I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11485

The Impact of Language on Socioeconomic Integration of Immigrants

Aslan Zorlu
Joop Hartog

APRIL 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11485

The Impact of Language on Socioeconomic Integration of Immigrants

Aslan Zorlu
University of Amsterdam and IZA

Joop Hartog
University of Amsterdam and IZA

APRIL 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Impact of Language on Socioeconomic Integration of Immigrants*

This study examines the causal effects of Dutch language proficiency of immigrants from four main source countries on their labour market and social integration outcomes. Language proficiency appears ranked according to linguistic distance to The Netherlands, a ranking that even holds for the gender gap in proficiency. We assess the effect of language proficiency on two objective indicators of integration (employment and income) and two subjective measures (feeling Dutch and feeling integrated). The analysis shows that endogeneity of language skills masks a substantial part of language effects. Once accounted for endogeneity, effects of Dutch language proficiency on social and economic integration of immigrants are more than double the estimates ignoring endogeneity.

JEL Classification: J15

Keywords: language skills, immigrants, integration, treatment effects

Corresponding author:
Aslan Zorlu
University of Amsterdam
Faculty of Social and Behavioural Sciences
Plantage Muidergracht 14
1018 TV Amsterdam
The Netherlands
E-mail: A.Zorlu@uva.nl

* Hartog acknowledges financial support from the Spanish Ministry of Economy and Competitiveness (grant# ECO2014-59055-R).

1. Introduction

Proficiency in local dominant language is an essential first step toward upward socioeconomic mobility of immigrants in the host countries. Language skills are not only a form of human capital to enhance individual productivity in economic actions (Chiswick and Miller 1995) but also a gateway for immigrants entering the receiving society. Language skills create a foothold for immigrants to explore opportunities and to integrate in the host society.

This study examines the causal effects of Dutch language proficiency of immigrants on their labour market and social integration outcomes. Comparing two different (linguistic) groups of immigrants, i.e. immigrants from Turkey and Morocco (Mediterranean) and immigrants from Suriname (a former colony of the Netherlands) and Dutch Antilles (still part of the Netherlands), the impact of a small language, Dutch, is assessed. Different from big languages such as English, Spanish, French and German, Dutch is a small language with a limited geographical coverage. This implies that Mediterranean immigrants will learn Dutch mostly after immigration and their incentives for learning Dutch would not be high when they intend to leave the Netherlands within a short period. In contrast, a vast majority of Surinamese and Antillean immigrants already speaks Dutch due to the colonial history. We explore this contrast to uncover the effects of language, using instruments affecting acquisition of language, such as age at migration, homeland education and homeland media orientation.[1]

Turkish, the official language of Turkey, belongs to the Altaic languages, a language family that also includes Mongolean, Korean and Japanese. In Morocco the official language, taught at school and used in official publications is standard Arabic, an Afro-Asiatic language. Mostly spoken is a dialect, Moroccan Arabic. But French is a frequently used second language, in business, government, shops and restaurants, and many TV programs are in French. At the Dutch Antilles, Dutch is still the dominant language in government and education, but the dominant mother tongue is Papiamento, a language with Iberian roots (Portuguese and Spanish). Less than 10 percent of the population has Dutch as mother tongue, and Dutch is generally experienced as an undesired colonially imposed foreign language. In Suriname, the position of Dutch is quite prominent. It's the official language since independence in 1975 and the mother tongue of over 60% of the population. Suriname is a multi-ethnic society, with many languages. Most inhabitants are fluent in both their ethnic language, Dutch and often also in Sranam, the other lingua franca. They take pride in their own species of Dutch. Based on these background facts[2], one would easily be inclined to rank immigrants by declining language distance to Dutch as Turks, Moroccans, Antilleans and Surinamese. And this is precisely the ranking of immigrants by Dutch proficiency (see Table A2, Appendix).

Since language skills have been conceptualized as a form of host-country specific human capital in the economic literature, effects of language acquisition have been extensively studied on labour market performance of immigrants (see Chiswick and Miller 2014 for an overview, Carliner 1981; McManus et al. 1983). Most empirical studies have predominantly examined the impact of English language proficiency on earnings of male immigrants in traditional immigration countries (Chiswick and Miller 1995; Dustmann and Fabbri 2003; Bleakley and Chin 2004; Miranda and Zhu 2013). Yet,

[1] The percentage shares of first and second generation immigrants in Dutch population are 2.4 for Turks, 2.3 for Moroccans, 2.1 for Surinamese and 0.9 for Antilleans.

[2] Information on languages was taken from Wikipedia and a number of websites specialised in language information.

few studies examine effects of German (Dustmann 1994; Dustmann and van Soest 2001, 2002; Aldashev et al. 2009), Hebrew (Chiswick 1998; Berman et al. 2003), Spanish (Budra and Swedberg 2012; Isphording 2013), Catalan (Di Paolo and Raymond 2012; Rendon 2011) and Dutch (Chiswick and Wang 2016; Yao and van Ours 2015). A few recent studies examined the effect of second language on social outcome of immigrants, such as health and demographic outcomes, children's education and residential choice (Bleakley and Chin 2010; Chen 2013; Guven and Islam 2015).

Immigrants' acquisition of second language is a function of their pre- and post-migration characteristics and in turn, language skills determine their socioeconomic outcomes. There is a broad consensus that language skills and socioeconomic outcomes are mutually determined for immigrants. Language skills are likely correlated with unobserved characteristics of immigrants, such as motivation, innate abilities, financial and cultural resources, as well as some contextual factors in the host country such as availability and quality of local language courses and density of encounters with native speakers. These unobserved factors are among the main sources of endogeneity. Another source is potential measurement errors in self-assessed language proficiency (Dustmann and van Soest 2002).

We use an instrumental variable approach to estimate the causal effects of Dutch language proficiency (Chiswick and Miller 1995; Yao and van Ours 2015; Guven and Islam 2015). We basically use three instruments: age at arrival, homeland education and having satellite antenna at home. The first variable is a well-established instrument, which is justified from linguistic evidence on a critical period of second language acquisition (Bleakley and Chin 2004, 2010; Chiswick and Miller 2008; Yao and van Ours 2015). We prove the role of age at migration in determining Dutch language proficiency and therewith reinforce its validity as an instrument. The other variables, homeland education and having satellite antenna at home, have not been used as instruments for language acquisition before, to the best of our knowledge. Homeland education is strongly correlated with post-immigration investment (Van Tubergen and Van de Werfhorst 2007). Language is likely a first part of this investment. The acquisition of host country language is more necessary for higher educated immigrants to be able to perform their original occupation as it more likely requires language skills, compared to low skilled occupations. Pre-migration education functions as an indicator of efficiency, lower opportunity costs and higher motivation to learn host country language. The variable "having satellite antenna" is associated with watching TV channels from homeland in mother tongue and points to a high intensity of mother tongue usage at home, possibly at the expense of host country language. This variable largely captures a high degree of home country orientation of immigrants and little inclination to learn host country language, assuming that media preferences reflect language preferences.

This paper contributes to the literature of language effects in various ways: First, we estimate determinants of the proficiency of a small language (Dutch) using very rich surveys. Second, this study examines the role of language proficiency in shaping the self-assessed integration outcome and feeling-Dutch in addition to two major structural determinants of integration, employment and income. Third, we explore differences between immigrants from former colonies and immigrants from Turkey and Morocco that are substantially different from the host country regarding language, cultural norms and economic development. Fourth, we account for endogeneity of language skills by using a well-established instrument, age at migration and two new instruments: pre-migration education and having a satellite antenna at home.

In this article, we use national survey data among four large ethnic groups in the Netherlands (i.e. Turks, Moroccans, Surinamese and Antilleans). Turkish and Moroccan immigrants arrive in the country without knowledge of the Dutch language and with a significant social and cultural distance from the host society. Surinamese and Antillean (Caribbean) immigrants are, on the other hand, familiar with the Dutch society due to colonial relations, even though they are not all fluent in Dutch.

The paper is organized as follows. First, in section 2, we give an overview of the relevant literature and derive our hypotheses. In section 3, we introduce our data and describe our variables. In section 4, we present our econometric strategy. In section 5 we examine determinants of language proficiency and the validity of our instruments. In section 6, we estimate the effects of language proficiency on integration outcomes of immigrants. Section 7 concludes.

2. Literature

Language proficiency

Empirical research on the second-language proficiency of immigrants relies on three conceptual variables: exposure to the host country language, economic incentives and efficiency in learning a new language (Chiswick and Miller 1991, 1995, 2014). *Exposure* refers to opportunities to hear, read speak and write the destination language. Exposure of Turkish and Moroccan immigrants to the host country language occurs predominantly in the Netherlands while a large share of immigrants from (former) colonies are already fluent in the destination language. The intensity of exposure in the host country is likely smaller for Turkish and Moroccan immigrants who live in the neighbourhoods where their co-ethnics are concentrated and where they can speak their mother tongue. Exposure increases with the duration of residence, the intensity of contacts with native speakers, and the younger the age at which immigrants arrive. Exposure may be more intense when there are children at school-going age in the household.

Economic incentives arise from lower costs of job search and cost of consumption, a higher rate of employment and a higher wage rate associated with a higher level of language proficiency. Economic incentives will increase with expected duration of residence and decrease with a stronger inclination to return and with stronger orientation toward the origin country which is associated with a low inclination to invest in host country-specific capital.

Efficiency refers to the ability of converting exposure into language proficiency. Empirical evidence suggests that the efficiency of learning the new host country language is inversely related to age at migration. Immigrants arriving at younger ages appear to learn language faster, as predicted by the theories of critical age of language acquisition. Efficiency may also positively relate to the level of pre-migration education. This efficiency arises from higher learning ability, learning skills acquired from higher schooling and origin country language skills (Chiswick and Miller 2003; Dustmann 1994; Isphording and Otten 2014). Linguistic distance between the origin and destination languages is another key variable. Immigrants with a closer linguistic origin face relatively low costs in acquiring the host-country language. If the mother tongue is very different from the destination language, a low efficiency is expected in learning the destination language (Isphording and Otten 2014).

Our data largely cover these three conceptual variables. *Exposure* is captured by two types of variables: first, a differential exposure effect in the country of origin among Caribbean (Surinamese and Antillean) and Mediterranean (Turkish and Moroccan) immigrants. A vast majority of Caribbean

migrants already speaks Dutch before migration, due to colonial relations while Mediterranean immigrants, who are from a distant linguistic background, have to learn Dutch in the Netherlands. Another part of exposure is approximated by intensity of contacts with native Dutch people outside the household and by contacts with Dutch speaking people within the household, i.e. children at school age and Dutch partner. Duration of residence in the host country (YSM) indicates the length of potential exposure to the host country language.

Economic incentives are related to inclination to return to the source country. This will be indicated by orientation toward the home country, and captured by having a TV-satellite at home and suffering from homesickness. Duration of residence in the host country (YSM) will be associated with falling likelihood of return migration and as such point to increased benefit of language proficiency.

Efficiency in learning Dutch is captured by variables measuring age at migration and education acquired in the home country. Age at migration is a strong predictor of learning a second language. Empirical studies indicate a clear inverse relationship between age at migration and language proficiency (Chiswick and Miller 2014; Isphording and Otten 2014; Bleakley and Chin 2010). Homeland education is likely an indicator of ability in processing information and learning new concepts, so higher educated immigrants will learn Dutch faster and better than low educated immigrants.

Effects of language

A vast majority of studies from predominantly English speaking countries on effects of language proficiency covers wage returns to the dominant language in the destination country of immigrants. Little attention has been paid to other integration outcomes. The primary focus on earnings has arisen from interest in economic integration of immigrants. Chiswick and Miller (2014) report an overview of the existing studies to measure language effects on earnings. Language proficiency leads to increase in earnings of immigrant males between 5 to 35%. Chiswick and Miller note that OLS estimates of language effects in Mincerian earnings functions provide structurally lower rates of language premium than studies accounting for endogeneity of language acquisition. Bleakley and Chin (2010) examined the effect of English language proficiency on marriage, fertility, and children's education and residential choice while Chen (2013) provided evidence on the link between language skills and health and wealth at retirement. Guven and Islam (2015) studied the impact of English language proficiency on the individual's health and satisfaction with partners and jobs as well as on a range of children's outcomes, such as test scores, club memberships, socialization, and their perceived importance of sports and hobbies.

3. Data

We use the Dutch Survey on the Integration of Minorities 2006 and 2011 (SIM 2006 and 2011). These survey data were collected among the four largest minority groups in the Netherlands by the Netherlands Institute for Social Research (SCP). They include random samples of Turkish, Moroccan, Surinamese and Antillean immigrants aged 15 years and older, drawn from the population records and cover both first and second generation immigrants (at least one parent born in the source country). We only use observations on individuals not born in the Netherlands. The data for Turks and Moroccans have been collected by bilingual interviewers and provide a self-assessed measure of language proficiency. Turkish and Moroccan participants who were expected not to be fluent in Dutch were interviewed in their native languages. The overall response rate is 53%, ranging from

46% for Surinamese to 60% for Turks[3]. Definitions of variables are given in Appendix Table A1 and summary statistics in Appendix Table A2.

The assessment of language proficiency is based on self-evaluations of the respondents, identifying the difficulties immigrants experience with speaking Dutch. The original variable includes the following categories[4]: (1) speaks no Dutch, (2) experiencing often difficulties (3) experiencing sometimes difficulties (4) experiencing no difficulties. Self-reported language skills is possibly subject to reporting errors (Chiswick and Miller 2014). In order to reduce potential misclassification errors when using a fine distinction, we dichotomize these categories into experiencing difficulties (1, 2 and 3) versus no difficulties (4). Accordingly, our binary variable takes on the value 1 if an immigrant has no difficulties in speaking Dutch, and 0 otherwise[5].

We use four dependent variables to measure integration outcomes of immigrants. First, two variables related to labour market position: a dummy variable for whether or not employed and a variable of household income, measured as eight classes of monthly household income in an interval of 500 Euros up to 3500 and a separate interval for amounts higher than 3500 Euro. As this is household income including both labour earnings and social transfers (unemployment, disability, welfare) the income measure is a measure of economic status and integration rather than individual productivity. Among interviewed men in a household with a partner, about ¾ of them are employed, among interviewed women in a household with a partner about half of them are employed. The sample contains only 68 children living in the parental household[6].

Two variables measure self-identification of immigrants. Respondents were asked to indicate whether they feel more or less Dutch. We dichotomize this variable by coding feeling Dutch (1) versus feeling ethnic (0). Immigrants who feel either Dutch or equally Dutch and ethnic are coded as 1 while immigrants who considered themselves as more ethnic than Dutch take 0. The last variable is a self-assessed integration variable, based on the question: Do you think you are well integrated? We dichotomize this variable by coding the answer "enough integrated" as 1, and 0 otherwise.

4. Empirical Strategy

A general framework

Level of proficiency in Dutch language D_i^*, a latent continuous variable, is determined by a vector of variables measuring exposure, economic incentives and efficiency to learn Dutch. However, we observe proficiency only as a binary variable D_i, good or bad as specified below:

$$D_i^* = \beta_D x_{1i} + \gamma_i z_i + \varepsilon_{Di} \ , \ D_i = 1 \text{ if } D_i^* > 0, \ D_i = 0 \text{ otherwise} \qquad (1)$$

[3] For more information about the survey, see Dagevos et al. 2007

[4] The original question in the survey is the following: Do you, when you have a conversation in Dutch, often, sometimes or never experience difficulty with the Dutch language?

[5] Charette and Meng (1994) collected data on French/English proficiency among Canadians from a written test and from self-assessment and found that the effect of measurement errors in self-assessed proficiency in an earnings function was annihilated when proficiency was instrumented (with inter alia education and age when the new language was learned).

[6] Sample size is 4169. Removing the 68 children has no effect on the income equations estimated below.

where x_{1i} and z_i are vectors of variables and ε_{Di} is a random error term. With ε_{Di} normally distributed, this is a probit model.

Host country language proficiency is a main determinant of integration outcomes y_i. We specify a general regression equation for the determinants of outcomes:

$$y_i^* = \delta D_i + \beta_{1y} x_{1i} + \beta_{2y} x_{2i} + \varepsilon_{yi} \quad (2)$$

We have two vectors of explanatory variables for outcomes, x_{1i} joint with equation (1) and x_{2i} disjoint with equation (1); the variables in vector z_i are excluded from equation (2), with some of them used as instruments for language proficiency D_i. We have four measures of outcome: income, self-identified as Dutch or ethnic, feeling integrated and employment. In case of income, we have a cardinal variable measured in intervals and we take $y_i = y_i^*$. In case of self-identification as Dutch/ethnic and feeling integrated y_i^* is a latent ordinal variable and the observed outcome y_i a binary variable with observation $y_i = 1$ if $y_i^* > 0$, i.e. a probit model if we assume ε_{yi} to have the normal distribution. In case of employment, the probability of employment has density function $f(\varepsilon_{yi})$ and $E(y_i^*)$ given by the deterministic part of equation (2). With a normal distribution for $f(\varepsilon_{yi})$, the binary model for observed employment status is again a probit model.

We first estimate probit model (1) for language proficiency. We then estimate outcome equation (2) under the assumption that ε_{Di} and ε_{yi} are correlated, with a bivariate probit model with an endogenous binary regressor (Angrist and Pischke 2009:197-205; Abadie 2003) for employment, feeling Dutch/ethnic and feeling integrated, and linear regression for income; language proficiency is instrumented by some variables in the vector z_i. To assess the effect of allowing for endogenous language proficiency, we will also estimate (2) under the assumption that ε_{Di} and ε_{yi} are uncorrelated.

An alternative approach

Alternatively, we also use a control function approach with endogenous treatment effects to estimate effects of language proficiency on integration outcomes (Blundell, Kristensen and Matzkin 2013; Imbens and Wooldridge 2009; Wooldridge 2015). The control function estimator uses residuals from the treatment model in the model for the potential outcome to estimate the average treatment effect (ATE) from observational data when treatment assignment is endogenous.

In the control function approach, we consider immigrants who are proficient in Dutch as the treated group. The potential integration outcomes of immigrants who are proficient in Dutch, y_{i1} and who are not, y_{i0} are determined by their expected value conditional on a vector of variables $\boldsymbol{x}_i$ and an unobserved random component ε_{ij}, with $j \in \{0,1\}$

$$y_{i1} = E(y_{i1}|\boldsymbol{x}_i) + \varepsilon_{i1}$$

$$y_{i0} = E(y_{i0}|\boldsymbol{x}_i) + \varepsilon_{i0}$$

Where $\boldsymbol{x}_i$ denotes a vector of variables affecting integration outcome. The Dutch language proficiency D_i is similarly determined by its expected value conditional on a set of conceptual variables $\boldsymbol{z}_i$ measuring exposure, economic incentives and efficiency and an unobserved random component ϑ_{ij}, with $j \in \{0,1\}$.

$$D_i = E(D_i|\boldsymbol{z}_i) + \vartheta_i$$

Where ϑ_i is an unobserved random component. In treatment effects setting, the unobserved random components in the potential outcomes are independent of $\boldsymbol{z}_i$, $E(\varepsilon_{ij}|\boldsymbol{x}_i, \boldsymbol{z}_i) = E(\varepsilon_{ij}|\boldsymbol{z}_i) = E(\varepsilon_{ij}|\boldsymbol{x}_i) = 0$ for $j \in \{0,1\}$. In case of endogenous language proficiency, unobserved components of the outcome equation are correlated with language proficiency, so that $E(\varepsilon_{ij}|D_i) \neq 0$ for $j \in \{0,1\}$.

Since we observe just one outcome for an immigrant, the potential outcome model (POM) is given as (Cerulli 2015; Imbens and Wooldridge 2009)

$$y_i = D_i y_{i1} + (1 - D_i) y_{i0}$$

The first term, $D_i y_{i1}$, denotes the integration outcome of immigrant i who is proficient in Dutch. Similarly, the second term, $(1 - D_i) y_{i0}$, denotes the outcome of immigrant i who is not proficient in Dutch. An immigrant can either speak Dutch well or not, but not both at the same time. Thus only one of these outcomes will be realized. For Dutch speaking immigrants, the second term is an ex post counterfactual outcome while the first term is counterfactual outcome for immigrants who are not proficient in Dutch.

The potential outcome model links unobservable with observable outcomes. The causal model of the realized integration outcome is given as an average treatment effect (ATE)

$$ATE = E(y_{i1} - y_{i0})$$

Acknowledging the dichotomous structure of endogenous language proficiency, we estimate endogenous treatment effects models with a probit estimator for the language equation and for binary outcomes, and an OLS estimator for the income outcome.

5. Proficiency in Dutch

Firstly, we aim to understand how the Dutch language proficiency is related to a set of background variables. We estimate binary probit model (1) with maximum likelihood estimator, including a full set of variables capturing the three broad concepts: exposure, efficiency and economic incentives. The estimated parameters are presented in Table 1. Both for men and for women, language proficiency by origin country inversely follows distance to Dutch (Surinamese, Antilleans, Moroccans, Turks) and is higher for men than for women, which can be understood from incentives and from exposure associated with higher labour force participation. Interestingly, the gap between male and female proficiency follows the ranking by language distance as a dichotomy (smallest for Surinamese and Antilleans, largest for Moroccans and Turks). The probability of Dutch proficiency increases with the duration of stay (YSM) and decreases with age at migration. As the positive effect of YSM is linear and the effect of later migration is quadratic, migration at higher age initially has negative effect on language proficiency at given age, and later a positive effect (with turning point at migration age 48). The initially negative effect of later migration may be countered by more

homeland education. If higher education would take from age 18 to age 24 to complete and migration would shift from age 18 to age 24, the balance of that shift would be negative. Immigrants who came for marriage have no different language proficiency than immigrants who came for work. The effect of other motives than work is substantial. Coming with parents or for reunification probably picks up another part of language advantage for children, study surely requires language proficiency and social advantages as migration motive presumably signals a desire to remove language barriers.

The probability of Dutch proficiency is negatively correlated with having TV-satellite at home and being homesick and it is positively associated with being married to a native Dutch partner. The probability of Dutch proficiency is almost 8 percentage points lower for an immigrant living in a home with TV-satellite compared to an immigrant without TV-satellite. Different than expected, the presence of minor aged children in the home, assumed to be associated with exposure, has no significant effect on language proficiency. The substantial improvement in proficiency between survey years is remarkable; it may be related to increased pressure to take courses for introduction to Dutch society, including language training.

Table 1. Probit estimations of Dutch language proficiency (Standard error in parentheses)

		coefficient	Marg.Eff
	Age at migration	-0.107***	-0.012
		(0.007)	(0.001)
	Age at migration squared	0.001***	
		(0.000)	
	Years Since Migration (YSM)	0.011***	0.003
		(0.002)	(0.001)
Homeland Educ	Primary		
	Unknown	-0.013	-0.003
		(0.064)	(0.016)
	Secondary	0.392***	0.097
		(0.057)	(0.014)
	High	0.599***	0.146
		(0.073)	(0.018)
Origin by gender	Surinamese Man		
	Turkish Man	-1.302***	-0.360
		(0.121)	(0.031)
	Moroccan Man	-0.964***	-0.258
		(0.121)	(0.031)
	Antillean Man	-0.163	-0.037
		(0.124)	(0.028)
	Surinamese Woman	-0.139	-0.031
		(0.115)	(0.026)
	Turkish Woman	-1.586***	-0.442
		(0.126)	(0.032)
	Moroccan Woman	-1.271***	-0.350
		(0.126)	(0.032)
	Antillean Woman	-0.194	-0.044
		(0.121)	(0.027)
Reason for migration	Work		
	Study	0.307**	0.079

		(0.094)	(0.025)
	Parents	0.460***	0.117
		(0.089)	(0.023)
	Marriage	0.029	0.008
		(0.087)	(0.023)
	Reunification	0.337***	0.087
		(0.077)	(0.020)
	Social advantages	0.353***	0.091
		(0.093)	(0.024)
	Other	0.412***	0.105
		(0.097)	(0.025)
Survey year	year=2006		
	2011	0.340***	0.080
		(0.048)	(0.011)
Number of	No child		
Children	1-2 children	0.068	0.016
		(0.052)	(0.012)
	3 to 7 children	0.083	0.020
		(0.066)	(0.016)
Home country	TV-satellite	-0.293***	-0.070
orientation		(0.066)	(0.016)
	Homesick	-0.207***	-0.050
		(0.047)	(0.011)
	Naturalized	0.473***	0.113
		(0.063)	(0.015)
Interethnic	Same origin partner		
Marriage	Single	0.112*	0.028
		(0.057)	(0.014)
	Dutch partner	0.395***	0.094
		(0.080)	(0.019)
	Other origin partner	0.111	0.027
		(0.127)	(0.031)
	Constant	1.267***	
		(0.182)	
	Pseudo R squared	0.371	
	BIC	4872.912	
	N	5468	

To illustrate the impact of age at migration, we plot the predicted probabilities over age at migration variable ranging from 0 to 16, holding the other covariates at their observed values, by country of origin, see Figure 1a. Language proficiency decreases monotonically with age at immigration. This pattern is much sharper for immigrants from Turkey and Morocco, compared to Antillean and Surinamese immigrants. It confirms the earlier empirical evidence for the prediction of the critical period hypothesis for learning a language (Chiswick and Miller 2008). Learning a new language is easier for younger immigrants and the effect is much stronger for greater language distance. Figures

1b and 1c show that there is also strong interaction with having a TV-satellite at home and with level of homeland education[7].

Figure 1a

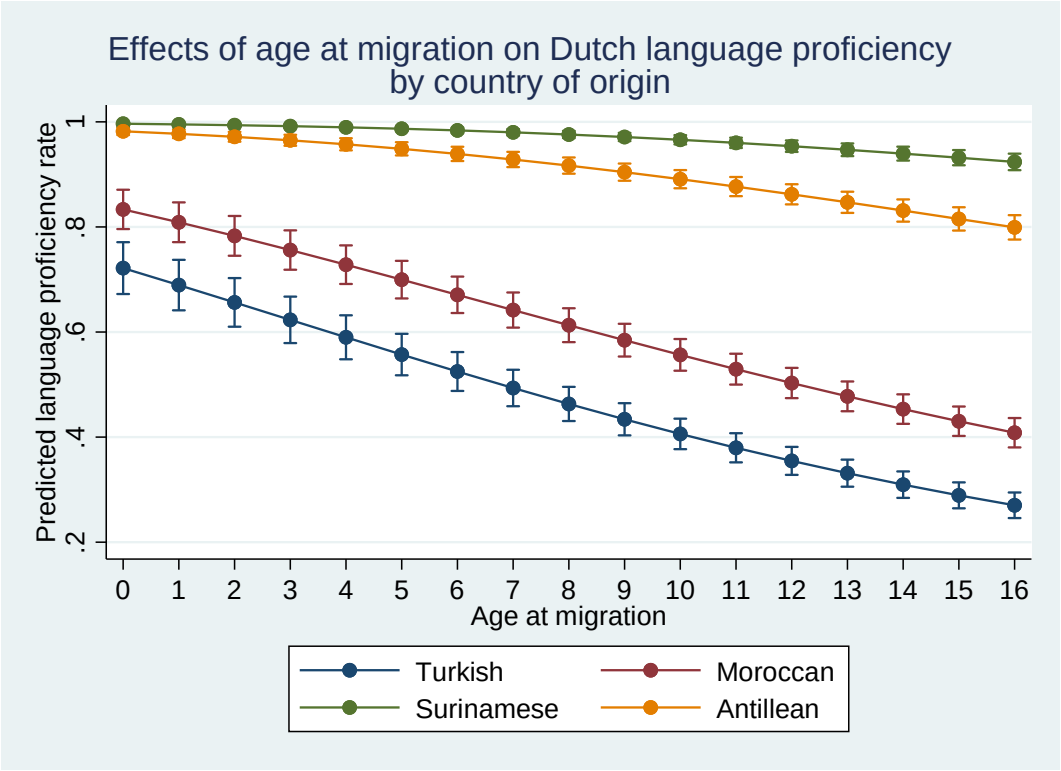


Figure 1b

[7] The depicted interaction with homeland education of course is a synthetic construct, as age of immigration and homeland education cannot independently vary.

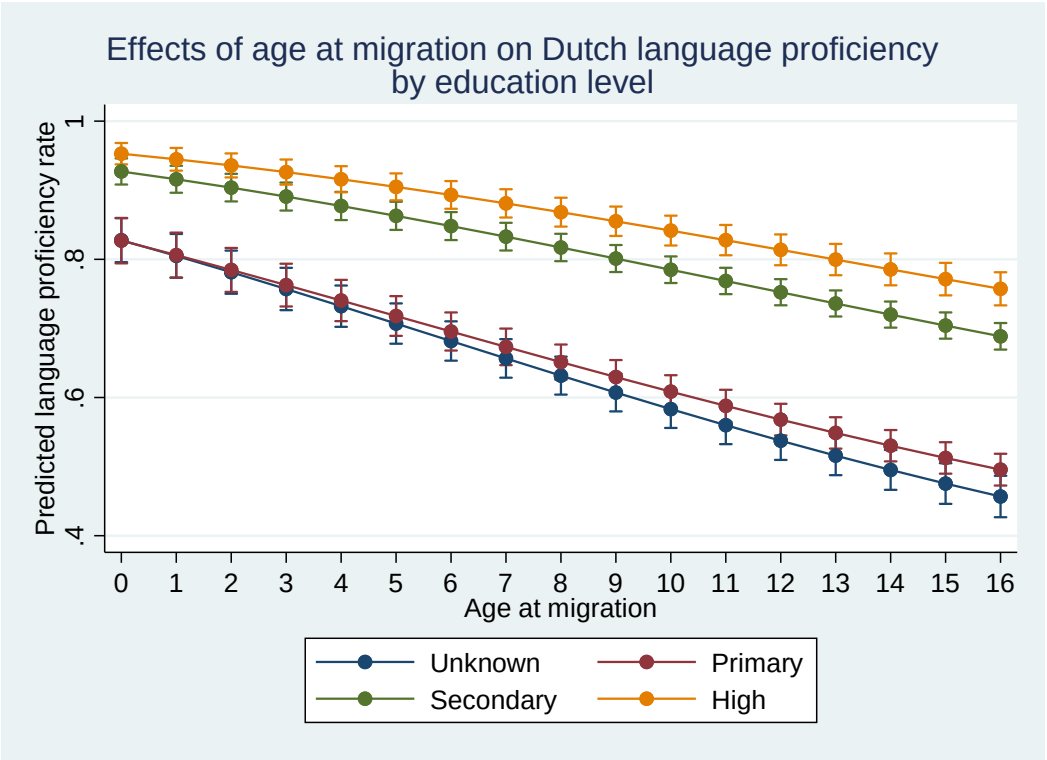
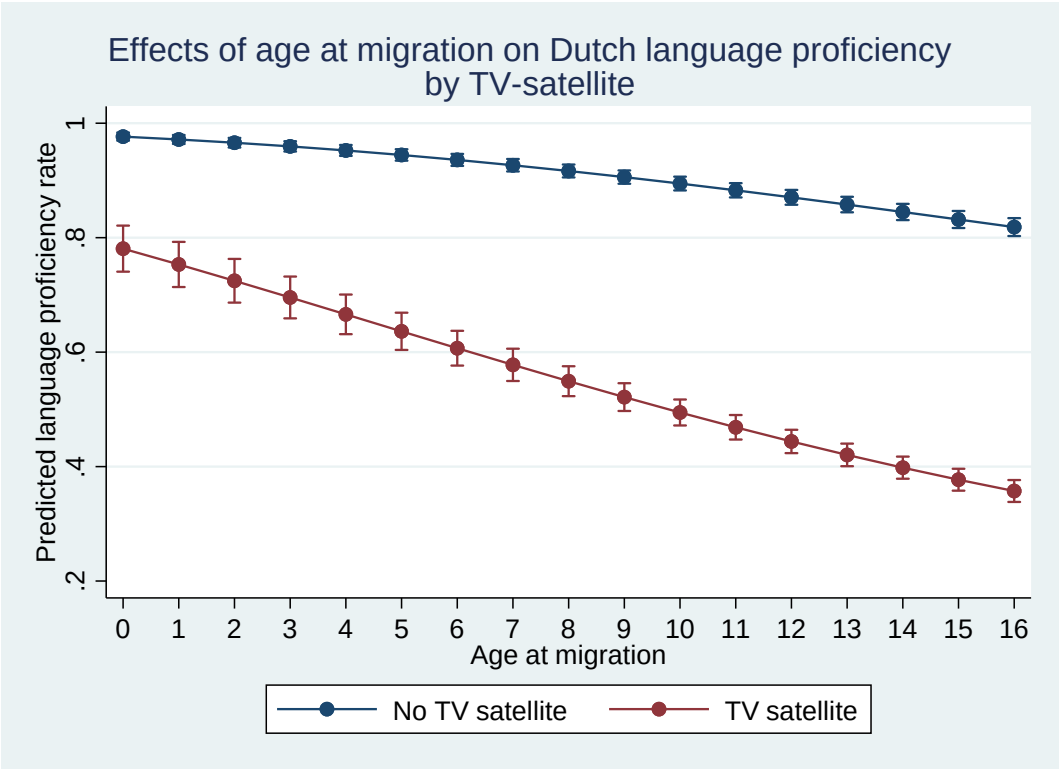

Effects of age at migration on Dutch language proficiency by education level
Predicted language proficiency rate
.4
.6
.8
1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Age at migration
Unknown
Primary
Secondary
High

Figure 1c


Effects of age at migration on Dutch language proficiency by TV-satellite
Predicted language proficiency rate
.2
.4
.6
.8
1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Age at migration
No TV satellite
TV satellite

6. The impact of language proficiency

To assess the impact of language, we start with estimating baseline binary probit models for binary integration outcomes: employment, feeling Dutch, and perceived integration and an OLS model for household income, assuming that language proficiency is exogenous to the integration outcomes. The estimates of these models are given in Table A3 in Appendix. We see clear distinction between men and women in our integration variables (women score lower than men) and among the ethnic groups (Antilleans are comparable to Surinamese, Moroccans and Turks perform worse). Language proficiency is statistically significantly related to all four outcome variables; these estimates are only given for comparison with our estimates acknowledging endogeneity.

In the next step, we estimate bivariate probit models with a binary endogenous variable for our binary integration outcomes and a linear regression model for household income. In all models, language proficiency is instrumented by age at migration, home country education and having TV satellite[8]. The parameter estimates of these models are given in Table 2. The estimated parameter for Dutch language proficiency is statistically significant and large in magnitude, indicating a substantial impact of language on the four integration outcomes of immigrants. The coefficients of Dutch proficiency are two to three times larger than the coefficients obtained from the baseline models, indicating the importance of accounting for a substantial endogeneity problem.

Years since migration has positive effect on all integration variables, higher education is associated with higher employment probability, higher household incomes and stronger feeling of integration, but not with intensity of feeling Dutch. Marriage composition has modest effects (higher income with a Dutch partner and, remarkably, feeling more Dutch with a non-Dutch partner). Experiencing discrimination has no significant effect on the economic variables (employment and income), but the effects on subjective integration is negative. Interestingly also, household income has no effect on subjective integration.

Considering differences by origin and gender, we note that employment probabilities do not differ among men, and that they are lower for all women, at about equal gap with Surinamese men for Caribbean women and, at larger distance, for Mediterranean women. In terms of household income, Antillean men are comparable to Surinamese men, while Turks and Moroccans do worse. Among the women, Surinamese, Antilleans and Moroccans do worse in that order, while Turkish women are not worse off (relative to Surinamese men). Subjective integration is only rated lower for Turks, both male and female.

Unobservables for proficiency and outcomes correlate negatively. This may be surprising if one would anticipate unexpectedly high language proficiency to come with unexpectedly good integration outcomes, but the correlation may also reflect the effect of language norms set by the respondent. Better integrated respondents may set higher proficiency norms for themselves, and hence, may judge more easily that they fall short of their own standard.

Table 2. Estimates of integration outcomes with endogenous language proficiency (bivariate probits; linear regression for income)

[8] The estimates for the language proficiency equation are given in Appendix Table A4. They barely deviate from the results in Table 1.

		Employment	HH income	Feel Dutch	Integrated
		Bivar. Prob coefficient	OLS-IV	Bivar. Prob coefficient	Bivar. Prob coefficient
	Speaks well	0.859***	0.898***	1.339***	1.630***
		(0.13)	(0.18)	(0.09)	(0.10)
	YSM	0.017*	0.023**	0.025***	0.021*
		(0.01)	(0.01)	(0.01)	(0.01)
	YSM squared	-0.000**	-0.000	-0.000*	-0.000
		(0.00)	(0.00)	(0.00)	(0.00)
Interethnic	same origin partner (ref.)				
Marriage	Single	-0.381***	-1.558***	0.084	0.026
		(0.05)	(0.06)	(0.06)	(0.07)
	Dutch partner	0.082	0.603***	0.089	0.175
		(0.07)	(0.07)	(0.07)	(0.09)
	other origin partner	-0.001	-0.005	0.265*	0.075
		(0.11)	(0.12)	(0.11)	(0.14)
Naturalized		0.156*	0.068	0.012	0.015
		(0.07)	(0.07)	(0.06)	(0.06)
Discriminated	Never				
	Sometime	0.046	0.039	-0.232***	-0.159**
		(0.04)	(0.05)	(0.04)	(0.05)
	Often	-0.111	-0.123	-0.339***	-0.093
		(0.06)	(0.07)	(0.06)	(0.07)
Origin by	Surinamese Man (ref.)				
gender	Turkish Man	0.013	-0.427***	-0.242*	-0.282*
		(0.11)	(0.12)	(0.10)	(0.14)
	Moroccan Man	-0.130	-0.687***	-0.024	-0.184
		(0.10)	(0.11)	(0.09)	(0.13)
	Antillean Man	-0.094	-0.103	-0.010	-0.227
		(0.11)	(0.11)	(0.10)	(0.14)
	Surinamese Woman	-0.354***	-0.297**	0.100	-0.122
		(0.09)	(0.09)	(0.08)	(0.13)
	Turkish Woman	-0.701***	-0.211	-0.282**	-0.404**
		(0.12)	(0.13)	(0.10)	(0.14)
	Moroccan Woman	-0.911***	-0.561***	0.063	-0.171
		(0.11)	(0.12)	(0.10)	(0.14)
	Antillean Woman	-0.358***	-0.435***	-0.065	-0.111
		(0.11)	(0.11)	(0.10)	(0.14)
Education	Low				
	Low Medium	0.206***	0.261***	0.066	0.163**
		(0.05)	(0.06)	(0.05)	(0.06)
	Medium	0.476***	0.599***	0.077	0.287***
		(0.06)	(0.06)	(0.05)	(0.07)
	High	0.761***	1.503***	-0.008	0.491***
		(0.07)	(0.08)	(0.07)	(0.10)
# Children	No child (ref.)				
	1-2 children	0.136**	0.293***	-0.037	0.060
		(0.05)	(0.05)	(0.05)	(0.06)
	3 to 7 children	-0.003	0.216**	-0.117	0.048
		(0.06)	(0.07)	(0.06)	(0.07)
Age	Age 15-19			0.040	0.394

				(0.17)	(0.24)
	Age 20-35	0.467***	-0.117	-0.075	0.008
		(0.09)	(0.10)	(0.08)	(0.10)
	Age 36-55	0.638***	0.205**	0.091	-0.004
		(0.07)	(0.07)	(0.06)	(0.07)
	Age 56 and older (ref.)				
Survey year	year=2006 (ref.)				
	year=2011	0.026	0.287***	-0.326***	0.100*
		(0.04)	(0.05)	(0.04)	(0.05)
Household	Up to 500 (ref.)				
income	500-1000			0.086	0.029
				(0.11)	(0.13)
	1000-1500			0.172	0.084
				(0.11)	(0.13)
	1500-2000			0.192	0.119
				(0.11)	(0.14)
	2000-2500			0.184	0.136
				(0.12)	(0.15)
	2500-3000			0.209	0.236
				(0.13)	(0.17)
	3000-3500			0.252	0.057
				(0.14)	(0.19)
	3500 and more			0.395**	0.109
				(0.14)	(0.20)
	Employed			-0.010	0.211***
				(0.05)	(0.06)
	_cons	-0.843***	2.795***	-1.168***	-0.730***
		(0.15)	(0.16)	(0.17)	(0.21)
	athrho	-0.319***	-0.275*	-0.780***	-0.623***
		(0.09)	(0.11)	(0.09)	(0.09)
	lnsigma		0.314***		
			(0.02)		
	BIC	9747	18320	9745	8156
	N	4773	4169	4659	4659

Instruments for language proficiency: age at migration, home country education, having TV satellite.

Comparing the ATE estimates from alternative estimators

Table 3 gives an overview of the effects of Dutch proficiency for four integration outcomes calculated from three alternative estimation methods, where estimated coefficients for the language variable have been converted into an average treatment effect (ATE). The first line gives the ATEs that are obtained from the baseline models without accounting for endogeneity (Table A3 in Appendix). The second line presents the ATEs that are estimated using bivariate probit models, as included in Table 2. The last line displays the ATEs that are obtained from endogenous treatment-effects estimation applied to all four outcomes[9].

[9] The estimations of the full models are available on request.

The table shows three clear conclusions. Accounting for endogeneity increases the treatment effect two- to fourfold. Effects estimated with bivariate probit and with endogenous treatment models are very similar. Treatment effects are larger for immigrants who speak well than for immigrants who do not speak well, with smallest gap for household income (20% higher) and largest effect for feeling Dutch (ATE almost four times as large). Clearly, allowing for endogeneity matters.

Table 3. Average treatment effects (ATE)

	Employed	HH income (range 1-8)	Feeling Dutch	Integrated
I. Baseline model				
ATE	0.139	0.323	0.120	0.203
II. Bivariate probit with endog. Dutch				
ATE	**0.289**	**0.898**[a]	**0.474**	**0.455**
III. Endogenous Treatment				
ATE	**0.285**	**0.678**	**0.521**	**0.487**
(speaks well)	0.661	3.991	0.704	0.945
(speaks not well)	0.376	3.313	0.183	0.458

[a] Linear regression with endogenous treatment.

To check the robustness of our ATE estimates, we apply the linear IV estimator disregarding the binary structure of our outcome variables (Table 4). Chiburis et al. (2012) show that bivariate probit and IV estimators can produce different estimates of treatment effects when sample sizes are below 5000 or treatment probabilities are close to zero or unity. This method is an alternative for maximum likelihood estimates of a bivariate probit model in which the outcome and treatment are assumed to be determined by a latent linear index model with jointly normally distributed errors (Chiburis et al., 2012). An inspection of the results of the first stage regression, the probit model of language proficiency indicates a pretty strong correlation of our instruments with language proficiency. We also report the F-statistic for the first stage regression, and use a limited information maximum likelihood estimator (LIML) as a check for over-identification of 2SLS models since LIML has better small sample properties than 2SLS with weak instruments for finite samples (Angrist and Pischke, 2008; Chiburis et al., 2012). F-statistics are far above the critical value 10, as suggested by Stock et al. (2002) and the LIML estimator generates rather higher ATE estimates than the two-step procedure. In addition to the IV estimator, we also use the probit estimator with endogenous Dutch proficiency (Angrist, 2001; Roodman 2011). The estimated average treatment effects (ATEs) from these two methods are quite similar to the earlier estimates. So, our main conclusions survive all these checks: the causal effects of Dutch proficiency are much higher than the estimates obtained from the baseline estimates assuming exogeneity of Dutch.

Table 4. Average treatment effects (ATE)

	Employed	HH income (range 1-8)	Feeling Dutch	Integrated
IV. Linear IV estimator				
ATE (2SLS)	**0.423**	**0.868**	**0.567**	**0.657**
ATE (LIML)	0.423	0.951	1.028	0.742
F-statistic	157.69	136.76	157.69	157.69
V. Probit with endogenous Dutch				
ATE	**0.347**		**0.417**	**0.440**
Speaks well	0.717		0.666	0.922

Speaks not-well	0.370		0.249	0.482
OBSERVED				
Speaks well	0.70	4.12	0.63	0.92
Speaks not-well	0.41	3.25	0.29	0.51

7. Conclusions

The Netherlands received its largest immigration groups from two Mediterranean countries, Turkey and Morocco, as an outflow of massive immigration of unskilled labour initiated half a century ago and from the Caribbean region, Suriname and Dutch Antilles, as a consequence of a long colonial history. Considering positions in the family of languages and linguistic institutions in these source countries, one might assume a ranking by increasing language distance to Dutch as Suriname, Antilles, Morocco, Turkey. And this is indeed the ranking by proficiency in the Dutch language in our sample of immigrants. In multivariate regression analysis this ranking is maintained. Language proficiency is higher for men than for women in all four immigrant groups, which can be understood from incentives and from exposure associated with higher labour force participation. Even the gap between male and female proficiency follows the ranking by language distance, now as a dichotomy (smallest for Surinamese and Antilleans, largest for Moroccans and Turks). Dutch language proficiency rises with the duration of residence and diminishes with age at migration. It is substantially lower for immigrants who have a TV-satellite and who suffer from homesickness.

To measure the impact of Dutch language proficiency on socioeconomic integration outcome it is essential to account for endogeneity. We have analysed the impact on four indicators, employment, household income, feeling Dutch and perceived integration). Dutch language proficiency has a statistically significant and economically substantial effect on the indicators. Allowing for endogeneity increases the estimated average treatment effect two- to fourfold. We estimated instrumental variable models using an established instrument, age at migration, and two new instruments, home country education and having TV satellite. We have also estimated the impact of language proficiency applying endogenous treatment-effects estimation, also known as control function approach. Our conclusion is robust under an alternative estimation strategy, the linear IV estimator as presented by Chiburis et al. (2012).

Our best estimates indicate that the impact of Dutch language on the socioeconomic integration outcome of immigrants is substantial. The Dutch language proficiency increases the probability of being employed about 30 percentage points and raises the household income index by almost one point on the scale of 1 to 8 (to be multiplied by 500 euro in monthly income). The causal effect of Dutch language is even larger for the social integration outcomes: the probabilities of feeling Dutch and perceived integration are about 50 percentage point higher for immigrants who speak Dutch well, compared to immigrants who report difficulties with speaking Dutch.

Overall, this paper has shown that that host country language is extremely important for the socioeconomic integration of immigrants. The role of language is much pronounced for social integration, compared to economic integration. This conclusion holds in particular for two large immigrant groups from a distinct linguistic background, Turkish and Moroccan in the Netherlands where immigrants from these groups can easily maintain daily life without Dutch language in immigrant neighbourhoods of large cities.

References

Abadie, A. (2003). Semiparametric Instrumental Variable Estimation of Treatment Response Models. *Journal of Econometrics*, 113; 231-63.

Aldashev, A., Gernandt, J. and Thomsen, S. (2009). Language usage, participation, employment and earnings: evidence for foreigners in West Germany with multiple sources of selection. *Labor Economics,* 16(3); 330-341.

Angrist, J.D. and Pischke, J.S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Berman, E., Lang K. and Siniver E. (2003). Language-skill complementarity: returns to immigrant language acquisition. *Labor Economics*, 10 (3); 265-290.

Bleakley, H. and Chin, A. (2010). Age at Arrival, English Proficiency, and Social Assimilation among US Immigrants. *Applied Economics*, 2 (1); 165-92.

Blundell, R., Kristensen, D and Matzkin, R.L. (2013). Control Functions and Simultaneous Equations Methods. *American Economic Review*, 103 (3); 563-569.

Budría, S. and Swedberg, P., (2012). The Impact of Language Proficiency on Immigrants' Earnings in Spain. *IZA DP No.* 6957.

Carliner, G. (1981). Wage differences by language group and the market for language skills in Canada. *Journal of Human Resources* 16; 384–399.

Cerulli, G. (2015), *Econometric Evaluation of Socio-Economic Programs: Theory and Applications*. Berlin: Springer.

Charette, M. and R. Meng (1994). Explaining language proficiency: Objective versus self-assessed measures of literacy. *Economics Letters*, 44 (3); 313-321

Chen, K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, 103; 690–731.

Chiburis, R., Das, J.; and Lokshin, M. (2012). A Practical Comparison of the Bivariate Probit and Linear IV Estimators. *Economics Letters*, 117; 762–766.

Chiswick, B. (1991). Speaking, Reading, and Earnings among Low-Skilled Immigrants. *Journal of Labor Economics,* 9(2); 149-170.

Chiswick, B. and Miller, P. (1995). The Endogeneity between language and earnings: international analyses. *Journal of Labor Economics*, 13(2); 246-288.

Chiswick, B. (1998). Hebrew language usage: Determinants and effects on earnings among immigrants in Israel, *Journal of Population Economics*, 11; 253-271.

Chiswick, B. and Miller, P. (2003). The complementarity of language and other human capital immigrant earnings in Canada, *Economics of Education Review*, 22(5); 469-480.

Chiswick B. and Miller, P. (2008). A test for critical period hypothesis for language learning, *Journal of Multilingual and Multicultural Development*, 29(1); 16-29.

Chiswick, B. R. and Miller, P.W. (2014). International Migration and the Economics of Language. *IZA DP No. 7880*

Chiswick, B. and Wang, Z. (2016), Social Contacts, Dutch Language Proficiency and. Immigrant Economic Performance in the Netherlands: A Longitudinal Study. *IZA DP No. 9760*.

Dagevos, J., Gijsberts, M., Kappelhof, J. and Vervoort, M. (2007). *Survey Integratie Minderheden 2006. Verantwoording van de opzet en uitvoering van een survey onder Turken, Marokkanen,*

Surinamers, Antillianen en een autochtone vergelijkingsgroep, The Hague: Netherlands Institute for Social Research.

Di Paolo, A. and Raymond, J.L. (2012). Language knowledge and earnings in Catalonia. *Journal of Applied Economics*, 15(1); 89-118.

Dustmann, C. (1994). Speaking Fluency, Writing Fluency and Earnings of Migrants. *Journal of Population Economics*. 7(2); 133-156.

Dustmann, C. and van Soest, A. (2001). Language fluency and earnings: estimation with misclassified language indicators. *The Review of Economics and Statistics*, 83(4); 663-674.

Dustmann, C. and van Soest, A. (2002). Language and the earnings of immigrants. *Industrial and Labor Relations Review*, 55(3); 473-492.

Dustmann, C. and Fabbri, F. (2003). Language proficiency and labor market performance of immigrants in the UK. *The Economic Journal*, 113; 695-717.

Guven, C. and Islam, A. (2015). Age at migration, language proficiency, and socioeconomic outcomes: evidence from Australia. *Demograph*, 52(2): 513-42 .

Imbens, G. W., and Wooldridge, J.M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47 (1); 5–86.

Isphording, I.E. (2013), Returns to Foreign Language Skills of Immigrants in Spain. *Labor*, 27(4); 443-461.

Isphording, I.E. and Otten, S. (2014). Linguistic barriers in the destination language acquisition of immigrants. *Journal of Economic Behavior and Organization*, 105; 30–50.

McManus, W., Gould, W., Welch, F. (1983). Earnings of Hispanic men: the role of English language proficiency. *Journal of Labor Economics*, 1; 101–130.

Miranda, A. and Zhu, Y., (2013). English Deficiency and the Native-Immigrant Wage Gap. *Economics Letters*, 118 (1); 38-41.

Rendon, S. (2007). The Catalan premium: language and employment in Catalonia. *Journal of Population Economics*, 20; 669-686.

Roodman, D. (2011). Fitting fully observed recursive mixed-process models with cmp. *Stata Journal,* 11; 159–206.

Stock, J. H., Wright, J.H. and Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, 20; 518-529.

Van Tubergen, F and van de Werfhorst, H. (2007), Postimmigration Investments in Education: A Study of Immigrants in the Netherlands, *Demography*, 44 (4); 883-898.

Yao, Y. and van Ours, J.C. (2015). Language skills and labor market performance of immigrants in the Netherlands. Labour Economics, 34; 76-85.

Wooldridge, J.M. (2015). Control Function Methods in Applied Econometrics. *Journal of Human Resources*, 50, 2; 420-445.

Appendix

Table A1. Definitions of variables

variable		structure
Dutch proficiency	Self-evaluations of Dutch language proficiency	Dummy for no difficulties in speaking Dutch
Employment	Having paid work	Dummy for being employed
Income	Self-reported net household income, regardless of sources	8 intervals of 500 Euro
Feeling Dutch	Self-identification of immigrants	Dummy for Feeling more Dutch
Integrated	Self-assessment of integration	Dummy for well integrated
Age at migration	Age when immigrated	Age at migration in years
YSM	Duration of residence in the Netherlands	Years Since Migration in years
		Dummies for
Gender	Gender	Man or Woman
Country of origin	Country of origin	Surinamese, Turkish, Moroccan, Antillean
Reason for migration	Self-reported main reason for migration	Work, study, parents, marriage, family-reunification, social advantages, other
Homeland Education	Education attended in the homeland	Primary, Secondary, High
Number of Children	Children in the household	No child, 1-2 children, 3 to 7 children
TV-satellite	having a TV satellite antenna at home	Having a TV-satellite
Homesick	Perceived homesickness	having homesickness
Naturalized	Dutch national	having Dutch nationality
Interethnic Marriage	Marital status	Single, Same origin partner, Dutch partner, Other origin partner
Discriminated	Perceived discrimination	Never, Sometime, Often
Education	Completed education level	Low, low-medium, medium, high

Table A2. Mean values of variables by ethnic origin

		Turkish	Moroc	Surinamese	Antillean
		Mean	Mean	Mean	Mean
Integration	Dutch Proficiency	0.32	0.44	0.92	0.78
Outcome	Employed	0.46	0.44	0.65	0.59
	Household income	3.65	3.39	4.31	3.80
	Feeling Dutch	0.27	0.42	0.75	0.57
	Integrated	0.56	0.68	0.95	0.87
Migration	Years Since Migration (YSM)	26.23	25.46	31.01	22.27
related	Age at migration	17.21	17.76	17.06	19.48
	Homesick	0.49	0.32	0.16	0.19
	TV Satellite	0.90	0.86	0.08	0.10
	Naturalized	0.65	0.73	0.96	0.02
Home	Unknown	0.24	0.50	0.15	0.11
education	Primary	0.42	0.19	0.24	0.23
	Secondary	0.25	0.20	0.39	0.42
	High	0.09	0.11	0.22	0.24
Children	No-child	0.32	0.31	0.49	0.54
	Children 1-2	0.47	0.37	0.42	0.37
	Children 3-7	0.21	0.33	0.09	0.09
Marital status	Single	0.19	0.22	0.43	0.55
	Married mono ethnic	0.72	0.69	0.37	0.19
	Married to Dutch	0.08	0.07	0.17	0.20
	Married to Other	0.01	0.03	0.03	0.06
Reason for	Work	0.19	0.21	0.05	0.10
migration	Study	0.01	0.02	0.13	0.34
	Parents	0.15	0.23	0.31	0.20
	Marriage	0.26	0.19	0.05	0.02
	Reunification	0.32	0.30	0.13	0.07
	Social advantage	0.04	0.02	0.19	0.15
	Other	0.03	0.04	0.14	0.12
Discrimination	Never	0.62	0.65	0.61	0.55
	Sometime	0.29	0.26	0.29	0.32
	Often	0.09	0.09	0.10	0.13
Education	Lower secondary (vbo/mavo)	0.24	0.17	0.26	0.26
	Upper secondary (mbo/havo/vwo)	0.20	0.20	0.32	0.37
	Tertiary (hbo/wo)	0.07	0.08	0.21	0.20
Age	Age 15-19	0.03	0.03	0.03	0.07
	Age 20-35	0.28	0.28	0.16	0.31
	Age36-55	0.50	0.50	0.55	0.42
	Age 56+	0.19	0.19	0.26	0.20
Gender	woman	0.52	0.50	0.56	0.53
	N	1,429	1,397	1,240	1,402

Table A3. Estimates of integration outcome (without accounting for endogeneity)

		Employment	HH income	Feel Dutch	Integrated
		Probit	OLS	Probit	Probit
	Speaks well	0.424***	0.323***	0.349***	0.797***
		(0.05)	(0.06)	(0.05)	(0.06)
	YSM	0.029***	0.032***	0.048***	0.038***
		(0.01)	(0.01)	(0.01)	(0.01)
	YSM squared	-0.001***	-0.000	-0.000***	-0.000
		(0.00)	(0.00)	(0.00)	(0.00)
Interethnic	same origin partner				
Marriage	Single	-0.362***	-1.530***	0.156**	0.080
		(0.05)	(0.06)	(0.06)	(0.07)
	Dutch partner	0.140*	0.648***	0.182**	0.275**
		(0.07)	(0.07)	(0.07)	(0.10)
	other origin partner	0.082	0.011	0.324**	0.115
		(0.11)	(0.12)	(0.11)	(0.15)
Naturalized		0.250***	0.169**	0.176**	0.161*
		(0.06)	(0.07)	(0.06)	(0.06)
Discriminated	Never				
	Sometime	0.057	0.039	-0.261***	-0.171**
		(0.05)	(0.05)	(0.05)	(0.05)
	Often	-0.114	-0.121	-0.372***	-0.095
		(0.07)	(0.07)	(0.07)	(0.08)
Origin by	Surinamese Man				
gender	Turkish Man	-0.078	-0.664***	-0.736***	-0.737***
		(0.10)	(0.10)	(0.09)	(0.13)
	Moroccan Man	-0.172	-0.856***	-0.388***	-0.535***
		(0.09)	(0.09)	(0.09)	(0.13)
	Antillean Man	-0.022	-0.059	0.043	-0.199
		(0.11)	(0.11)	(0.10)	(0.15)
	Surinamese Woman	-0.372***	-0.296**	0.122	-0.115
		(0.09)	(0.09)	(0.09)	(0.14)
	Turkish Woman	-1.015***	-0.505***	-0.853***	-0.925***
		(0.10)	(0.10)	(0.09)	(0.13)
	Moroccan Woman	-1.105***	-0.783***	-0.367***	-0.588***
		(0.10)	(0.10)	(0.09)	(0.14)
	Antillean Woman	-0.374***	-0.387***	-0.011	-0.054
		(0.10)	(0.11)	(0.10)	(0.15)
Education	Low				
	Low Medium	0.277***	0.304***	0.154**	0.245***
		(0.05)	(0.06)	(0.06)	(0.06)
	Medium	0.549***	0.655***	0.188**	0.399***
		(0.06)	(0.06)	(0.06)	(0.07)
	High	0.774***	1.573***	0.119	0.632***
		(0.07)	(0.07)	(0.08)	(0.11)
# Children	No child				
	1-2 children	0.186***	0.300***	-0.024	0.084
		(0.05)	(0.05)	(0.05)	(0.06)
	3 to 7 children	0.008	0.221**	-0.101	0.078
		(0.06)	(0.07)	(0.07)	(0.07)
Age	Age 15-19			0.526**	0.903***

				(0.18)	(0.25)
	Age 20-35	0.695***	0.286***	0.264***	0.147*
		(0.07)	(0.07)	(0.06)	(0.07)
	Age 36-55	0.186***	0.300***	-0.024	0.084
		(0.05)	(0.05)	(0.05)	(0.06)
	Age 56 and older (ref.)				
Survey year	year=2006				
	year=2011	0.051	0.294***	-0.342***	0.118*
		(0.04)	(0.05)	(0.04)	(0.05)
Household	Up to 500				
income	500-1000			0.083	0.035
				(0.12)	(0.14)
	1000-1500			0.158	0.077
				(0.12)	(0.14)
	1500-2000			0.183	0.112
				(0.12)	(0.15)
	2000-2500			0.199	0.171
				(0.13)	(0.16)
	2500-3000			0.222	0.243
				(0.14)	(0.18)
	3000-3500			0.282	0.061
				(0.15)	(0.20)
	3500more			0.399**	0.067
				(0.15)	(0.21)
	Employed			0.023	0.254***
				(0.05)	(0.06)
	_cons	-1.015***	2.844***	-1.206***	-0.707**
		(0.15)	(0.15)	(0.18)	(0.22)
	R-square	0.196	0.419	0.171	0.276
	BIC	5474.515	14543.115	5641.621	4023.127
	N	4773	4169	4659	4659

Table A4. Dutch Language Proficiency (first equation of estimates of integration outcomes with endogenous language proficiency)

		Employment	**HH income**	**Feel Dutch**	**Integrated**
	Age at migration	-0.125***	-0.129***	-0.111***	-0.117***
		(0.01)	(0.01)	(0.01)	(0.01)
	Age at migrat. squared	0.002***	0.002***	0.002***	0.002***
		(0.00)	(0.00)	(0.00)	(0.00)
	YSM	0.012***	0.017***	0.013***	0.015***
		(0.00)	(0.00)	(0.00)	(0.00)
Homeland	Primary (ref.)				
Education	Unknown	-0.027	-0.096	-0.127	-0.140*
		(0.07)	(0.08)	(0.07)	(0.07)
	Secondary	0.364***	0.383***	0.366***	0.429***
		(0.06)	(0.06)	(0.06)	(0.06)
	High	0.560***	0.585***	0.533***	0.601***
		(0.08)	(0.08)	(0.08)	(0.08)
Origin by	Surinamese Man (ref.)				

gender	Turkish Man	-1.335***	-1.325***	-1.312***	-1.327***
		(0.13)	(0.15)	(0.13)	(0.13)
	Moroccan Man	-0.961***	-0.946***	-0.992***	-0.977***
		(0.13)	(0.15)	(0.13)	(0.13)
	Antillean Man	-0.109	-0.101	-0.126	-0.073
		(0.14)	(0.15)	(0.13)	(0.13)
	Surinamese Woman	-0.138	-0.107	-0.111	-0.093
		(0.13)	(0.15)	(0.12)	(0.12)
	Turkish Woman	-1.674***	-1.662***	-1.608***	-1.627***
		(0.14)	(0.16)	(0.13)	(0.13)
	Moroccan Woman	-1.304***	-1.240***	-1.239***	-1.232***
		(0.14)	(0.15)	(0.13)	(0.13)
	Antillean Woman	-0.149	-0.139	-0.143	-0.151
		(0.13)	(0.15)	(0.13)	(0.13)
Reason for	Work (ref.)				
migration	Study	0.321**	0.343***	0.237*	0.242*
		(0.10)	(0.10)	(0.09)	(0.10)
	Parents	0.440***	0.410***	0.394***	0.399***
		(0.09)	(0.10)	(0.09)	(0.09)
	Marriage	0.065	0.105	0.102	0.066
		(0.09)	(0.09)	(0.09)	(0.09)
	Reunification	0.324***	0.352***	0.268***	0.364***
		(0.08)	(0.08)	(0.08)	(0.08)
	Social advantages	0.375***	0.378***	0.317***	0.375***
		(0.10)	(0.10)	(0.09)	(0.10)
	Other	0.417***	0.337**	0.267**	0.345***
		(0.10)	(0.11)	(0.10)	(0.10)
Number of	No child (ref.)				
Children	1-2 children	0.030	0.059	0.073	0.095
		(0.06)	(0.06)	(0.05)	(0.06)
	3 to 7 children	0.033	0.081	0.117	0.121
		(0.07)	(0.07)	(0.07)	(0.07)
Home country	TV-satellite	-0.217**	-0.222**	-0.237***	-0.247***
orientation		(0.07)	(0.07)	(0.07)	(0.07)
	Homesick	-0.160**	-0.158**	-0.375***	-0.227***
		(0.05)	(0.05)	(0.05)	(0.05)
	Naturalized	0.516***	0.478***	0.454***	0.435***
		(0.07)	(0.07)	(0.07)	(0.07)
Interethnic	same origin partner (ref.)				
Marriage	Single	0.093	0.090	0.090	0.095
		(0.06)	(0.07)	(0.06)	(0.06)
	Dutch partner	0.391***	0.368***	0.328***	0.352***
		(0.08)	(0.09)	(0.08)	(0.08)
	another origin partner	0.114	0.070	0.080	0.050
		(0.13)	(0.14)	(0.14)	(0.14)
Survey year	2006 (ref.)				
	2011	0.315***	0.324***	0.303***	0.331***
		(0.05)	(0.05)	(0.05)	(0.05)
	_cons	1.365***	1.257***	1.369***	1.275***
		(0.21)	(0.23)	(0.19)	(0.20)

	athrho	-0.319***	-0.275*	-0.780***	-0.623***
		(0.09)	(0.11)	(0.09)	(0.09)
	lnsigma		0.314***		
			(0.02)		
	BIC	9747	18320	9745	8156
	N	4773	4169	4659	4659