group's language in a household setting instead of using its own native language is a strong proof of integrated individuals. Next to this, having good friendships with its classmates displays the ability to accept cultural differences and cope with them, as well as the willingness to put effort into participating in social life.

Baumgartner (1998) states that in today's society immigrants are free to choose their peers and do not have to forge friendships with natives if they find their conduct unappealing. In addition to this, if immigrants find native's "behavior offensive, unpredictable or otherwise questionable, they are engaging in avoidance" (Baumgartner, 1998). This allows us to draw the conclusion that if immigrants report having good relationships with their classmates, this is solely due to personal taste and the share of values and morals. There is thus no perceived obligation to be friendly towards the majority group that affects our variable, the quality of the relationships. This variable thus only measures the degree at which immigrant children accept and enjoy the Italians as supported by Baumgartner (1998). If this would not hold and social pressure would push immigrant children to socialize with natives, this variable would not be a relevant measure of integration anymore. Finally, previous research found a high correlation between speaking the majority group's language and having more contact with them (Fong and Isajiw, 2000; Weijters and Scheepers, 2003). This implies that households speaking the country's official language participate more in social life and therefore we consider it as an adequate measure of integration. More than a
measure of integration, the literature also found that immigrants who speak the official language have higher incomes (Chiswick, 1991; Dustmann 1994, Chiswick and Miller, 1992) which motivates studying language. Since we look at the language spoken at home and not Italian language skills in general, we cannot draw conclusions on the effects of education on language skills. Even though, previous findings are interesting to keep in mind throughout this research.

The quality of the relationships is reported by the children themselves answering the question: "with how many of your classmates do you get along well". It is important to note that this measure looks at whether a child gets along well with its classmates and not if they are pure friends. Friendships are not a direct consequence of how the child fits the group and are highly influenced by the degree of socialization of the child which is the result of inherent characteristics instead. Some children are more introvert and not inclined to have a lot friendships. Another way to look at integration would be to look specifically at the amount of friendships one has with native children. Our data does not allow this and therefore our results could be biased by the fact that one child might have a lot of friends in its classroom but at the same time be friends only with fellow immigrants and not with natives, which would still be interpreted as well integrated following our methodology. We do not consider this as a threat since the average amount of immigrant children per classroom remains low and amounts to 2.29 immigrants per classroom.

## 5. Data

To analyze the effect of parental education on the integration of their children, I use administrative data from the National Institute for the Evaluation of the Education system (INVALSI). This institute tests mathematical and Italian language knowledge of all Italian 5th grade students. All students need to take the test, which is held every year at the end of the academic year. The Mathematical test consists of both logical reasoning and calculation problems, while the Italian test questions about grammar, vocabulary and reading comprehension. Next to the test scores, the INVALSI also collects survey data from all students, asking for and about personal information, family composition and parental characteristics, how the students felt about the tests, parental and peer pressure, household belongings and school and classroom atmosphere.

We analyze the period from academic year 2011-2012 to 2016-2017. In this research, using the academic year 2013-2014 as example, we refer to this academic year as year 2013. Regarding the first hypothesis, data limitations only allow us to use the years 2013 and 2014. Since we only study two years, this variable is a dummy variable and displays values $0=2013$ and $1=2014$. The second hypothesis will be investigated using the years 2011 until 2016 included. Here, the time variable will be a set of dummies and will have as starting value $0=2011$.

Table 1 shows the breakdown of the dataset into observations for the analysis. The data was available from two different datasets, both from the INVALSI. The first one called "Italian test scores" reveals information about the student and its parents, the second one named "Survey" displays the amount of friends, the language spoken at home and other self-reported information about the daily activities and school impressions of the student. Table 1 shows the amount of observations lost by the merging of the datasets for hypothesis 1 .

Considering the years 2013 and 2014, our data consists of $28.72 \%$ first generation and $71.28 \%$ second generation immigrants for a grand total of 56513 observations. Following the INVALSI, a first generation immigrant is defined as being born in a foreign country as well as its parents whereas the second generation immigrant is described as being born in Italy while its parents were born abroad. Around $24 \%$ of the immigrants come from European Union countries, $26 \%$ non-European Union countries in the Europe area (for example Albania, which is the largest migrant community in Italy (ISTAT, 2017)), $46 \%$ come from non-European countries. Out of all subjects, $38.63 \%$ speak Italian at home or an Italian dialect. Sex and year of enrollment in 5th grade are evenly distributed. There are more highly educated (i.e. pertaining to categories 5 and 6) mothers than fathers as table 3 shows, $15.59 \%$ and $13.30 \%$ respectively. Category 1 represents the lowest educated share of the parents while category 6 reports the highest educated share of parents. As discussed below, each classroom counts on average 2.29 immigrants. The average total amount of students is 19.16 per classroom. Including all the years from 2011 to 2016, we count a total amount of classrooms of 28,978.

We need to consider the fact that there exist countries that have a large Italian speaking population. Such countries are Malta, Romania and Albania. In Malta and Albania, over 60\% of the inhabitants speak Italian. Following the Romanian National Institute of Statistics (2018), over one-and-a-half million Romanians speak Italian. We cannot check if the household already spoke Italian before entering the country, but we can see from table 8 that $26,69 \%$ of the mothers and
$25,97 \%$ of the fathers come from non-European Union countries of the Europe area. One of the non-European Union countries we are interested in is Albania. We will run a separate regression for the first hypothesis in which we control for children whose parents come from non-European Union countries in the Europe area. While this also includes observations from countries where few people speak Italian, we do capture a lot of Italian speaking subjects. While we cannot remove countries such as Malta or Romania from the dataset, we will still get a good indication whether removing Italian speaking immigrants has an effect on our estimates. Since these immigrants already know Italian, speaking Italian at home is not an effective measure of integration anymore.

Table 4 shows the distribution of the amount of students with whom the immigrant gets along in the classroom. For the ease of the research this variable will from now on be called "amount of friends". The five categories are defined as follows: no friends, a few friends, quite some friends, a lot of friends, all are friends. Keep in mind that here we analyze how well a student fits in the classroom and not the amount of friends he ended up making. The latter could bias our estimates by the fact that some individuals are better at making friends and more sociable than others as discussed previously. Using this measure allows us to see how well an individual managed to integrate in the classroom and to become accepted. Only around $15 \%$ of the students report getting along with only few classmates or none. Close to $25 \%$ of the immigrants report that they get along with all of their classmates which is surprisingly high, and incites us to believe that in Italy a big share of the $5^{\text {th }}$ grade immigrants is well integrated into the national culture.

Our control variables take value 1 when the individual falls into the category specified by the name of the control variable. This holds for the variables "female", "European_nonEU_immigrant", "first_generation", "nursery" and "kindergarten". For example variable "female" as a dummy variable is specified as $0=$ male and $1=$ female.

When analyzing the years 2011, 2012, 2015 and 2016 no relevant difference in frequencies and descriptive statistics was found.

## 6. Methodology

Fixed effects are an important aspect of our methodology as we will discuss more in detail below. Regarding our first hypothesis stated in the introduction, our dependent variable "amount_of_friends" is a categorical variable which requires ordered logit models. The problem
we face is that there is yet a consensus to be found in literature on how to integrate fixed effects in ordered logit models (Baetschmann, 2011) ${ }^{1}$. For simplicity and security reasons we will transform our categorical variable into a dummy variable referred to as "lot_of_friends". This new variable will take value zero for categories one and two and take value one for categories three, four and five of the variable "amount_of_friends". Thanks to this new variable we can use logistic fixed effects regressions. To answer our first hypothesis we want to measure the effect of parental education on the amount of friendships of the child. To test this we use a basic model expressed as follows:
amount_of_friends $_{\mathrm{i}, \mathrm{c}, \mathrm{t}}=\alpha+\beta \operatorname{sex}_{\mathrm{i}}+\beta$ year $_{\mathrm{t}}+\beta$ i.education_father $\mathrm{r}_{\mathrm{i}}+\mathrm{c}+\mathrm{e}$

Where i refers to each individual student, t to the year and c to the classroom. We start with a logistic regression but end up using panel data OLS regressions for the majority of our research. The reasons for this will be discussed in the results section. While this is the basic model, for comparison we will first use a model without fixed effects. We will also run regressions where we distinguish between first and second generation migrants. The dataset will be split the in two, one dataset will have only first generation immigrants which will allow to control as well for the age of the child at immigration. Table 5 shows the distribution of the first and second generation immigrants. The age at which the child immigrated to Italy potentially has an impact on its relationships with natives as supported by previous research. First, children being in the country for several years already had more time to make friends. In the classrooms of small schools, composition barely changes throughout the years and following Hallinan (1989), the more opportunities students have to interact, the more friendships are created. The second subset of data will only include second generation immigrants and will allow us to control for whether the child went to the nursery or kindergarten. Table 6 and 7 show the distribution of these variables. For the same reasons as stated before, children who had more contact with Italian children previous to entering primary school might have created more friendships beforehand. This is supported by the literature that shows that going to the nursery encourages playing with other children (Allen et al, 1964) and that participating in kindergarten allows higher skill development (Ladd et al, 1999).

[^0]The ease for immigrants to make friends might be impacted by the public opinion which changes by year. If events such as terrorist attacks occur, people's opinion towards immigration might alter in that particular year. While it is unrealistic believing 5th grade primary school children are affected by such events directly, the attitude of the parents plays a major role in the socialization process of the children.

Including classroom fixed effects allows to account for neighborhood differences. Schools in expensive suburbs have a lower migrant concentration and following Hallinan (1982), the racial majority in a classroom is more likely to be popular with the other races than the minority. This implies that in certain geographical areas it is easier for immigrants to make friends. In addition to this neighborhoods concentrating more educated citizens are advantageous for immigrants because of the more universalistic view of life educated individuals have (Kalmijn, 1998). The classroom fixed effects also differentiate between public and private school, account for differences in funding per school and for other differences that can affect student interactions. Classroom fixed effects are our most effective measure for teacher effects. Teachers may or may not create a classroom environment in which interracial friendship is promoted. Allport (1954) suggests that teachers may make abstraction from status and Cohen (1975) found that they sometimes ignore status hierarchy, both practices that help the integration of immigrants. Teacher intervention has a significant effect on the interracial relationships (Hallinan, 1982). Classrooms also have different class hours per week. We believe that spending more hours with its classmates largely facilitates friendships. Fixed effects will allow us to control for this.

Structural differences in between regions might create different attitudes of natives towards immigrants, which has a direct impact on the interactions between natives and foreigners. Regions with a lot of migrants might be more open-minded and less hostile, while more separatist regions such as Sicily or South Tyrol are by definition more austere towards foreigners. This is also captured by classroom fixed effects.

As can be seen in table 2, previous research used similar controls such as sex, length of stay (which is similar to our variable "age at immigration") and first or second generation immigrant. Table 2 displays the hypothesized sign of the effects of our control variables and the previous findings and significance of these control variables. We decide to use only "education_father" and not "education_mother" because of multicollinearity concerns. We believe this is important because a broad literature showed that individuals tend to marry someone belonging to the same
social class (see Buss, 1985). Section 8 will show robustness checks where the regressions are run with the education of the mother as independent variable instead.

The second regression will measure the effect of parental education on the language spoken at home, controlling for the same variables. Our variable "language_at_home" is a dummy variable taking value zero if the language spoken is not Italian and value one if the child speaks Italian at home. Since this variable is a dummy variable we start with a simple logit model:

$$
\text { speaking_italian_at_home }_{i, c, t}=\alpha+\beta \operatorname{sex}_{i}+\beta \text { i.year }{ }_{t}+\beta \text { i.education_father }{ }_{i}+\mathrm{c}+\mathrm{e}
$$

We will follow the same steps as with the first hypothesis regarding the different regressions used, but here we add one more step where we remove migrants coming from non-EU countries to account for the fact that the majority of Albanian migrants already spoke Italian at their entry. 40 717 observations remain after this manipulation. Again, we end up using panel data OLS regressions for the majority of the models as will be discussed further below.

We use the same controls as for our first regressions since all of these variables affect participation in social life of immigrants as justified above. In this case, age at immigration might have a much larger effect since individuals being in the country for a longer time had more time to master the language and adopt it in their daily lives. Moreover, several studies suggest that language learning is more difficult after early childhood (Lenneberg, 1967; Newport, 1990) which implies that children immigrating at an early age have a significant advantage on mastering and using the language. Though, we look at the language spoken at home and not solely at how well a child speaks Italian. The language spoken at home is not only dependent on the child's language skills but also those of his parents. Parents decide which language is spoken in the household more than the children and even if children may learn faster following Lenneberg or Newport, in the majority of the cases parents master a specific language better than their 11 years old child does. This precision was necessary to show that some of our control variables only affect the language at home through the impact on a child's ability to speak Italian. Other controls such as being a first generation immigrant or being a European Union immigrant might still have a very large effect since those variables do not only affect the child but the households as a whole.

Multicollinearity is not at stake here, since education is a predetermined characteristic. We make the assumption that immigrated parents do not resume their studies in the host country. This assumption is made realistic by the fact that immigrated parents first need to accommodate and organize their income, especially since migrating has a large cost, and therefore will at least in the short-term not restart studying but work. Furthermore, raising children is costly and only few families can afford educating their children while quitting their job.

## 7. Results

We first analyze the effect of parental education on the amount of friends of a child. We hypothesized that higher parental education engenders a higher probability for a child to have a lot of friends. Table 9 shows the results obtained for our first hypothesis.

Table 9: Effects of parental education on the probability of a child having a lot of friends


Models (1) and (2) show the coefficients for a simple OLS regression and a logit regression without fixed effects. When transforming these coefficients into probabilities, by using the odds ratio for the logit regression, we obtain similar results. Table 10 compares the effects obtained with both methods.

Table 10: Increase in probability of having a lot friends compared to the category above (expressed in percentage points)

|  | logit | OLS |
| :---: | :---: | :---: |
| constant | 79.51 | 79.82 |
| category |  |  |
| 2 | 3.48 | 3.25 |
| 3 | 1.07 | 1.01 |
| 4 | 3.22 | 3.01 |
| 5 | -0.55 | -0.53 |
| 6 | 0.29 | 0.28 |

There is no substantial statistic difference between the results obtained with both regressions. Therefore we choose to run the following regressions as panel data regressions, and not logit. The regression number (3) displays the coefficients of the corresponding panel OLS regression where classroom fixed effects were accounted for. When observing regressions (1) and (3) we observe various disparities. Fixed effects thus have a large impact on the estimation of the effects of our independent variables. Therefore we will include them in all of our following analysis. Only the sex dummy is insignificant as was expected due to previous findings (ten Teije et al, 2013). The next step is to add an extra control variable to our model, the "firstgeneration" dummy that describes whether the child is a first generation or second generation immigrant. We expected second generation immigrants having more friends than first generation immigrants because of their longer stay in Italy providing them with more time to integrate and make friends. Also, these children being born in Italy do not have to face a culture shock. This hypothesis seems confirmed in regression (4), which we will also consider as our main model controlling for the most variables and including all the observations. The sign of the coefficient being negative and highly significant implies that second generation immigrants have on average $2.99 \%$ points more chances of having a lot of friends comparing to first generation immigrants. All the categories of
the education of the father are positive and increasing up to education category 4 . Children whose parents have a category 2 education are on average $3.65 \%$ points more likely to have a lot of friends compared to children whose parents are in category 2 . Similarly, category 4 gives $3.71 \%$ points extra ( $=7.36-3.65$ ) chances of having a lot of friends compared to category 2 children and $0.62 \%$ points extra probability compared to children of categories 5 or 6 . This latter finding will be discussed more in details further below.

Model (5) is run only on second generation immigrants and additionally controls for "nursery" and "kindergarten", two dummies that inform whether the child went to the nursery or kindergarten respectively. The effects of education remain similar to our previous regression with regards to the sign of the effect, while we observe small changes in the magnitude of the effect. The interesting feature of this model are the coefficients of our two additional control variables. Children who went to the nursery have $0.82 \%$ points extra and children who went to kindergarten have $4.43 \%$ points extra probability to have a lot of friends in their classroom. While the nursery estimates are insignificant, the effect of kindergarten is significant and important as we expected. The higher effect of going to kindergarten compared to going to nursery can be explained by the fact that kindergarten is both closer in time to primary school and often linked to the particular primary school the child will enter after its kindergarten years. Children that went together to kindergarten are thus often in the same class or at least school a couple of years later. Thus if children have a lot of friends at kindergarten they have a higher probability of having a lot of friends in primary school as well.

Finally, regression (6) looks only at first generation immigrants and controls for an additional variable, the age of the child at immigration. Here we see a higher disparity of the probabilities of having a lot of friends between the different education categories. This illustrates that in the process of integration, the level of education plays a bigger role for first generation immigrants than for second generation immigrants. We can interpret this in the way that at arrival, higher educated immigrants are at first more curious and more willing to participate in social life, because of more openness to the world are more greed to have control over their lives. Once this initial period of arrival has passed, what has a higher impact on the integration of the children in the long-run might become the neighborhood in which the household lives or the attitude of the teachers towards foreign students rather than the education of the parents. The difference is as big
as $7.47 \%$ between children who immigrated at age 1 and lower and children who immigrated at age 10 .

We previously said that we cannot research the second part of the "immigrant paradox" because we do not have information about reported perceived discrimination. The results of regressions (4), (5) and (6) showcase a similar pattern as the "immigrant paradox" states. We assume that the socializing behavior is a direct consequence of the perceived discrimination of individuals. Both in the way that individuals feeling discriminated are less willing to put effort in making friendships, and in the way that individuals that have a hard time making friends may feel discriminated because of this. If this holds, we can still make same comments regarding the paradox. Up to education category 4 (5 for regression (6)), children have a higher probability of having a lot of friends if they are in a higher category. But after this, this probability decreases again and we see from regression (5) that children whose parents are in education category 6 have a $0.61 \%$ points lower probability of having a lot of friends compared to children whose parents belong to category 5 . This would show evidence for the fact that higher education increases the integration of the immigrants but also increases their hostility towards natives (or the perceived discrimination, or the natives' hostility towards immigrants). What exactly causes this decreasing integration cannot be observed nor derived from this dataset, but it seems evident that category 4 or 5 are the tipping point at which the negative effect of education on integration becomes larger than the positive effect education has on integration and would therefore support the "immigrant paradox" encountered in previous literature.

While these results reflect the effects of parental education on the amount of friends of children, in the next section we look at the effects on the language spoken at home.

Secondly, we analyze the effects of parental education on the language spoken at home. Table 11 displays the results regarding our second hypothesis which states that higher educated immigrants have a higher probability of speaking Italian at home.

Table 11: effects of parental education on the probability of children speaking Italian at home

| Dependant variable: Italian_at_home |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Female | 0.0037* | 0.0044* | 0.0045* | 0.0038 | 0.0043 | 0.0007 |
|  | (0.0022) | (0.0024) | (0.0024) | (0.0024) | (0.0034) | (0.007) |
| Education father |  |  |  |  |  |  |
| 2. | 0.0649*** | 0.0518*** | 0.0462*** | 0.0426*** | 0.0484*** | 0.028** |
|  | (0.0048) | (0.0052) | (0.0052) | (0.0052) | (0.0074) | (0.0139) |
| 3. | 0.0675*** | 0.0606*** | 0.0559*** | 0.0596*** | 0.0669*** | 0.0524*** |
|  | (0.0054) | (0.006) | (0.006) | (0.0059) | (0.0084) | (0.0156) |
| 4. | 0.1137*** | 0.0924*** | 0.0846*** | 0.0793*** | 0.0892*** | 0.0454*** |
|  | (0.0048) | (0.0053) | (0.0053) | (0.0053) | (0.0075) | (0.0144) |
| 5. | 0.0813*** | 0.0676*** | 0.062*** | 0.0601*** | 0.0713*** | 0.0331 |
|  | (0.0072) | (0.0078) | (0.0078) | (0.0077) | (0.0106) | (0.0207) |
| 6. | 0.1327*** | 0.0932*** | 0.0885*** | 0.0858*** | 0.0976*** | 0.0205 |
|  | (0.0058) | (0.0022) | (0.0064) | (0.0064) | (0.0091) | (0.0178) |
| European but non-EU | No | No | 0.0537*** | 0.0456*** | 0.0488*** | 0.0415*** |
|  |  |  | (0.003) | (0.0029) | (0.004) | (0.0086) |
| First generation | No | No | No | $\begin{aligned} & -0.1576 * * * \\ & (0.0028) \end{aligned}$ | No | No |
| Nursery ${ }^{2}$ | No | No | No | No | 0.0266*** | No |
|  |  |  |  |  | (0.0038) |  |
| Kindergarten | No | No | No | No | $\begin{array}{r} 0.0641 * * * \\ (0.0079) \end{array}$ | No |
| Number of observations | 173,834 | 173,834 | 173,834 | 173,834 | 100,355 | 28,097 |
| Adjusted/Pseudo/Overall | 0.0734 | 0.1159 | 0.1178 | 0.136 | 0.1052 | 0.1899 |
| Age at immigration | No | No | No | No | No | Yes |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Classroom Fixed Effects | No | Yes | Yes | Yes | Yes | Yes |
| Including first generation | Yes | Yes | Yes | Yes | No | Yes |
| Including second generation | Yes | Yes | Yes | Yes | Yes | No |
| All panel data regressions are including the intercept. <br> ***, **, * denote statistically significant effects at a $1 \%, 5 \%$ and $10 \%$ level respectively. |  |  |  |  |  |  |

The regressions follow the same logic as for the first hypothesis but this time the tests whether logit or xtreg regressions give different results have been excluded. Model (3) shows the effect of adding a control for non-EU European area countries. We see that individuals coming from non-EU European area countries have on average $5.37 \%$ points higher probability of speaking Italian at home compared to individuals from other countries. This result is highly significant and the effect substantially high. Including this control in all of our following models is important for future representative results. Model (4) includes an additional control for first generation

[^1]immigrants and will therefore be considered as our most accurate model. A first generation immigrant has $15.76 \%$ points less chance of speaking Italian at home compared to second generation immigrants. This coefficient is much higher than for our first hypothesis where it was close to $3 \%$ points only. Being born in a country has thus a very big effect on the probability of speaking the official language of that country at home. This is not a surprising result, since children being born in Italy have Italian as their mother tongue regardless of their origins. Italian is the language they hear the most and therefore the language they master the most. Opposed to what we discussed, the language abilities of a child seem to have a big impact on the language spoken at home. Regarding our main variable of interest, education, we see a similar effect as for our first hypothesis. Higher educated individuals speak more often Italian at home than lower educated individuals, even if this effect doesn't follow a straight line. The probability of speaking Italian increases from categories 1 until 4 , decreases from category 4 to 5 and increases again from category 5 to 6 , ending at a probability being higher than the probability at category 4 . Individuals whose father belongs to category 4 have a $7.93 \%$ points higher probability of speaking Italian at home compared to individuals of category 1 . At the same time, individuals of category 6 have a $2.57 \%$ points extra higher probability of speaking Italian at home compared to category 5 individuals but this percentage only amounts up to $0.65 \%$ percentage points comparing category 4 to 6 . This implies there is a negative effect of belonging to category 5 compared to category 4 . Those individuals have a $1.82 \%$ point's lower probability of speaking the national language at home.

The output of regressions (5) and (6) is obtained after splitting the dataset in two. Model (5) only includes second generation immigrants and controls for nursery and kindergarten attendance. These coefficients are significant at $1 \%$ and imply that children who went to both nursery and kindergarten have on average $9.07 \%$ points extra probability of speaking Italian at home compared to children that did not go to any of these two institutions. All coefficients of paternal education are higher, which was expected because of the negative sign of the control "first_generation". Removing the first generation immigrants increased the average effect of education on the probability of speaking Italian at home. Finally model (6) allows us to control for the age at immigration. This time most education estimates are insignificant and follows a different pattern. The effect of being a category 6 individual compared to category 1 is this time much lower and only equals $2.05 \%$ points.

These results show a different pattern than the pattern of the "immigrant paradox" since category 6 individuals have a higher probability of speaking Italian at home compared to all other individuals. The probability still decreases at category 5 but increases back at category 6 . There is no evidence of a tipping point. This suggests that the negative effect of category 5 is due do the small amount of observations in that category, $3.94 \%$. Regardless of the pattern observed, I recall that this paradox looks at the perceived discrimination of the immigrants. While the amount of friends of immigrants can be related to the discrimination immigrants discern, the language they speak at home is independent of this. Still, we see that up to category 4, education has a large effect on the integration. After that category, this effect becomes smaller. This shows evidence that education seems to matter for lower category individuals but that after a certain education level, being higher educated does not seem to have any effect. One possible explanation is that it is not education but employment of the parents that affects the integration of the children. Since education is highly correlated with employment, including an employment control variable to our model is not possible. Lower educated immigrants are less likely to find a job. The higher immigrants are educated the higher their chances of being employed. After a certain education level, all immigrants find an employment and thus being more educated does not give more probability of being employed. Employment could impact social integration since active immigrants might have more interactions with natives than the unemployed share of immigrants.

## 8. Robustness Check

In section 6 we assumed that the education of the father was highly correlated with the education of the mother. The correlation between these two variables equals 0.608 . We therefore run model (3) from our first hypothesis with maternal education instead of paternal education. A comparison of the results can be found in table 12.

Table 12: coefficient difference of paternal (1) education and maternal education (2)

| Dependent variable: lot_of_friends |  |  |
| :---: | :---: | :---: |
|  | (1) | (2) |
| Female | $\begin{aligned} & -0.00008 \\ & (0.0034) \end{aligned}$ | $\begin{aligned} & -0.00012 \\ & (0.0034) \end{aligned}$ |
| Year | $\begin{array}{r} 0.0295 * * * \\ (0.0037) \end{array}$ | $\begin{gathered} 0.0296 * * * \\ (0.0037) \end{gathered}$ |
| Education parent $2 .$ | $\begin{aligned} & 0.0373 * * * \\ & (0.0074) \end{aligned}$ | $\begin{aligned} & 0.033 * * * \\ & (0.0069) \end{aligned}$ |
| 3. | $\begin{array}{r} 0.0467 * * * \\ (0.0085) \end{array}$ | $\begin{gathered} 0.039 * * * \\ (0.0085) \end{gathered}$ |
| 4. | $\begin{aligned} & 0.0748 * * * \\ & (0.0075) \end{aligned}$ | $\begin{array}{r} 0.0617 * * * \\ (0.007) \end{array}$ |
| 5. | $\begin{array}{r} 0.0686 * * * \\ (0.0110) \end{array}$ | $\begin{aligned} & 0.0515 * * * \\ & (0.0105) \end{aligned}$ |
| 6. | $\begin{array}{r} 0.0685 * * * \\ (0.0091) \\ \hline \end{array}$ | $\begin{array}{r} 0.0643 * * * \\ (0.0083) \\ \hline \end{array}$ |
| Number of observations | 56,513 | 56,513 |
| Adjusted R-squared | 0.0315 | 0.0305 |
| Fixed Effects | Yes | Yes |
| All panel data regressions are including the intercept. ***, **, * denote statistically significant effects at a $1 \%$, $5 \%$ and $10 \%$ level respectively. |  |  |

The difference in effects are quite substantial and equals $1.71 \%$ points for category 5 . We observe that the coefficients of paternal education (1) are higher than the coefficients of maternal education (2). This suggests that fathers have more impact on the integration of their children than mothers. Besides this disparity, the differences of coefficients between categories are similar and follow a similar pattern, except for category 6 . Furthermore the R-squared of the model with paternal education is slightly higher. Our choice to include only paternal education in the models seems justified, and would not have led to big differences with regards to the interpretation of our results.

Secondly, during our data manipulations the year 2012 dataset presented reporting mistakes. In one dataset the student numbers all contained an identical extra code number. When removing this common number the merging was made possible, but because of this change we had to make ourselves we cannot be certain that the student numbers matched well. We thus compared two models in which we first included and later removed the year 2012 to check for differences. The first model is the same model as model (3) used for our second hypothesis in table 11. If the merge of the year 2012 matched students to their wrong characteristics results obtained with year 2012 would be significantly different. Table 13 shows the comparison of both results:

Table 13: coefficient differences including (1) and excluding (2) year 2012

| Dependent variable: Italian_at_home |  |  |
| :---: | :---: | :---: |
|  | (1) | (2) |
| Female | $\begin{aligned} & \hline 0.0045^{*} \\ & (0.0024) \end{aligned}$ | $\begin{aligned} & 0.0068^{* *} \\ & (0.0027) \end{aligned}$ |
| Education father |  |  |
| 2. | $\begin{aligned} & 0.0462 * * * \\ & (0.0052) \end{aligned}$ | $\begin{aligned} & 0.0563 * * * \\ & (0.0058) \end{aligned}$ |
| 3. | $\begin{gathered} 0.0559 * * * \\ (0.006) \end{gathered}$ | $\begin{aligned} & 0.0671 * * * \\ & (0.0066) \end{aligned}$ |
| 4. | $\begin{array}{r} 0.0846 * * * \\ (0.0053) \end{array}$ | $\begin{gathered} 0.1006 * * * \\ (0.0059) \end{gathered}$ |
| 5. | $\begin{aligned} & 0.062^{* * *} \\ & (0.0078) \end{aligned}$ | $\begin{gathered} 0.0756 * * * \\ (0.0087) \end{gathered}$ |
| 6. | $\begin{aligned} & 0.0885 * * * \\ & (0.0064) \end{aligned}$ | $\begin{aligned} & 0.1051 * * * \\ & (0.0071) \end{aligned}$ |
| European but non-EU | $\begin{array}{r} 0.0537 * * * \\ (0.003) \\ \hline \end{array}$ | $\begin{array}{r} 0.0606 * * * \\ (0.0033) \\ \hline \end{array}$ |
| Number of observations | 173,834 | 148,316 |
| Adjusted R-squared | 0.1178 | 0.0643 |
| Year Fixed Effects | Yes | Yes |
| Classroom Fixed Effects | Yes | Yes |
| All panel data regressions are including the intercept. ***, **, * denote statistically significant effects at a $1 \%$, $5 \%$ and $10 \%$ level respectively. |  |  |

We observe a significant difference in the magnitude of the effect between both models. Model (1) displays the results obtained when including the observations of year 2012. The second model shows results without that specific year for which we have uncertainty about the reliability of the data. While the magnitude is different, the sign is identical and the pattern of the effects, i.e. the difference in effects between categories, is very similar. The adjusted R-squared is higher for the first model and all standard errors are inferior. Omitting year 2012 in model (2) did not lead to a difference in results but reduced the reliability and accuracy of the model. Therefore we confirm there is no bias due to the uncertain merge and that the matching was accurate.

## 9. Suggestions for future research and policy implications

Several limitations of this research have already been discussed previously. Data and econometric limitations limit a more detailed and analysis. The suggestions for future research go
in two ways. First I would like to address the data limitations, and secondly the omissions in the analysis that could be investigated in future research, or how the results could be extended.

The survey attached to the INVALSI test provides very detailed information about each Italian student and his household. Though, the questions vary per year and only some remain for the full period. The answers to some very interesting questions are only available for a limited time period and does not allow for consistent analysis over time. Furthermore, each child is given a student number for a specific year. The next year the same student numbers are used and given to other students (or are incorrectly reported) which does not allow to follow a student over the years and analyze time variant characteristics and investigate their evolution. A big share of the observations have missing values for a number of variables which are due to reporting mistakes. The INVALSI test is mainly a tool for education comparisons and statistics but its potential for other purposes for economic and social analysis is large.

This research used large and precise data publicly available about this topic, but other data could allow the use of more precise measures of integration and more efficient methodology. The use of ordered logistic models with fixed effects is certainly a more adequate approach but we will have to wait for a consensus in the field of econometrics. Information about perceived discrimination can be obtained through a similar survey ten Teije et al. (2013) used and allows to investigate the existence of an immigrant paradox. The failure to control for individuals who spoke Italian before entering the country might be an obstacle to the relevance of our results, and data with this additional information will tell us to what extent. In future research on the effects in scholastic settings, it is important to take school and classroom characteristics into account. This study shows that fixed effects have an impact on inter classroom relationships of the students. Overall the possibilities are numerous but impossible to exploit in one paper, leaving room for other curious and creative minds.

Various instances pressure the government to increase education spending because of its well-known benefits on poverty, innovation, employment or growth. This paper through its results provides evidence in favor of a broader perspective: society's well-being. Higher educated individuals would integrate better and increase society's cohesion.

## 10. Conclusion

Integration of immigrants is important for employment and career opportunities. The literature showed that highly educated immigrants are better in the labor market. But the "immigrant paradox" also shows that education does not necessarily lead to better social integration. Hence, this paper aims to research the effects of parental education on the integration of immigrant $5^{\text {th }}$ grade children in Italy. The INVALSI yearly tests, from the period 2011-2016, offers to better study these type of effects. In line with our first hypothesis we found that better education of the parents increased the probability that a child has a lot of friends in his classroom. We also found that education increases the probability for a household to speak Italian at home. We therefore can accept both of our hypotheses to the extent that individuals with lower parental education score in both cases a lower average probability than individuals with higher parental education. However, we cannot confirm that these effects remain when comparing highly educated individuals. After a certain level of education, increasing education affects the integration of the child to a smaller extent. We did not find what this level of education exactly was.

Overall, the answer to the research question is that parental education affects the integration of immigrant children. Higher educated parents have more integrated children compared to lower educated parents. Taking the average of the effects obtained for both hypothesis we find that individuals belonging to the highest education category have on average an approximated probability of being integrated $7 \%$ points higher compared to the lowest education category, all else being equal.

Besides our main findings, interesting effects of the control variables have been found ${ }^{3}$. We obtain that second generation immigrants are more integrated than first generation immigrants. As previous literature found, the "female" estimates are not significant. Going to the nursery and kindergarten increases the probability that the child will be well integrated in the local culture and social life. Non-EU European area country immigrants are better integrated than other immigrants because of their historically proximity to the Italian culture and language. The absence of a repeated survey over several years does not allow us to analyze the integration and what affects the

[^2]integration of immigrants over time and to check whether this type of short and long-term effects exist. It is interesting to note that children immigrating at a later age have a lower probability to have a lot of friends compared to children who immigrated at an early age, even if the age at immigration shows insignificant coefficients. While this was not the aim of the paper, we found possible evidence for the "immigrant paradox" if the assumption that perceiving discriminations retains one from making friends holds. This paradox remains an interesting subject to discuss.

While some data limitations restricted our analysis we were able to derive significant and valuable results. An extensive share of immigrants are doing their best to integrate into the host country. Whether the fact that higher educated immigrants succeed better is due to the acquisition of better skills and tools that allowed them to become accepted or due to the fact that they are seen as having a better influence and bringing more utility to the natives compared to lower categories is not known. While it is likely that both effects are present, if the latter, which we will call relative education, is the major effect driving higher educated to be more integrated, increasing education for all individuals would not solve many problems of integration. However, the idea that better educated individuals have larger openness to the world proposed by Kalmijn (1998) and are thus more willing to make effort towards the native population is likely to be true. In that case, a better education for everyone is a possible solution for better integration and increased social cohesion. Regardless of the question of what drives higher educated individuals to be more integrated, this paper supplies additional evidence for the benefits of education and hopefully helps to build relevant welfare increasing policies, both for native populations as for immigrants.

## References

Ainsworth, H. L., \& Eaton, S. E. (2010). Formal, Non-Formal and Informal learning: The Case of Litteracy, Essential Skills and Language Learning in Canada. ERIC.

Akcigit, U., Grigsby, J., \& Nicholas, T. (2017). Immigration and the rise of American ingenuity. American Economic Review (107), 327-331.

Allen, K. E., Hart, B., Buell, J. S., Harris, F. R., \& Wolfe, M. A. (1964). Effects of social reinforcement on isolate behavior of a nursery school child. Child Development(35), 511-518.

Allport, G. W. (1954). The Nature of Prejudice. Cambridge, MA: Addison-Wesley.
Baetschmann, G., Staub, K. E., \& Winkelmann, R. (2011). Consistent Estimation of the Fixed Effects Ordered Logit Models. Bonn: The Institute for the Study of Labor (IZA).

Baumgartner, M. P. (1998). Moral life on the cultural frontier: Evidence from the experience of immigrants in modern America. Sociological Focus(31), 155-179.

Bertola, G., Checchi, D., \& Oppedisano, V. (2007). Private School Quality in Italy. Giornale degli Economisti e Annali di Economia, 66(3), 375-400.

Black, D. (1993). The Social Structure of Right and Wrong. San Diego: AcademicPress.
Bodvarsson, O. B., \& Van den Berg, H. (2009). The Economics of Immigration: Theory and Policy. Heidelberg: Springe.

Booth, A. L., \& Van Ours, J. C. (2008). Job satisfaction and family happiness: the part-time work puzzle. Economic Journal, 114, 77-99.

Borjas, G. (1995). The Economic Benefits From Immigration. Journal of Economic Perspectives(2), 3-22.
Borjas, G. (1996, November). The New Economics of Immigration. The Atlantic Monthly, 72-80.
Borjas, G. (2003). The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. Quarterly Journal of Economics(118), 1335-1374.

Card, D. (2009). Immigration and Inequality. American Economic Review(2), 1-21.
Chiswick, B. R. (1991). Speaking, Reading, and Earnings among Low-Skilled Immigrants. Journal of Labor Economics(2), 149-170.

Chiswick, B. R. (2005). The Economics of Immigration: Selected Papers of Barry R. Chiswick. Cheltenham, UK: Edward Elgar Publishing.

Chiswick, B. R., \& Miller, P. W. (1992). Language in the Immigrant Labor Market: The Immigrant Experience in Canada and The United-States. In B. R. Chiswick, Immigration, Language and Ethnicity: Canada and the United States (pp. 229-296). Washington, DC: American Enterprise Institute.

Cohen, E. (1975). The Effect of Desegragation on Race Relations. Law and Contemporary Problems, 39, 271-299.

Cohen, F. S. (1940). Exclusionary Immigration Laws: Their Social and Economic Consequences. Contemporary Jewish Record, 3(2).

Cooney, M. (2009). Ethnic conflict without ethnic groups: A study in pure sociology. British Journal of Sociology(60), 473-492.

Das, M., \& Soest, A. (1999). A panel data model for subjective information on household income growth. Journal of Economic Behavior and Organization, 40, 409-426.

Dustmann, C. (1994). Speaking Fluency, Writing Fluency and Earnings of Migrants. Journal of Population Economics(2), 133-156.

Dustmann, C., Fabbri, F., \& Preston, I. (2005). The Impact of Immigration on the British Labour Market. Economic Journal(115), 324-341.

Ferrer-i-Carbonell, A., \& Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? Economic Journal, 114, 641-659.

Fong, E., \& Isajiw, W. W. (2000). Determinants of Friendship Choices in Multiethnic Society. Sociological Forum, 15, 249-269.

Gijsberts, M., \& Vervoort, M. (2007). Wederzijds Beeldvorming [Mutual Stereotyping]. In J. Dagevos, \& M. Gijsberts, Jaarrapport Integratie (pp. 282-308). The Hague, The Netherlands: SCP.

Hallinan, M. T. (1982). Classroom racial composition and children's friendships. Social Forces(61), 5672.

ISTAT. (2017). Population and Households data. Italy: National Institute of Statistics.
Kalmijn, M. (1998). Intermarriage and Homogamy: Causes, Patterns, Trends. Annual Review of Sociology, 24, 395-421.

Ladd, G. W., Birch, S. H., \& Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? Child Development(70), 1373-1400.

Lenneberg, E. (1967). Biological foundations of language. New-York: Wiley.
Lin, N. (1999). Social Networks and Status Attainment. Annual Review of Sociology(25), 467-487.
Mahuteau, S., \& Junankar, P. N. (2008). Do Migrants Get Good Jobs in Australia? The Role of Ethnic Networks in Job Search. Economic Record, 84(1), 115-130.

Martinovic, B., Van Tubergen, F., \& Maas, I. (2009). Dynamics of interethnic contact: A panel study of immigrants in The Netherlands. European Sociological Review(25), 303-318.

Newport, E. L. (1990). Maturational constraints on language learning. Cognitive Science(14), 11-28.
OECD. (2018). International Migration Database. Paris: OECD.
Ottaviano, G. I., Peri, G., \& Wright, G. C. (2013). Immigration, offshoring, and American jobs. American Economic Review(5), 1925-1959.

Romanian National Institute of Statistics. (2018). Population and Housing Report. Romania: INS.
Sam, D., Vedder, P., Liebkind, K., Neto, F., \& Virta, E. (2008). Immigration, acculturation and the paradox of adaptation in Europe. European Journal of Developmental Psychology(5), 138-158.
ten Teije, I., Coenders, M., \& Verkuyten, M. (2013). The paradox of integration: Immigrants and their attitude toward the native population. Social Psychology(44).

Tolsma, J., Lubbers, M., \& Gijsberts, M. (2012). Education and cultural integration among ethnic minorities and natives in the Netherlands: A test of the integration paradox. Journal of Ethnic and Migration Studies(38), 793-813.

Van Doorn, M., Scheepers, P., \& Dagevos, J. (2013). Explaining the integration paradox among small immigrant groups in the Netherlands. Journal of International Migration and Integration, 14, 381-400.

Weijters, G., \& Scheepers, P. (2003). Verschillen in Sociale Integratie tussen Etnische Groepen: Beschrijving en Verklaring [Differences in Social Integration between Ethnic Groups: Description and Explanation]. Mens en Maatschappij, 78, 144-157.

## Appendix

Table 1: effect of manipulations on the amount of observations for hypothesis 1

| Hypothesis 1 <br> Years 2013 to 2015 | Initial observations: 995024 |  |
| :--- | :---: | :---: |
|  |  |  |
| Manipulation | Lost observations | Remaining observations |
| dropping non merged | 138163 | 856861 |
| dropping natives | 765846 | 91015 |
| dropping missing education | 28596 | 62419 |
| dropping missing information about child relationships | 1730 | 60689 |
| dropping missing citizenship | 4176 | 56513 |

Table 2: hypothesized signs of the effects of the independent and control variables

| Variable | Hypothesis 1 | Hypothesis 2 | Litterature | Significance |
| :---: | :---: | :---: | :---: | :---: |
| Dependent |  |  |  |  |
| Amount of friendships |  |  |  |  |
| Language spoken at home |  |  |  |  |
| Independent |  |  |  |  |
| Sex | 0 | 0 | - (ten Teije et al, 2013) | not significant |
| Year | 0 | 0 |  |  |
| Education mother | + | + | + (Martinovic et al, 2009) | significant at 0.1 |
|  | + | + | + (ten Teije et al, 2013) | significant at 0.001 |
| Education father | + | + | + (Martinovic et al, 2009) | significant at 0.1 |
|  | + | + | + (ten Teije et al, 2013) | significant at 0.001 |
| European non-EU immigrant |  | + |  |  |
| Age at immigration | + | + | + (Martinovic et al, 2009) | not significant |
|  | + | + | + (ten Teije et al, 2013) | significant at 0.001 |
| Is a first generation immigrant | - | - | - (ten Teije et al, 2013) | significant at 0.001 |
| Went to nursery | + | + | + (Allen et al, 1964) | not applicable |
| Went to kindergarten | + | + | + (Ladd et al, 1999) | significant at 0.001 |

Table 3: distribution of maternal education (1) and paternal education (2), displaying their highest degree obtained

|  | $(1)$ <br> Education <br> mother | $(2)$ <br> Education <br> father |
| :--- | :---: | :---: |
|  | Freq | Freq |
|  | (Percent) | (Percent) |
|  |  |  |
| Elementary school diploma | 4,255 | 3,674 |
|  | $(7.529)$ | $(6.501)$ |
| High school diploma | 19,803 | 19,374 |
|  | $(35.04)$ | $(34.28)$ |
| Professional diploma | 5,252 | 7,125 |
|  | $(9.293)$ | $(12.61)$ |
| Bachelor | 18,394 | 18,825 |
|  | $(32.55)$ | $(33.31)$ |
| Higher than bachelor | 2,414 | 2,227 |
|  | $(4.272)$ | $(3.941)$ |
| Master, Phd degree or higher | 6,395 | 5,288 |
|  | $(11.32)$ | $(9.357)$ |
| Total |  |  |
|  | 56513 | 56513 |

Table 4: distribution of the variable "friends"

| Friends | (1) <br> Freq <br> (Percent) |
| :--- | :---: |
| No friends | 908 |
|  | $(1.607)$ |
| A few friends | 6,983 |
| Quite some friends | $(12.36)$ |
|  | 12,400 |
| A lot of friends | 22,9496 |
|  | $(39.63)$ |
| All are friends | 13,826 |
|  | $(24.47)$ |
|  |  |
| Total | 56513 |

Table 5: distribution of the variable "first_generation"

|  | $(1)$ <br> Freq <br> (Percent) |
| :--- | :---: |
| First generation | 16,230 |
|  | $(28.72)$ |
| Second generation | 40,283 |
|  | $(71.28)$ |
|  |  |
| Total | 56513 |

Table 6: distribution of the variable "nursery"

|  | $(1)$ <br> Freq <br> (Percent) |
| :--- | :---: |
| Went to nursery | 15,881 |
|  | $(28.10)$ |
| Did not go | 29,511 |
| Missing information | $(52.22)$ |
|  | 11,121 |
| Total | $(19.68)$ |
|  | 56513 |

Table 7: distribution of the variable "kindergarten"

|  | $(1)$ <br> Freq <br> (Percent) |
| :--- | :---: |
| Went to kindergarten | 46,200 |
|  | $(81.75)$ |
| Did not go | 5,726 |
|  | $(10.13)$ |
| Missing information | 4,587 |
|  | $(8.117)$ |
| Total | 56513 |

Table 8: distribution of the nationality of the mother (1) and of the father (2)
$\left.\begin{array}{|lcc|}\hline & \begin{array}{c}(1) \\ \text { Nationality } \\ \text { mother }\end{array} & \begin{array}{c}\text { (2) } \\ \text { Nationality } \\ \text { father } \\ \\ \\ \\ \\ \text { Freq } \\ \text { (Percent) }\end{array} \\ \hline & & \text { Freq } \\ \text { (Percent) }\end{array}\right]$

Table 14: hypothesized signs and results of the control variables

| Variable | Hypothesis 1 | Results | Significance | Hypothesis 2 | Results2 | Significance2 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Dependent |  |  |  |  |  |  |
| Amount of friendships |  |  |  |  |  |  |
| Language spoken at home |  |  |  |  |  |  |
| Independent |  |  |  |  |  |  |
| Sex | 0 | 0 | insignificant | 0 | 0 | insignificant |
| Year | 0 | + | significant at 0.01 | 0 | 0 | insignificant |
| Education Mother | + | + | significant at 0.01 | + | + | significant at 0.01 |
| Education Father | + | + | significant at 0.01 | + | + | significant at 0.01 |
| European non-EU immigrant |  |  |  | + | + | significant at 0.01 |
| Age at immigration | + | + | insignificant | + | + | insignificant |
| Is a first generation immigrant | - | - | significant at 0.01 | - | - | significant at 0.01 |
| Went to nursery | + | + | insignificant | + | + | significant at 0.01 |
| Went to kindergarten | + | + | significant at 0.01 | + | $+$ | significant at 0.01 |


[^0]:    ${ }^{1}$ A two-step minimum distance method (see Das and van Soest, 1999) or the Ferrer-i-Carbonell and Frijters (FF) estimator (Ferrer-i-Carbonell and Frijters, 2004) are frequently used in the literature (see Booth and van Ours, 2008) but are heavily criticized (Baetschmann, 2011).

[^1]:    ${ }^{2}$ The correlation between going to the nursery and going to the kindergarten is equal to 0.095 . A majority of the children who went to the nursery went also to kindergarten but that amount is very low. Most children did not go to the nursery but went to kindergarten. Both variables are not multicollinear.

[^2]:    ${ }^{3}$ A summary of the results can be found in table 14 of the appendix.

