Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

Language proficiency of migrants

The relation with job satisfaction and skill matching

Hans G. Bloemen

DP 11/2023-052

Empirical Economics (2023) 65:2899–2952
https://doi.org/10.1007/s00181-023-02445-3

Language proficiency of migrants: the relation with job satisfaction and skill matching

Hans G. Bloemen[1,2]

Received: 27 September 2022 / Accepted: 22 May 2023 / Published online: 17 June 2023

Abstract
We empirically analyse the relation between language proficiency and job level of migrants in the Netherlands. A lack of language skills may induce the migrant to work in jobs of a lower level leading to lower job satisfaction. We analyse information about job satisfaction, the fit between the migrant's education and skill level and the job, and professional level. Men with a higher proficiency level are more satisfied with their type of work and are employed at a higher professional level. For women, no impact of language proficiency can be found.

Keywords Discrete regression and qualitative choice models · Economics of immigrants · Occupational choice · Job satisfaction

JEL Classication C33 · J15 · J24 · J28

1 Introduction

Are migrants with a lack of proficiency in the destination country's language more likely to have jobs of a lower level that are less satisfactory? The aim of this study is to analyse this question using data from the Longitudinal Internet Studies for the Social Sciences (LISS) for the Netherlands as a destination country. The literature, of which Chiswick (2007) provides an overview, shows that the language proficiency of migrants has a potential impact on labour market outcomes. Most often analysed is the relation between language proficiency and earnings, recognizing that language skills are part of the individual's human capital.

✉ Hans G. Bloemen
h.g.bloemen@vu.nl

1 Department of Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

2 IZA, Bonn, Netspar, Tinbergen Institute, Amsterdam, The Netherlands

The literature addresses the determinants of the migrant's destination country's language proficiency. Chiswick and Miller (1995) provide a comparison for the USA, Canada, Israel and Australia and find that destination country's fluency increases with exposure, speaking the language of origin decreases with duration of residence, destination fluency rates increase with schooling level and decreases with age at migration, while earnings increase with fluency. (Chiswick and Miller 2001, focus on Canada and find comparable results, while Chiswick, 1998, studies the case of Israel). Dustmann (1994) with an application for Germany finds that language abilities considerably improve the earnings position of migrants. Dustmann and van Soest (2001), in an analysis for Germany, use panel data to econometrically deal with classification errors in language proficiency and show that their methodology improves results. Dustmann and Fabbri (2003) not only analyse earnings as an outcome variable but also employment and find that language proficiency leads to a positive effect on employment, while lack of proficiency leads to earnings losses. Gonzales (2010) does an analysis for Spain and finds effects of host language skills on employment but not on earnings. Chiswick and Miller (2010) address the role of occupational (English) language requirements for US data and find that there is an earnings premium for jobs with higher English requirements, in addition to an effect of migrants' language proficiency itself. They also note that low-skilled occupations have a low required level of English. A detailed analysis of occupational skill requirements is done by Imai et al. (2016) using data for Canada. A unique element in their data is the observation of occupation and skill requirements of migrants before and after migration. They find that migrants after migration more often find themselves in jobs with lower cognitive skill requirements (while cognitive skill requirements are associated with high skilled jobs) than before migration, but also more often in jobs with higher manual skill requirements than before. The migrants' levels of destination language proficiency amplify this.

A recent branch of the literature uses quasi-experimental designs to analyse the impact of language proficiency on labour market outcomes. In France, migrants were offered a language training if their test score was below a certain level, which is exploited by Lochman et al. (2019) to do a regression discontinuity analysis, finding that hours of training increases labour force participation. Foged et al., forthcoming, use a regression discontinuity design for a reform in Denmark that aimed to improve the language skills of refugees after January 1, 1999, by language training classes. Permanent positive effects on earnings are found, as well as an increase in the incidence of working in communication-intensive jobs. Foged and van der Werf (2022) use migrants' proximity to language training centres for the case of Denmark as a source of exogenous variation and find positive effects on subsequent human capital accumulation and local integration. Schmid, forthcoming, exploits the boundaries in Swiss between the German and French language cantons applied to French speaking African refugees and obtains difference-in-differences estimates showing a positive impact of language proficiency on the employment probability. Heller and Slungaard Mumma (2022) study the impact of an English language training in the USA, to which participants were randomly assigned by a lottery, and find an increase in annual earnings for participants of 56%.

If a lack of destination country language proficiency can be completely compensated by lower earnings, we can still have the migrant with the appropriate skills in the right job, albeit with a lower wage. The acquisition of additional destination language skills, possibly supported by the employer, then could help the migrant to grow in his or her current job. We have an inherently different labour market situation if a migrant with a given set of skills ends up in a lower-level job, at the wrong place and maybe with the wrong employer. Although such a situation may provide the migrant with an additional stimulus to acquire destination country language skills (if he or she is aware of the mechanism), a lack of attractive job offers and the existence of labour market frictions can be discouraging. Moreover, in this situation we cannot expect much of a current employer in playing a role in supporting a migrant in acquiring destination language skills. Instead, government policy may be required to obtain a more efficient skill matching of migrants in the labour market.

We contribute to the literature by providing insight in the relation between destination language skills and job match quality. To this purpose, we analyse outcome variables related to job satisfaction and the match between the required skills and the migrant's educational level. These variables are related to the migrant's valuation, or, utility, of the job match. The importance of satisfaction as an outcome has been recognized before in the educational mismatch literature (Chevalier 2003; Mavromaras et al. 2013). A migrant who consciously incorporates ending up in lower-level jobs upon migration as a consequence of a deficiency in the proficiency of the destination country's language, may be satisfied with aspects of the job, reflecting the choices made. But someone involuntarily ending up in a lower-level job due to low destination country language skills can be dissatisfied. As an alternative, more objective, determinant of job match quality, we look at the professional level.

Secondly, the destination language of our analysis is Dutch. As pointed out by Chiswick (2007), it is interesting to analyse destination languages that are less common than the English language. The Dutch language is used among a relatively small part of the world's population so for many migrants it is unlikely that they are familiar with the Dutch language prior to migration to the Netherlands. Moreover, the Netherlands has a rich variation in the migrant population. A recent study by Yao and van Ours (2015) analyses how Dutch language skills of migrants affect their labour market performance.

Thirdly, we address some econometric issues in our model, such as accounting for the fact that in the data we observe only an indicator of an underlying latent language proficiency level, and allowing for simultaneity in labour market outcomes and language proficiency.

Fourthly, we use a combination of econometric techniques, instrumental variables, control variables and robustness checks to enhance identification. Particularly difficult is finding suitable exclusion restrictions (see e.g. Chiswick and Miller 1995; Dustmann and van Soest 2001; Bleakley and Chin 2004): we need variables that plausibly affect labour market outcomes only via their impact on language proficiency. Incorporating information on linguistic distance (Bakker et al. 2009) looks promising, but suffers potentially from country-of-origin confounders. In line with Bleakley and Chin (2004), we include cross-effects of age at migration and linguistic distance as an instrumental variable. We employ several methods to reduce the potential impact of potential confounders of linguistic distance. Exploiting the panel nature of our data, we incorporate unobserved individual specific time-invariant random effects in our econometric model, allowing for unobservable correlation between labour market outcomes and language proficiency (see also Dustmann and van Soest 2001). Outcomes are robust with respect to various sensitivity analyses with confounders of our instrumental variable.

A first analysis of the determinants of language proficiency identifies variables that are in line with determinants discussed in the literature (Chiswick 2007), which gives confidence in our observed measure of language proficiency. In our main analysis of labour market outcomes, we find for men a positive relationship between language proficiency and satisfaction with the type of work and career (which is quite robust across different exclusion restrictions), and we find that for men language proficiency adds to the match between education/skills and the job. Moreover, men are less likely to end up in a low-skilled manual job if they have a higher proficiency. For women, we do not find any robust effects. However, an additional analysis using employment as an outcome variable shows that for women language proficiency may influence selection into employment, whereas for men such a selection effect is absent.

In Sect. 2, we describe the data from the LISS survey. Section 3 presents the econometric model and discusses identification. Section 4 contains the results. Section 5 concludes.

2 Data

Data are drawn from the LISS panel, a panel survey drawn from the population in the Netherlands, consisting of roughly 5000 households (8000 individuals).[1] We use four waves for the years 2008 through 2011.[2] The LISS survey collects information on a great deal of topics, including the household's economic situation (income, assets), work and schooling, religion and ethnicity, and health. Individuals reporting to be born outside the Netherlands are defined as migrants. We exclude individuals born in Belgium as one of the major languages in Belgium is Flemish, similar to Dutch. All Belgian immigrants in the survey have the highest proficiency level according to our survey indicators. In our base sample, we select individuals older than 22 and younger than 65 for which the relevant information is observed. This results in a sample of 1303 individuals-years observations (pooled over the four waves) of 549 different individuals. We use this as our base sample for analysing the determinants of language proficiency. Appendix B provides an analysis of nonresponse and particularly addresses whether nonresponse is related to language proficiency (which could lead to attrition bias), using the panel nature of our data. For the second part of our analysis (the labour market outcomes), we use smaller subsamples, depending on the availability of

1 A detailed description of the sample selection procedure can be found in Scherpenzeel (2009).

2 The panel started in October 2007 and 2008 was the first complete year of data collection. In 2011, LISS introduced the 'Immigrant Panel'. This is a new panel consisting of 'around 2400 individuals, of which 1700 are of non-Dutch origin' (source: LISS. The remaining 700 of Dutch origin serve as a control group). This is not the panel we are using for our current study. Our study uses the regular panel, initiated in 2007, which also contains immigrants. In 2011, these immigrants are still in the regular panel (the 'Immigrant Panel' was newly drawn), but no refreshment was added. The 'Immigrant Panel' provides fewer details about country or language of origin (the emphasis is on the bigger groups of migrants in terms of country of origin) and also does not contain the same question on language proficiency.

the information on the outcome variables, which are typically observed for individuals with a job.

It is good to realize that the time span is limited to 4 years (2008–2011) and the number of different individuals is limited. For this reason, the results mainly identify short-run effects, which can affect their interpretation.

2.1 Base sample for analysing the determinants of language proficiency

Respondents in the LISS are drawn from the municipal registers (Scherpenzeel 2009). The consequence of this for the selection of migrants in the survey is limited since migrants not included in these registers are staying on a so-called 'short stay visa',[3] for a period of at most 3 months. Migrants with the intention to stay in the Netherlands for more than 3 months need to register at their municipality to receive a residence permit (whether temporary or permanent). Scherpenzeel (2009) reports that the sample is biased towards households in which at least one adult is capable of understanding the Dutch language[4] and provides some rough numbers indicating the consequence of this selection: she shows that 3% of the gross sample (i.e. the addresses initially drawn from the municipal registers) is classified as 'non usable' which includes addresses that are dropped due to language problems, in addition to 'among other things, non-existing or non-inhabited addresses, companies, long term infirm or disabled respondents'. This relatively small percentage shows that the impact on selection into the panel was limited.

Information about fluency is obtained by the following survey question:

> 'When having conversations in Dutch, do you ever have trouble speaking the Dutch language?'[5]

1. yes, often have trouble/do not speak Dutch
2. yes, sometimes
3. no, never

The empirical analysis of language proficiency in Sect. 4.1 will shed more light on the quality of the data obtained.

For our base sample, we selected individuals that show no nonresponse to this question, and for which basic characteristics (education level, gender, and the number of years they live in the Netherlands) are observed.[6]

[3] Visum Kort Verblijf.

[4] The questionnaire is computer based and questions appear in Dutch to respondents. However, questionnaires in English are downloadable from the LISS site. It is not known to what extent respondents exploit this opportunity.

[5] A similar question is asked for reading: 'When reading newspapers, letters or brochures, do you ever have trouble understanding the Dutch language?' In the data migrants less often report problems with reading than with speaking, while for the native Dutch in the sample it is the other way around. We therefore consider the fluency indicator as the more reliable indicator of proficiency, as it requires more active skills of the migrant. In all the analyses, we use the fluency indicator. A separate analysis of the determinants of fluency and literacy showed similar results.

[6] Only very few observations are lost by these latter selections. Appendix B provides an analysis of nonresponse.

Table 1 presents descriptive statistics for our sample. The first column shows the sample selected on age (22 < age < 65). The second column shows observations that are more attached to the labour market (we dropped students, retired, disabled, and housewives). The first line shows information about the country or region of birth. The biggest groups of migrants in the Netherlands originate from Turkey, Morocco, the Dutch Antilles, Suriname, and Indonesia, the latter three being (former) Dutch colonies. Individuals from other origins are classified into groups, based on the region of origin, where we group together countries with a similar background. Bigger areas need to be created for countries of origin with smaller numbers of migrants. We distinguish Anglo-American countries (including USA, UK, Australia), with a western cultural and economic background, Germanic and Nordic countries (including German and Scandinavian countries), with cultural, political and economics systems closest to the Dutch, Latin countries (including France, Italy, Spain) with a Southern European culture and economic system. Large regions we consider are Asia, the Middle East, Africa, Eastern Europe, and countries with English as a second language. We thus obtain a set of region-of-origin fixed effects which will also play a role in our identification strategy (see Sect. 3.2).

About 57% of the migrants reports to experience no speaking problems.[7] More detailed descriptives in Table 6 of the Appendix by region of origin show considerable variation by origin in an intuitively appealing ordering. For instance, 77% of migrants from the German and Nordic category report not to experience speaking problems, whereas for migrants from Asia the percentage is 22. The subsample of respondents attached to the labour market shows somewhat better outcomes for the fluency indicator. Note, though, that education levels are also higher for this subsample.

In our analysis, we use a binary indicator for speaking proficiency. This indicator, named ‘speak’, takes the value 1 for those who never have problems in speaking, and is zero otherwise.[8] Thus, we aggregate the two gradations of ‘yes’ when it comes to having troubles with speaking Dutch.

Respondents are asked whether they speak Dutch at home or an other language, and if the latter holds, they are asked to report this other language. Around 70 per cent of the migrants speak Dutch at home, which is a larger percentage than the percentage of migrants who never experience any troubles in speaking Dutch. This suggests that there are people experiencing trouble in speaking Dutch who nevertheless speak Dutch at home.[9]

---

[7] For survey respondents from Dutch/Belgian origin, not part of the sample of migrants, this percentage is 92.1.

[8] Dustmann and van Soest (2001) discuss the issue of measurement error since in their data (based on a 5 points response scale) they observe the phenomenon that migrants may adjust downward the evaluation of their own fluency the longer they are in the country. We checked whether in our data such a pattern appears as well, by estimating the transition probability of going from value ‘speak = 0’ in one period to value ‘speak = 1’ in the next, including the number of years since migration as a regressor. We did a similar regression for the reverse, a transition from ‘speak = 1’ to ‘speak = 0’. We found that the number of years since migration has a positive effect on the first transition (an improvement in proficiency), while it has a negative effect on the second transition (a deterioration in proficiency). Thus, unlike Dustmann and Van Soest (2001), there is no evidence of a downward adjustment in the evaluation of one’s own proficiency.

[9] The data show that high (low) fluency does not necessarily lead to (not) speaking at home. For the entire sample of migrants, we find that 53 per cent reports to have no problems in speaking Dutch and speak Dutch

Table 1 Sample of migrants: sample percentages and means of pooled sample

Variable Observations:	22<age<65 $NT = 1303$	Attached to lab. market $NT = 943$
Origin (language group/area)		
Turkey	12.2	11.0
Morocco	7.6	7.0
Dutch Antilles	9.3	9.2
Suriname	11.2	13.0
Indonesia	5.5	5.5
German and Nordic countries	9.7	9.1
Anglo-American countries	8.0	8.1
Latin countries	10.2	10.4
Countries with English as a second language	3.6	3.8
Asia	4.6	3.9
African	3.5	3.9
Eastern Europe	9.4	9.2
The Middle East	5.1	5.7
Troubles speaking Dutch?		
Yes, often/don't speak Dutch	4.4	3.6
Yes, sometimes	38.6	36.9
No, never	57.0	59.5
Speak (never problems)	57.0	59.5
Speak Dutch at home or other language?		
Dutch at home	68.9	72.1
Dutch dialect	0.5	0.4
Two languages	0.4	0.1
Female	58.0	50.2
Age (mean, std)	43.0	41.8
Household type		
Single	16.4	17.4
Couple without children	23.9	20.6
Couple with children	47.4	49.8
Lone parent	10.9	10.6
Other household type	1.5	1.6
Number of household members (mean, std)	2.9	2.9
	(1.3)	(1.3)
Number of children (mean, std)	1.1	1.1
	(1.1)	(1.0)
Has partner	71.2	70.4
Urbanization		
Extremely urban	26.0	25.7

Table 1 continued

Variable	22<age<65	Attached to lab. market
Observations:	$NT = 1303$	$NT = 943$
Very urban	33.7	34.2
Moderately urban	23.0	24.5
Slightly urban	12.2	10.4
Not urban	5.0	5.2
Occupational status		
In paid employment	57.2	79.8
Works/assists in family business	1.5	2.0
Autonomous professional, freel, self-empl.	4.4	6.0
Job seeker following job loss	5.5	7.6
First time job seeker	1.2	1.9
Exempted from job seeking following job	1.0	
Attends school or is studying	4.5	
Takes care of the housekeeping	11.8	
Is pensioner, [voluntary] early retirem	3.8	
Has (partial) work disability	5.8	
Performs unpaid work while retaining ben.	0.6	0.9
Performs voluntary work	1.3	1.8
Does something else	1.2	
Is too young to have an occupation	0.1	
Education level		
Primary education	12.7	9.5
Lower vocational/professional training	20.3	17.7
Higher sec and middle voc/prof training	32.8	34.9
Higher voc/prof training, university	34.2	37.9

The remaining variables in our sample are the usual demographic control variables. Couples with children are more prevalent among the subsample of migrants attached to the labour market, whereas the reverse holds for singles. Table 1 also shows the occupational status variable on basis of which the subsample of those attached to the labour market was made. Removing those who are taking care of the housekeeping causes a reduction in the share of women. Education levels between countries are difficult to compare. Therefore, we only use a broad categorization of education levels where we distinguish four levels.

Footnote 9 continued
at home; 20 per cent reports both to have problems in speaking Dutch and do not speak Dutch at home; 17.9 per cent reports to speak Dutch at home, even though they experience problems sometimes; 9.2 per cent never experience problems but do not speak Dutch at home.

2.2 Subjective information on educational match and job satisfaction

To shape our thoughts about the relationship between job match quality, job satisfaction, and language proficiency, Appendix C presents a theoretical model providing interpretation and background to the concepts. The main analysis, though, is empirical. The theory assumes that a deficiency in destination country language proficiency potentially leads to downgrading in attainable job level. Absence of an effect of destination country language proficiency may imply that the job level is not affected by it, or the migrant fully acknowledges and incorporates the fact that a lower level of language skills results in a lower-level job in forming job satisfaction. Presence of a negative effect of lack of language proficiency on job satisfaction signals that the actual job level is affected downward, while this is felt either as 'unfair' by the migrant or the migrant fails to perceive that a lower job level is due to a deficiency in language skills and therefore is unexpected and not incorporated in the migrant's decisions.

The survey contains subjective questions to collect information about the match between education, skills, and the job. The first question is about education:
'Please indicate on a scale from 0 to 10 how your highest level of education suits the work that you now perform',
with zero indicating 'does not at all suit my work' and ten indicating 'suits my work perfectly'. A similar question is asked for knowledge and skills:
'Please indicate on a scale from 0 to 10 how your knowledge and skills suit the work you do'.
A final question that we use in our analysis is
'Can you indicate on a scale from 0 to 10 whether your knowledge and skills create any problems in fulfilling your position'
with zero indicating 'very serious problems' and ten indicating 'no problems at all'. All these questions are asked to respondents with a paid job at the moment of the interview.

As far as job satisfaction is concerned, information about the following aspects is collected and used in our analysis:
'How satisfied are you with:

(*a*) *your wages or salary*
(*b*) *the type of work that you do*
(*c*) *your working hours*
(*d*) *your career so far'*

Respondents could answer by indicating a number in the range of zero to ten, ranging from 'not at all satisfied' to 'fully satisfied'. Table 9 in the Appendix shows sample frequencies of the outcomes, also by gender.

Job satisfaction is also considered as an outcome variable in the education mismatch literature, albeit in different ways. Chevalier (2003) uses information about job satisfaction, together with other job features, to construct a measure of mismatch. Mavromaras et al. (2013) use job satisfaction directly as an outcome variable, like we do in our analysis. Their motivation for using job satisfaction as an outcome is that it not only provides information about restrictions faced by the worker, but also incorporates the worker's preferences: a low job level relative to the education level may have been the result of choice, rather than restriction.[10] Applying it to the context of migration, a relatively low job level need not lead to dissatisfaction if working conditions are favourable to the migrant. Accepting a lower job level can also be a conscious consequence of migration. But if someone involuntarily ends up in a lower-level job due to low destination country language skills, this may result into a lower satisfaction with various aspects of the job.

2.3 Language proficiency and professional level

Table 2 shows the various professional levels that are distinguished in the LISS questionnaire. We show sample percentages, both for native Dutch and for migrants, in the age range of 22 to 65, for which also the language proficiency indicator is observed. At the top of the labour market, there is a relatively high representation of migrants in higher academic professions. It is likely that for this group, Dutch language skills are of minor importance for their job characteristics, especially if they work at universities or multinational companies. But the higher educated is a small and specific group, also among the natives. Among higher supervisory professions, the migrants show smaller sample frequencies than the Dutch, and actually for almost all intermediate level professions, ranging from intermediate academic down to skilled and supervisory manual work, we see lower sample frequencies of migrants, compared to the native Dutch. The reverse holds for the lower three categories in the table, consisting of semi-skilled and unskilled manual work and agrarian professions. Adding them together, there is a much larger representation of migrants among these professional levels. This observation is comparable to the findings for Canada by Imai et al. (2016).

2.4 Measures for linguistic distance and genetic distance

In the analysis use will be made of a measure for linguistic distance by Bakker et al. (2009).[11] The linguistic distance is measured using a lexicostatistical approach. A list of 40 stable elements from a list of words commonly used in linguistics[12] is compared between two languages to determine the distance measure. The distance measure is based on the ‘minimum total number of additions, deletions, and substitutions of symbols necessary to transform one word into another’ (Bakker et al. 2009). This number is normalized by dividing it by the maximum necessary changes (thus, it becomes a fraction). Finally, a correction is made for arbitrary coincidences between words of different languages, based on the combinations of words from the 40 words list with different meaning.[13] Holman (2011) provides software and a database to

[10] For instance if lower-level jobs allow for more flexibility or less stress.

[11] Adsera and Pytlikova (2015) use an alternative measure of linguistic distance, based on the language tree.

[12] The Swadish list, see Bakker et al. (2009).

[13] After this final correction, the resulting number is not necessarily a fraction any longer, but it is unlikely to exceed 1 by much. Holman (2011), expresses it as a ‘percentage’ by multiplying it by 100.

Table 2 Sample statistics professional level

	Both genders		Male		Female	
	Native	Migrant	Native	Migrant	Native	Migrant
Number of observations (NT):	13,538	920	6383	418	7155	502
Professional level:						
Higher academic or independent profession	6.8	8.0	9.4	9.3	4.5	7.0
Higher supervisory profession	8.3	5.8	12.8	9.1	4.2	3.0
Intermediate academic or independent profession	25.4	21.4	17.0	14.4	32.9	27.3
Intermediate supervisory or commercial profession	14.0	8.8	18.4	9.8	10.1	8.0
Other mental work	24.9	24.4	15.1	15.6	33.7	31.7
Skilled and supervisory manual work	7.0	6.5	13.1	13.4	1.6	0.8
Semi-skilled manual work	6.8	9.8	9.8	17.0	4.1	3.8
Unskilled and trained manual work	5.2	12.8	2.2	7.7	7.8	17.1
Agrarian profession	1.7	2.5	2.3	3.8	1.2	1.4
Total bottom 3 categories:	13.7	25.1	14.3	28.5	13.1	22.3
Further explanation levels:						
Professional level	e.g.					
Higher academic or independent profession	Architect, physician, scholar, academic instructor, engineer					
Higher supervisory profession	Manager, director, owner of large company, supervisory civil servant					
Intermediate academic or independent profession	Teacher, artist, nurse, social worker, policy assistant					
Intermediate supervisory or commercial profession	Head representative, department manager, shopkeeper					
Other mental work	Administrative assistant, accountant, sales assistant, family carer					
Skilled and supervisory manual work	Car mechanic, foreman, electrician					
Semi-skilled manual work	Driver, factory worker					
Unskilled and trained manual work	Cleaner, packer					
Agrarian profession	Farm worker, independent agriculturalist					

compute the distance measure between any pair of languages.[14] We can assign the linguistic distance to each survey respondent since we observe the country of origin and the language in which they grew up speaking. In an analysis of language proficiency of migrants in Germany with the GSOEP Isphording and Otten (2014) applied the measure by Bakker et al. (2009) for linguistic distance and found it to be a strong predictor for their language proficiency indicator.

Below we will argue that linguistic distance between Dutch and the language of the country of origin is likely to affect Dutch language proficiency, while it is plausible that it will not have a direct effect on labour market outcomes. However, other distance measures based on the country of origin may act as confounders influencing economic outcomes. Thus, it is important to allow for such confounders to prevent invalidating the identification strategy.

Well-known distance measures used in the economic literature are genetic distance and geographic distance. Spolaore and Wacziarg (2009) and Ashraf and Galor (2013) emphasize the importance of differences in genetic distance of people born in different countries as a measure of genetic diversity. They measure its impact on differences in economic development between countries. Spolaore and Wacziarg (2009) made data on genetic distance available and we use this information in the estimation. Chiswick and Miller (2001) include geographic distance as an indicator for language proficiency. We will also include an indicator for geographical distance, based on the shortest distance between the capital cities of the countries.

Muthukrishna et al. (2020) constructed an indicator for cultural difference, based on various of dimensions, including financial and economic dimensions, science and innovation, political values, norms, and beliefs. They present is as a 'theoretically defensible and robust method of measuring cultural distance'. They made available an online tool (www.culturaldistance.com) allowing us to determine the cultural distance between the Netherlands and 98 other countries. The cultural distance measure will be used as a confounder in an additional sensitivity analysis. A fixed effect is included for countries for which the value is not available.

3 Methods

3.1 The econometric model

The properties of the data require a suitable econometric model. There are three important issues that need to be addressed. Firstly, in our data we observe an indicator of language proficiency, but we need to acknowledge that the language proficiency itself is a latent variable. Secondly, unobserved individual specific effects that influence language proficiency may also have an impact on labour market outcomes. Thirdly,

[14] To give an impression of the values (expressed in 'percentages'): for German, we have 50.2, for English 63.22, Sranan Tongo (spoken in Suriname) 74.2, Papiamento (spoken at the Antilles) 90.51, Spanish 91.1, Russian 92.2, Standard Arabic 100, Mandarin 100.3, Turkish 102.33. Thus, we see that for languages far away from Dutch, the distance measures are relatively close together (with Spanish remarkably close to Russian), whereas for languages closer to Dutch, like German and English, the differences in the distance measure are relatively far apart. Thus, the distribution of distance measures will be skewed, as also noted by Isphording and Otten (2014).

we wish to fully exploit the panel nature of our data and control for unobserved time-invariant individual effects. These issues are combined in the following model specification. Fourthly, we need to incorporate the ordered nature of the satisfaction and other indicators.

Define l^*_{it} as a latent variable indicating language proficiency, whereas l_{it} is a binary indicator for it (like the indicator 'speak' in our data). Then, we may define the equation

$$l^*_{it} = z'_{it}\beta + m_i + \epsilon_{it} \text{ with } l_{it} = \iota(l_{it}* > 0) \tag{1}$$

with z_{it} a vector of observable characteristics, uncorrelated with the (zero mean) random variables m_i and ϵ_{it}, of which m_i is an individual specific (zero mean) random effect, with variance $Em_i^2 = \sigma_m^2$, and ϵ_{it} an (zero mean) idiosyncratic error, with variance normalized to one ($E\epsilon_{it}^2 = 1$).

Let r_{it} denote an outcome variable of interest, such as the job suitability or job satisfaction indicators, or professional level, and let r^*_{it} be the underlying latent variable. The outcome variable can be an ordered or a binary variable.

$$r^*_{it} = \alpha l^*_{it} + g'_{it}\gamma + \theta_i + v_{it} \tag{2}$$

with

$$r_{it} = j \text{ if } c_j < r^*_{it} \leq c_{j+1},\ j = 0, ..., K \tag{3}$$

with $c_0 = -\infty$ and $c_{K+1} = +\infty$, with K related to the number of ordered outcomes the observed indicator can take (with $K = 1$ as a special case for a binary variable, and $K = 10$ for the outcomes on job satisfaction that range from zero to 10). In (2) g_{it} is a vector of observable characteristics, uncorrelated with θ_i and v_{it}, which are (zero mean) random variables, with $E\theta_i^2 = \sigma_\theta^2$ and $Ev_{it}^2 = 1$. We allow for $Em_i\theta_i = \sigma_{m\theta} \neq 0$ and $E\epsilon_{it}v_{it} = \sigma_{\epsilon v} \neq 0$ with corresponding correlation coefficients $\rho_{m\theta}$ and $\rho_{\epsilon v}$.

The random variables m_i and θ_i capture time-invariant individual specific variation in the individual's language proficiency and labour market performance, respectively. We allow for a nonzero correlation $\rho_{m\theta}$ between the two, and the value of $\rho_{m\theta}$ will be determined by the data in the estimation of the model.

An example of time-invariant individual specific variation represented by m_i and θ_i is individual capabilities that affect both language proficiency and labour market outcomes. But we may also think of individual effects related to the culture or economic situation in the country of origin, or individual effects related to the exposure to education and destination language during childhood. The correlation $\rho_{m\theta}$ is identified because of the panel nature of our data with same individuals observed in multiple periods.

The identification of the correlation coefficient $\rho_{\epsilon v}$ of the idiosyncratic errors ϵ_{it} and v_{it} is closely related to identifying the causal effect of language proficiency from an effect running through unobservables. This identification relies on instrumental variables and exclusion restrictions (Sect. 3.2). Equations (1) and (2) will be estimated simultaneously for all labour market outcomes. Obtaining estimates of $\rho_{m\theta}$ and $\rho_{\epsilon v}$

allows for testing for the endogeneity of language proficiency ($H_0 : \rho_{m\theta} = \rho_{\epsilon v} = 0$) and the results show the relevance of allowing for endogeneity.

In linear panel data models, the trade-off between random effects (efficiency) and fixed effects (robustness) is made. A fixed effects variant of (2) has the advantage of robustness against correlation between l^*_{it} and θ_i. However, this still does not solve the potential correlation between l^*_{it} and v_{it}. The efficiency loss comes from the possible lack of within group variation of outcome variables and regressors, while the ordered structure[15] of the response variable and the latent nature of the proficiency variable cannot be allowed for in a linear model. Indeed, the within group variation in the observed language proficiency indicator is quite small relative to the between group variation. This is reflected by the large share of the variance σ^2_m of m_i in the total variance $\sigma^2_m + 1$ of $m_i + \epsilon_i$ in the language proficiency Eq. (1). Parameter estimates show values of 85%, implying that fixed effects estimation will remove a major part of the observed variation.

3.2 Identification

In the equation for the labour outcome (2), α is the parameter of interest, since it measures the impact of language proficiency on the labour market outcome. The hardest part is finding a suitable instrument for the identification of $\rho_{\epsilon v}$ to separate the causal effect from correlation in unobservables. Instrumentation and the potential endogeneity issue of language proficiency were addressed by Chiswick and Miller (1995) and Dustmann and van Soest (2001). The former use theoretical exclusions restrictions (family variables affect proficiency but not earnings), while the latter use parental education to instrument proficiency. None of these exclusion restrictions are completely convincing in our application. Bleakley and Chin (2004), in a study of US immigrants, instrument destination language skills (English in their application) by cross-effects between age at migration and country of origin (English speaking versus non-English speaking). They explain that if migration takes place at later age it is more difficult to acquire destination language skills and this relation between age at migration and destination language skills depends on whether the migrant comes from an English or a non-English speaking country or origin.

We base our instrument on the linguistic distance measure introduced in Sect. 2.4. It is appealing to assume that the linguistic distance affects the individual's language proficiency l^*_{it}, without directly affecting labour market outcomes. Nevertheless there are various problems of using the linguistic distance measure that need discussion.

First, if only linguistic distance based on country of origin would be used as an instrument, it would fail to predict outcomes specific to the individual, as all individuals with the same country of origin would have the same predictor. In the spirit of Bleakley and Chin (2004) we include a cross-effect between age and migration and language skills in Eq. (1), while we include age at migration itself in both Eqs. (1) and (2). The

[15] In fact, applying a linear model to an ordered response variable with ten categories implicitly places a cardinal interpretation to the values of the categories, implying, for instance, that the distance between category 1 and 2 (or 1 and 4) is the same as the distance between categories 7 and 8 (or 6 and 9).

underlying assumption is that the effect of age at migration on the outcome variable does not vary by linguistic distance, except through language proficiency itself.

Next, since linguistic distance is based on the country of origin, there are potential confounders related to linguistic distance. Linguistic distance may relate to the choice of destination country (see Adsera and Pytlikova 2015), cultural differences, international trade, tourism, and FDI flows. To reduce the effects of potential confounders, we include the region-of-origin fixed effects in both the language proficiency Eq. (1) and the labour market outcome Eq. (2). The region-of-origin indicators, as described in the data section, relate in a lesser detail to the country of origin than the linguistic distance measure, since, for instance, several countries in Africa are aggregated to one category. To be able to fully capture confounders of the linguistic distance measure requires the assumption that differences between countries in the same category, on the one hand, and the Netherlands, on the other hand, are not relevant (e.g. cultural differences between Tanzania and the Netherlands or between Nigeria and the Netherlands, both countries from Africa, have a comparable effect). On the other hand, for 'big' migrant countries of origin, such as Morocco, Turkey, and Indonesia, there is a unique correspondence between the region-of-origin fixed effect and country.

To further reduce the effects of potential confounders we include in the base specification the measures for genetic distance and geographic distance introduced and discussed in Sect. 2.4, while cultural distance is added in sensitivity analyses. These alternative distance measures can also be crossed with age at migration and included in both Eqs. (1) and (2). All in all, it is hard to see why the region-of-origin specific fixed effects, augmented with the genetic and geographic distance measures, would not be able to capture a large part of the potential confounders.

In addition to the observed confounders, the nonzero correlation in random effects between Eqs. (1) and (2) can capture the impact of unobserved confounders, such as the individual decision to select into migration to a specific destination country.

In summary, to come to identification as close as possible we apply a combination of different variables and methods: (i) we apply different instrumental variables to analyse the robustness of outcomes; (ii) we allow for nonzero correlation $\rho_{m\theta}$ in time-invariant individual specific random effects, which will capture part of unobserved individual specific confounders between language proficiency and labour market outcomes; (iii) we include region-of-origin specific fixed effects and country-of-origin specific distance measures in both Eqs. (1) and (2) to capture the effects of confounders that relate to linguistic distance.

4 Results

In Sect. 4.1, the determinants of fluency are analysed. Section 4.2 shows results on the analysis of the subjective indicators for job satisfaction and job suitability. Section 4.3 presents the analysis of the professional level.

4.1 Determinants of fluency

Chiswick (2007) discusses the relevant determinants of language proficiency in terms of the 3 E's: exposition, education, and economic incentives. Using our data for migrants in the age range older than 22 and younger than 65 (Table 1) we analyse the various determinants of our fluency indicator. The aim of this first analysis is twofold. First, we want to shed light on the performance of the indicator 'speak' as a measure for language proficiency and check whether signs of regressors are as may be expected. Next, we analyse the predictive value of the linguistic distance and its cross-effect with age at migration.

Table 3 presents probit regressions results for fluency (dependent variable is 'speak', introduced in Sect. 2.1). The table includes the average marginal effects of the regressors on the probability that 'speak' equals one, meaning that the migrant has no problems speaking Dutch. Table 7 shows the underlying parameter estimates. All presented standard errors are robust to correlation in unobserved errors across time for the same individual (clustering).

Destination language proficiency is lower for the lower two education levels and decreases with age at migration,[16] both according to expectation. Language proficiency rises with age and migrants speaking Dutch at home do better as well.

In the literature, there is a discussion on whether or not to separate the analysis for men and women, since men and women may have different incentives for learning a language, especially if women are less attached to the labour market. The dummy indicator for female gender is not significant.[17] Table 3 includes regressors for household composition. Notably, the impact of children got attention in the literature: on the one hand, children may stimulate the fluency of parents, as they learn the language quickly at school, while on the other hand, the children may serve as interpreter for their parents, such that the parents themselves exercise the language less actively. Moreover, there may be a differential impact by gender. We included the number of children, as well as indicators for household type (couples without children, couples with children, lone parents, other households, and singles as reference category). The fluency of lone parents seems to be significantly lower than for other household types. Not reported is a regression which includes cross effects of the family indicators with gender. The value of the likelihood ratio test statistic for testing the joint significance

[16] In order to address the question whether age at migration merely approximates the difference between migrants who entered the Netherlands during youth, and therefore were educated in the Dutch schooling system, and migrants who entered during adulthood we did an analysis with a selected subsample of migrants who entered at a later than 12 (and therefore did not attend primary school in the Netherlands) and another analysis with a subsample of migrants who entered at age older than 18 (and thus did not attend secondary school in the Netherlands). We found a similar pattern as for the entire sample (a significant negative effect of age at migration and a small positive squared effect). The impact of age at migration on the pseudo R-squared is still substantial, but smaller, also because the impact of area of origin has a relatively bigger impact for those who entered at adulthood.

[17] Not reported here are regressions where we included cross-effects for female gender and other variables. We included cross-effects of female with the indicators Turkish and Moroccan origin, as these countries are dominantly Islamic, and the position of women may be different in these countries. We did not find any significant effects. Later we report on cross-effects of gender with indicators for household composition.

of the cross-effects with gender is 5.8, indicating that we cannot reject that there are no gender specific household composition effects.

The region-of-origin fixed effect, presented earlier, is included with Asia as a reference group. They measure effects not yet captured by the linguistic, genetic, and geographic distance measures based on country of origin (discussed later). They reveal a higher proficiency for migrants from the (former) Dutch colonies Suriname and the Antilles, and also migrants with German, Nordic, or Latin background do better.

In Sect. 2.4, we discussed the linguistic distance. The survey respondents were asked the question ‘which language or languages did you grow up speaking?’. Some respondents include both Dutch and a foreign language as an answer. We experimented with linguistic distance measures based on Dutch or on the foreign language, and the variant with Dutch outperformed in terms of explanatory power (higher pseudo R-squared) so we use that measure. In Sect. 3.2, we proposed, based on the literature (Bleakley and Chin (2004); Isphording and Sinning (2013)), to use the cross-effect of linguistic distance and age at migration as an exclusion restriction in the equations for labour market outcomes. The first regression in Table 3 only includes the linguistic distance measure itself. It has a positive and significant effect on our measure for language proficiency. As regressors we also included the genetic distance and the geographic distance based on country of origin as potential confounders to linguistic distance (Sect. 2.4). Both distance measures have a negative, but insignificant effect on fluency. Linguistic distance remains a strong predictor for our language proficiency indicator even after controlling for region-of-origin fixed effects, geographic distance, and genetic distance. This is exactly what a good measure of linguistic distance is supposed to do: it predicts language proficiency without suffering from collinearity with other distance measures based on country of origin.

In column 2 of Table 3, we include the cross-effect of linguistic distance and age at migration. It also shows a significant and negative effect, implying that the negative impact of linguistic distance on language proficiency is more important for immigrants who move in at a higher age. Column 3 includes both linguistic distance and its cross-effect with age at migration. Here we see that the coefficient estimate of linguistic distance itself gets imprecise, while the cross-effect remains significant. The joint effect remains significant, as can be seen from the likelihood ration (LR) test statistic at the bottom of the table. (The degrees of freedom for the final column is 2, while 1 for the first two regressions).

Bound et al. (1995), in the context of the linear regression model, pointed at the weak instruments problem: instrumental variables with insufficient predictive power to the endogenous regressor may blow up the variance and small sample bias of the linear Instrumental Variables estimator. In the linear model a widely applied rule of thumb for the instruments to have sufficient predictive power is that the F test statistic for their significance in the first stage regression is not smaller than 10. Now, we do not have a linear model and we apply full information maximum likelihood rather than the linear IV estimator, but it would nevertheless be reassuring if our data and measures would also be applicable in that context. Therefore, we also estimated the regressions in Table 3 for the linear probability model and computed the F-test statistic on basis of the sum of squared residuals of the linear model. It is this F that we report at the

Table 3 Probit regressions speaking fluency for sample of migrants, 22 < age < 65, NT=1303, estimates expressed as average marginal effects

Variable	Coef.	Std.	Coef.	Std.	Coef.	Std.
Linguistic distance	−0.154**	0.041			−0.025	0.073
Ling. Dist. x Age at Migr.			−0.075**	0.018	−0.066**	0.033
Geographic distance	−0.006	0.007	−0.004	0.007	−0.005	0.007
Genetic distance	−0.664	0.542	−0.708	0.528	−0.703	0.532
Prim. Education	−0.136**	0.055	−0.138**	0.056	−0.137**	0.056
Lower voc./prof.	−0.128**	0.043	−0.132**	0.042	−0.132**	0.042
Higher sec./middle voc.	−0.040	0.037	−0.044	0.037	−0.044	0.037
Age at migration	−0.226**	0.042	−0.216**	0.043	−0.214**	0.043
Sqr. Of age at migr.	0.018*	0.009	0.023**	0.009	0.022**	0.010
Age	0.040**	0.016	0.045**	0.016	0.044**	0.016
Female	0.002	0.031	0.004	0.031	0.004	0.031
# Children	0.014	0.023	0.018	0.022	0.017	0.022
couple without children	−0.042	0.048	−0.043	0.048	−0.043	0.048
Couple with children	−0.053	0.060	−0.058	0.059	−0.057	0.059
Lone parent	−0.134*	0.075	−0.143*	0.073	−0.142*	0.074
Other household	−0.171	0.152	−0.176	0.144	−0.175	0.145
Speak Dutch at home	0.135**	0.030	0.132**	0.030	0.131**	0.030
Region-of-origin fixed effects:						
Turkey	0.153*	0.088	0.137	0.087	0.140	0.088
Moroccan	0.085	0.099	0.078	0.099	0.079	0.099
Dutch Antilles	0.385**	0.082	0.365**	0.083	0.366**	0.083
Suriname	0.333**	0.093	0.311**	0.093	0.311**	0.093
Indonesia	0.137	0.094	0.146	0.092	0.144	0.092
German/Nordic	0.247**	0.113	0.245**	0.111	0.244**	0.112
Anglo-American	0.093	0.101	0.094	0.100	0.094	0.100
Latin, western	0.213*	0.110	0.214*	0.111	0.215*	0.110
Latin, nonwestern	0.245**	0.083	0.242**	0.082	0.243**	0.082
English 2nd lang.	0.086	0.089	0.076	0.092	0.077	0.091
Africa	0.252*	0.137	0.270*	0.137	0.269*	0.137
Eastern Europe	0.162	0.103	0.169	0.103	0.170	0.103
The Middle East	0.160	0.110	0.162	0.111	0.164	0.111
Likelihood value	−562.7		−559.1		−559.0	
Pseudo R2	0.368		0.372		0.372	
LR test statistic for (joint)						
Significance ling. dist.	24.1		31.4		31.6	
and Ling. Dist. x Age at Migr.						
F-test lin. prob. model						

Table 3 continued

Variable	Coef.	Std.	Coef.	Std.	Coef.	Std.
Instr. strength ling. dist. and Ling. Dist. x Age at Migr.	32.4		50.0		25.0	

Reference categories dummy variables: Asia; Higher professional/vocational training, university; Single; **/*: significant at 5/10 % level; Standard errors are robust to correlation in unobserved errors across time for the same individual (clustering)

bottom of the table. For all three regressions, the F-statistic meets the criterion.[18] Also in the linear model the cross-effect of linguistic distance and age at migration shows up as the stronger predictor. In the sequel, we continue with the latter variable as our instrumental variable. We provide sensitivity analysis of our results to using linguistic distance as the instrument and to using both. Results turn out to be robust.

Our base specification includes the cross-effect of linguistic distance and age at migration as the instrumental variable, while genetic and geographic distance serve as confounders. In a sensitivity analysis, we include more confounders by adding the cross-effect between age at migration with linguistic distance, with genetic distance, with a dummy for originating from a Western country, and with a dummy for originating from a former Dutch colony (with remaining non-western countries as a reference category). We added these regressors to the analysis to recompute the F-statistic for instrumental strength.[19] The F's are, respectively, 21.2, 27.9, and 14.0, showing that the instruments remain strong enough.

4.2 Results job suitability and satisfaction

Estimation of results for job suitability and job satisfaction is conditional on employment. In Sect. 4.4, an additional analysis to estimate the effect of language proficiency on employment is presented. No effect of fluency of men is found, while the employment rate for women is higher the higher is the fluency of women. This should be kept in mind while interpreting the differences found for men and women. Section 4.4 discusses the differences between men and women and addresses how selectivity into employment may affect the interpretation of the results, especially for women.

Estimates for the parameters of the simultaneous equation model (1) and (2) have been obtained by maximum likelihood estimation for the various outcomes of job suitability and job satisfaction. The parameter of interest for each outcome is the parameter α of language proficiency in (2). To get a feeling for the sensitivity of the estimates, the model was estimated including different combinations of confounders and alternative combinations of instruments. Our base specification (model 1) includes

[18] A sensitivity analysis with a reduced sample of individuals with paid work only (NT = 749) led to respective F-statistic values of 19.1, 30.8, and 15.4, all higher than 10.

[19] Results are not explicitly shown, but Table 13 focusses on the parameters of the cross-effects for the base specification. The cross-effect of age at migration and linguistic distance remains significant. The individual coefficients of the additional cross-effects with age at migration and the confounders do not show up significantly.

the cross-effect of linguistic distance and age at migration as an instrumental variable, and genetic an geographic distance as major confounders. Table 4 shows the coefficient estimate of the parameter of interest α in (2) for various labour market outcomes. The Appendix contains tables with the complete regression results for Eq. (2) for our base specification (model 1), showing the list of regressors included, and the coefficient estimates of the error structure.[20]

The left column of Table 4 shows the results for the base specification, model 1. The coefficient estimate of fluency on satisfaction with work type is positive, suggesting that a higher fluency level leads to job types with a higher level of satisfaction. Separate estimation for men and women shows that this effect is attributed to men. Estimates of fluency on the satisfaction with career are significantly positive for both men and women. A positive coefficient estimate of language proficiency on the fit between education and work, and also on the fit between skills and work, is found for men. Note that this outcome is consistent with the theoretical model in Appendix C. In this model an effect of language proficiency on job satisfaction runs via an effect on the match between the migrant's education or skill level and the level of the job. Thus, for the model to hold, finding an effect on job satisfaction outcomes should go together with finding an effect on the match between education and skill level. Only weak (and not robust to all models) effects for men are found for the response to the statement that there are no problems with knowledge and skills in performing the current job.

As an alternative (model 2), we include as additional confounders cross-effects of age at migration with geographic and genetic distance, and with a dummy for western countries and (former) Dutch colonies, for reasons discussed in Sects. 3.2 and 4.1. Table 4 (model 2) shows the results obtained after adding the additional confounders to the model. The qualitative effects are rather robust to the inclusion of the additional cross effects. Table 14 in the Appendix shows the coefficient estimates of the cross-effects in the labour market outcomes equation. We see positive cross-effects of age at migration and coming from a western country on satisfaction with work type, satisfaction with career and the fit of education and work. A possible interpretation is that migrants from western countries moving to the Netherlands at higher age are often expats who are moving in because of their job. Therefore, they have a higher satisfaction with their type of work and career compared to migrants from western countries with a lower age at migration. The size of this cross-effect is not high enough to have a large impact on the results, and since it has a negative impact on language proficiency (Table 13), the effect of language proficiency is underestimated somewhat by omitting this cross effect.[21] We see negative effects of the cross-effect of age at

[20] Tables 10 through 12 show the estimation results for the full sample and by gender. For reasons of conciseness, we do not show complete results for the variant with additional confounders, but we show the coefficient of interest in Table 4. We also exclude results for the language proficiency equation, since they are not fundamentally different from the results discussed in Sect. 4.1. Like in Sect. 4.1, both genders are included in the estimation of the proficiency equation, also in the estimates in Tables 11 and 12. A higher level of aggregation of categories was used, since we are estimating with fewer observations than in Sect. 4.1: the Middle East and English as a second language were merged to the Asian reference group, and we do not make a distinction between Latin western and non-western countries, on basis of the results of Sect. 4.1. We also aggregated family composition by including a dummy for couples versus the remaining household type, as Sect. 4 showed little impact of family composition. The number of children is maintained.

[21] For some background on the size and the direction of omitted variable bias see, e.g. Dougherty (2016).

Table 4 Parameter estimates for the coefficient α (Eq. 2) of language proficiency on various labour market outcomes, for different estimation methods (migrants, 22<age<65, with paid job)

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Instruments:										
Ling. Dist.				X	X	X	X			
Ling. Dist. × Age at migr		X	X		X		X	X	X	X
Including x-effects										
Confounders and		No	Yes	No	No	Yes	Yes	No	Yes	Yes
Age at migration:										
Dependent variable:	Gender:									
Job suitability:										
Fit education to work	All	0.35**	0.32**	0.30**	0.35**	0.30**	0.32**	0.38**	0.39**	0.29**
	M	0.37**	0.44**	0.43**	0.41**	0.53**	0.51**	0.37**	0.54**	0.42**
	F	0.32*	0.24	0.08	0.33*	0.03	0.10	0.14	0.16	0.15
Fit skills to work	All	0.27**	0.21*	0.19*	0.27**	0.15	0.21	0.28**	0.23*	0.18*
	M	0.32**	0.41**	0.28**	0.32**	0.35**	0.42**	0.33**	0.38**	0.40**
	F	0.08	−0.02	−0.04	0.05	−0.13	−0.08	0.02	−0.07	−0.04
No problems with knowledge	All	−0.12	−0.20*	−0.17*	−0.15*	−0.22**	−0.21**	−0.17*	−0.25**	−0.17*
and skills in performance job	M	−0.19	−0.33*	−0.24*	−0.22*	−0.33**	−0.33**	−0.21*	−0.39**	−0.29**
	F	−0.05	−0.08	−0.13	−0.09	−0.17	−0.14	−0.13	−0.08	−0.05

Table 4 continued

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Job satisfaction:										
Satisfaction with wage	All	0.13	0.13	0.10	0.12	0.09	0.12	0.06	0.08	0.11
	M	0.29	0.29	0.04	0.29	0.27	0.28	0.17	0.29	0.22
	F	0.17	0.19	−0.07	0.17	0.16	0.18	0.09	0.22	0.15
Satisfaction with work type	All	0.32**	0.31**	0.30**	0.32**	0.32**	0.31**	0.22**	0.32**	0.27**
	M	0.54**	0.63**	0.46**	0.54**	0.55**	0.64**	0.45**	0.62**	0.60**
	F	0.23	0.14	0.19	0.23	0.14	0.13	0.13	0.16	0.12
Satisfaction w. working hours	All	0.15	0.17	0.09	0.14	0.11	0.16	0.14	0.17	0.15
	M	0.25*	0.32**	0.22	0.24*	0.27	0.32*	0.23*	0.28	0.32**
	F	0.19	0.13	0.04	0.18	0.01	0.09	0.18	0.16	0.14
Satisfaction with career	All	0.38**	0.47**	0.35**	0.39**	0.41**	0.47**	0.43**	0.40*	0.41**
	M	0.51**	0.68**	0.44**	0.51**	0.59**	0.68**	0.52**	0.72**	0.64**
	F	0.36**	0.37*	0.31	0.35**	0.33	0.36	0.42**	0.37*	0.37*

*/**Significance at 5/10% level. Selection of complete regression equations in the Appendix, Tables 10 through 12 for base variant, model 1
Instruments: Ling. Dist. = Linguistic Distance; Ling. Dist. × Age at migr = Linguistic Distance × Age at migration
x-effects confounders and age at migration, with confounders: geographic distance; genetic distance; cultural distance (model 8); Western country of origin; Dutch colony country of origin
Variables included in g_{it} **base variant** (model 1): region fixed effects, genetic distance, geographic distance, education dummies age at migration, age at migration squared, age, (female), # children, couple
• Models 2, 5, 6, 8, and 9 add x-effects confounders and age at migration
• Models 7 and 8 add cultural distance. Model 9 adds 'Hard Being Accepted with Foreign Culture'
Variables included in z_{it}: same as g_{it} + instruments + speak Dutch at home

migration and coming from a Dutch colony on satisfaction with work type and career: migrants who move in at older age from a Dutch colony do worse than those who move in at younger age. The impact of the cross-effect of age at migration and coming from a Dutch colony on language proficiency (Sect. 4.1) was positive but very low, however, so no big effect of omitted variable bias is expected.

Models 3 includes linguistic distance as an instrumental variable, instead of the cross-effects of linguistic distance and age at migration, while model 4 includes both linguistic distance and the cross-effect of linguistic distance and age at migration. The qualitative outcomes are robust to these choices. Model 5 and model 6 extend models 3 and 4, respectively, by adding as additional confounders the cross-effects with age at migration (see also model 2). The results are again robust.

The measure for cultural distance by Muthukrishna et al. (2020) has been added to base model 1 in Table 4, resulting in model 7. We first checked the impact of adding cultural distance on the strength of the base model's instrument for fluency, the cross effect of linguistic distance and age at migration, by adding it to the fluency equations in Table 3. The results in the Appendix, Table 8, show no significant effect of cultural distance, while the instrument's strength remains unaffected. Muthukrishna et al. (2020) discuss that cultural distance for some countries can have opposing effects to other distance measures. An example is South Korea, with a relatively large linguistic distance, but a moderate cultural distance (for instance, the educational environment is similar to the Netherlands). But for the same reason, cultural distance can have an impact on labour market outcomes, as it refines, for instance, the difference between countries that have a similar linguistic distance but a different cultural distance. Table 4 shows that the effect of fluency on labour market outcomes is robust to the inclusion of cultural distance. However, we do find that cultural distance have separate effects on several labour market outcomes. The satisfaction with work type and with career decrease with cultural distance, as does the degree to which knowledge and skills suit the present job. As before, the results hold for men. Thus, we do find effects of cultural distance on labour market outcomes, but they do not replace the effects of fluency. Model 8 in Table 4 adds cultural distance to model 2, thereby extending the set of regressors in model 7 with the cross-effects of the various confounders (including cultural distance) with age at migration. Results are robust.

An alternative indicator for cultural differences is obtained from survey information. The LISS survey includes the statement 'It is difficult for a foreigner to be accepted in the Netherlands while retaining his/her own culture', to which the respondent can agree or disagree. We construct an indicator for 'disagree' and add it to model 2 to obtain model 9. The results are not affected much by its inclusion.

Some additional sensitivity checks have been carried out. Results with linguistic distance as an instrument (model 3) were also robust to a more flexible specification with linguistic distance, its square, and a separate dummy variable for whether Dutch was among the language(s) someone grew up with. Results were also robust to more flexible specifications in age at migration: we added dummy variables for age at migration below six (meaning that the migrant followed primary and subsequent education in the Netherlands), age at migration below 12 (meaning that secondary and subsequent education was followed in the Netherlands), and age at migration below 18. Apparently the quadratic in age at migration was flexible enough.

In conclusion, we can say that the simultaneous models notably find effects of fluency for men, especially for satisfaction with work type, satisfaction with career, and the fit between ability and work. A positive effect of fluency for men is also found for the fit between education and work, as long as linguistic distance is included as an instrument. This outcome has important implications. It suggests that a good job match in terms of type of work, career opportunities, and skills is more important to the (male) migrant than the wage, as satisfaction with the wage was not affected. Migrants feel restricted in terms of their job opportunities in this respect. Increasing job match quality can be beneficial both for the individual migrant, reducing his dissatisfaction, and for society, as it can lead to getting the right people at the right place. Apart from policy aimed at increasing destination country language skills at the migrant level, one may think of facilitating the presence of workers with lesser fluency at the workplace, for instance by adapting the package of tasks or by allowing for world languages, such as English, at the job. Although theoretically wage increases may compensate for work-type dissatisfaction, the question is whether these touch the core of the problem.

4.3 Results language proficiency and professional level

The analysis so far considered the direct effect of destination language proficiency on subjective outcomes of job satisfaction and indicators for job suitability, and we notably found a robust effect of fluency on the satisfaction with the type of work. The type of work may be related to the professional level of the job, which is a more objective measure of job type. We do an analysis in two steps: we first check how the subjective satisfaction and suitability indicators are related to professional level by including professional levels in ordered probit regressions for the satisfaction and suitability indicators. Next, we analyse the impact of language proficiency on the professional level. Since basically anybody is able to perform semi- or unskilled manual work, irrespective of the education, we narrow down the analysis to the question whether migrants with a lower language proficiency level are more likely to end up in a manual job. In the analysis we will again allow for unobservable correlation between proficiency and the probability to end up in such a job.

In the first step, the dummy variables for the professional levels (introduced in Sect. 2.3) were included in an ordered probit analysis of job satisfaction and job suitability, taking the semi-skilled, unskilled manual work, and agrarian professions as one reference category. Tables 15 through 17 in the Appendix show the estimation results (for both genders pooled, and men and women separately). For satisfaction with career and satisfaction with work time, both for men and for women, most professional levels lead to a higher satisfaction than the manual reference category. We also find a better fit of education and the job, and of knowledge and skills and the job if the professional level is higher than manual. For men, we do not find much effect of professional level on satisfaction with wage, except that migrants with a higher academic profession are more satisfied with their wage than migrants with manual professions. For women, we find a somewhat stronger relation between professional level and satisfaction with wage. For men, we find no relation between the professional level and satisfaction with work time, whereas women with an intermediate professional level seem to be more satisfied than manual workers. For men we do not find that migrants with a higher professional level have more or less problems in performing their job than migrants with a manual profession, whereas women with a higher academic profession seem to experience more problems in performing their job than women in manual professions. Over all, the impression is that if there is any relation between job satisfaction and professional level, migrants in manual jobs are less satisfied.

The second step is to analyse whether there is a relationship between language proficiency and having a manual profession. Table 5 shows the estimation results.[22] The table shows univariate regressions, where the indicator 'speak' is simply plugged in among the right hand side variables, both ignoring that it is an indicator of an underlying latent proficiency level and ignoring possible correlation in unobservables. Next to that the results of the simultaneous equations estimates presented.[23] Table 5 presents results with the cross-effect of linguistic distance and age at migration as an instrument. The univariate model shows a negative parameter estimate of fluency in the equation for the probability of having a manual job. Estimation by gender shows that this effect is attributed to men. In the simultaneous estimation, the parameter estimate becomes less precise, but is still significant at the 10% level. Moreover, we find that the correlation coefficient between the equations for language proficiency and the probability of ending up in a manual job, ρ is not significantly different from zero. Therefore, we also estimated the model with ρ restricted to zero[24] and found that the parameter of fluency became more precise (significantly different from zero at the 5% level) whereas the likelihood ratio test statistic for testing the hypothesis $\rho = 0$ took the value 0.86 (in the estimation for men), such that the null hypothesis is not rejected.

For women, the simultaneous equation model shows a 10% significant effect, but as opposed to men, this effect disappears when alternative instruments are used. Thus, for women we do not find a robust effect.

In a sensitivity analysis, we included again the cross-effects of age at migration with the aforementioned confounders. For men, this results in a coefficient estimate of α of -0.70, significant at the 5% level, indicating a slightly stronger effect than our base specification. The individual coefficients of the cross-effects of age at migration are not estimated precisely. For women, the coefficient estimate of α becomes -0.36 and is not significantly different from zero, in line with earlier results. The picture provided by the sensitivity analysis compares to the earlier results: the quantitative impact on the parameter of interest is of second order nature, and if there is any effect, it points in the direction of a slightly stronger effect for men.

In conclusion, we may say that at the least we find a negative correlation between the probability of ending up in a manual job for men. The result is consistent with Imai et al. (2016).

22 For the simultaneous equations, we again suppress results for the language proficiency equation.

23 We do not present random effects estimates, exploiting the panel nature. The wave to wave within individual variation in manual work turned out to be so small that it is not possible to identify random effects. The variance of the random effect in the manual work equation, which also measures the within individual correlation across time, grew very large during the maximization procedure.

24 Note that with ρ equal to zero we still allow for the latent nature of language proficiency.

Table 5 Dependent variable: Manual Work, (migrants, 22 < age < 65, with paid job), parameter estimates of coefficients

	Both genders pooled				Male				Female			
	Univariate		Simultaneous		Univariate		Simultaneous		Univariate		Simultaneous	
Variable	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Speak	−0.43**	0.17	−0.46**	0.21	−0.66**	0.26	−0.50*	0.27	−0.25	0.22	−0.50*	0.28
Turkey	0.50	0.36	0.47	0.39	0.81*	0.47	0.59	0.60	0.48	0.51	0.64	0.49
Moroccan	−0.45	0.41	−0.50	0.41	−0.24	0.56	−0.46	0.64	−0.59	0.56	−0.30	0.58
Dutch Antilles	−0.42	0.38	0.11	0.56	0.36	0.59	0.84	0.71	−0.64	0.49	0.13	0.74
Suriname	−0.73**	0.36	−0.07	0.57	−0.38	0.47	−0.02	0.57	−1.06*	0.56	0.32	1.14
Indonesia	0.49	0.44	0.75	0.48	1.29*	0.66	1.63**	0.67	−0.08	0.57	0.29	0.59
German/Nordic	−0.59	0.45	−0.26	0.52	−1.04	0.65	−0.97	0.70	−0.20	0.59	0.59	0.80
Anglo-American	−1.20**	0.41	−1.06**	0.45	−1.60**	0.67	−1.65**	0.75	−0.56	0.63	−0.08	0.72
Latin	−0.34	0.37	−0.06	0.42	−0.52	0.51	−0.36	0.57	0.05	0.43	0.66	0.60
Africa	1.62**	0.47	1.91**	0.61	2.83**	0.91	3.23**	1.08	1.26**	0.60	1.39*	0.81
Eastern Europe	−0.64	0.44	−0.53	0.46	−0.77	0.71	−0.90	0.76	−0.21	0.50	0.21	0.58
Geographic distance	0.04	0.04	0.03	0.04	0.03	0.05	0.02	0.06	0.05	0.06	0.02	0.06
Genetic distance	−7.43**	2.40	−7.95**	2.76	−10.49**	2.75	−11.94**	3.27	−4.81	4.35	−3.31	5.06
Prim. Education	1.97**	0.28	1.68**	0.42	1.82**	0.44	1.44**	0.65	2.53**	0.44	2.05**	0.66

Table 5 continued

Variable	Both genders pooled				Male				Female			
	Univariate		Simultaneous		Univariate		Simultaneous		Univariate		Simultaneous	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Lower voc./prof.	1.84**	0.25	1.58**	0.37	1.69**	0.38	1.46**	0.46	2.16**	0.38	1.64**	0.65
Higher sec./middle voc.	0.82**	0.22	0.75**	0.25	0.59*	0.33	0.52	0.36	1.16**	0.32	0.98**	0.39
Age at migration/10	0.31	0.24	−0.08	0.37	0.37	0.37	0.06	0.48	0.40	0.33	−0.14	0.52
Age at migration/10 squared	−0.04	0.06	−0.01	0.06	−0.07	0.09	−0.07	0.11	−0.05	0.07	0.00	0.08
Age	−0.16*	0.09	−0.12	0.11	0.06	0.14	0.05	0.15	−0.34**	0.13	−0.20	0.18
Female	−0.06	0.16	−0.08	0.16								
# Children	0.51	0.82	0.37	0.86	2.35*	1.28	1.96	1.35	−0.50	1.22	−0.44	1.20
Couple	−0.21	0.18	−0.10	0.19	−0.48*	0.27	−0.34	0.30	−0.11	0.24	0.06	0.26
Intercept	−0.58	0.61	−0.37	0.70	−1.25	0.87	−0.72	1.19	−0.66	0.71	−0.78	0.69
ρ (corr. coef. error terms)			0.28	0.25			0.23	0.31			0.46	0.30

Reference categories dummy variables: Asia; Higher professional/vocational training, university; Single;
In the univariate regressions ‘speak’ is included as a binary indicator; in the simultaneous estimation, the latent specification is used, as in Eq. (2)
For the simultaneous estimation, results for the language proficiency equation are omitted from the table
**/*Significant at 5/10 % level;
Standard errors are robust to correlation in unobserved errors across time for the same individual (clustering)

4.4 Reason for different outcomes by gender?

The previous analyses showed that, both for subjective and objective measures of job level outcomes, we mainly found an impact of language proficiency for men, but not for women. The results of the outcome equations are all conditional on employment. This may potentially affect the interpretation of results. Since in general, labour market participation rates for men are higher than for women, we did an additional analysis to check whether there is a difference between men and women as far as the impact of fluency on selection into employment is concerned. We did an analysis with employment as the outcome variable, both for a full sample (i.e. measuring employment versus non-employment) and a sample of individuals attached to the labour market (i.e. measuring employment versus unemployed participants). For men we did not find an impact of language proficiency on employment for any sample.[25] For women, we found a positive impact on employment for both subsamples if we estimate an employment equation simultaneously with an equation for proficiency. Thus, it seems that for women language proficiency plays a more pronounced role in selection into employment, so once selected into employment proficiency does not have an additional impact on job level outcomes. Men seek to enter employment, irrespective of their language proficiency, and within employment outcomes for job level seem to move together with proficiency. For women, the results of not finding an effect of fluency on job satisfaction should be interpreted with care. If women with a potentially high degree of job dissatisfaction remain out of employment, the selection of employed women may expose a lower degree of variation in job satisfaction such that no effect for women will be found. We can compare, though, the variation in job satisfaction between men and women and check whether the variation among women is lower than among men. There is no evidence for the latter. As an example we consider the outcomes for satisfaction with work type. For men/women, the mean score is 7.2/7.1, the median is 8/8, while the 10th and 90th per cent quantiles are 5/5 and 9/9. Thus, no large difference in the variation of satisfaction with work type across men and women is found.

5 Conclusions

We find that a higher Dutch language proficiency leads to a higher degree of satisfaction with the type of work and career, notably for male workers. Male workers also report a better fit between their educational attainment and the work they do. These results are robust to the inclusion of various combinations of confounders. An analysis of the impact of professional level on job satisfaction as a possible explanation for the underlying mechanism shows a lower satisfaction with various job aspects for migrants in jobs that require lower skill levels. Male workers with a lower Dutch proficiency have a higher probability of ending up in a manual job. For women, we do not find (robust) effects. However, it seems that women are affected by language proficiency through the selection into employment.

[25] For reasons of conciseness we do not show Tables with results for this sensitivity analysis.

Finding an impact of language proficiency on job satisfaction is interesting, since satisfaction results from the interaction of preferences and restrictions. A migrant who, prior to migration, is fully aware what lack of destination language proficiency means for the type of work and incorporates this in the migration decision need not necessarily be less satisfied as a result of a lower language proficiency. Finding an effect of language proficiency on job satisfaction is indicative of restrictions felt by the migrant.

The outcome of the analysis has important implications. It suggests that a good job match in terms of type of work, career opportunities, and skills is more important to the (male) migrant than the wage, as satisfaction with the wage was not affected. Migrants feel restricted in terms of their job opportunities. Increasing job match quality can be beneficial both for the individual migrant, reducing his dissatisfaction, and for society, as it can lead to getting the right people at the right position. Apart from policy aimed at increasing destination country language skills at the migrant level, one may think of facilitating the presence of workers with lesser fluency at the workplace, for instance by adapting the package of tasks or by allowing for world languages, such as English, at the work floor. Although theoretically wage increases may compensate for work type dissatisfaction, the question is whether these touch the core of the problem. We should be aware, though, that the suggested implications may only hold for the short run, due to the limited time span of the data of four years.

The contribution of these findings largely depends on the extent to which we have been able to isolate the causal effect of language proficiency on the labour market outcomes. This in turn depends on the validity of the exclusion restrictions and instrumental variables, but also on the methods employed, while sensitivity analyses shed light on the robustness of the results. Let us recapitulate. In the main analysis, a cross-effect of linguistic distance and age at migration are used as an exclusion restriction (although conclusions do not change if we only use linguistic distance). By itself, it seems appealing that the (cross) effect of linguistic distance on labour market outcomes runs through language proficiency, and not directly via a separate channel. But since linguistic distance is largely based on the country of origin, confounders that are also based on country of origin may invalidate the procedure (see discussion Sect. 3.2). Whenever there are potential confounders, it is important to try and control for them. We include region-of-origin fixed effects which have a higher degree of aggregation than country of origin (see Sect. 3.2), except for the 'bigger' countries of origin. We include country-of-origin based genetic and geographic distance measures which in the literature have been used as controls for economic outcomes of countries. The linguistic distance measure turns out to be a very strong predictor of fluency, while genetic, geographic and cultural distance do not add to the explanation of fluency, once region-of-origin fixed effects and linguistic distance have been controlled for (Sect. 4.1). Moreover, adding genetic and geographic distance does not reveal traces of multicollinearity with the linguistic distance measure, which we would typically expect if confounders play a major role, since all distance measures would capture at least part of country-of-origin specific confounders. The linguistic distance measure is constructed in an intricate way using characteristics of the language (Sect. 2.4) which explains its power in predicting fluency. We use an econometric framework that allows for individual specific time-invariant correlation between unobservables in fluency and the labour market outcomes (as specified in Sect. 3.1 and motivated in Sect. 3.2). By including cross-effects of age at migration and confounders of linguistic distance, we check the sensitivity of our results. In particular, we put our exclusion restriction to a test and see whether our results are 'explained away' by including these additional cross-effects. The sensitivity analysis shows that results are robust. In conclusion, we may say that much care has been spent on controlling for the possible impact of confounders and correlation in unobservables, and with more than a reasonable degree of plausibility results can be considered to reveal the prevalence of causal effects of language proficiency of job level.

Declarations

Conflict of interest The author declares to have no conflict of interest concerning this study. This paper draws on data of the LISS panel of CentERdata. This paper benefited from comments by two anonymous referees, Federica Teppa (seminar and conference) participants at UAB Barcelona, the MESS Workshop, ESEM, Université Paris I, Oded Galor and other participants at the IZA Migration Meeting, University of Basque Country in Bilbao, Alexander Humboldt Seminar Mannheim.

Appendix A: Fluency by origin

Table 6 shows sample statistics by region of origin for different sample selections for the fluency indicator. The sample in the left panel was selected on the basis of age. The first two lines compare the native Dutch/Belgium with the migrants. Among the native Dutch/Belgium, the percentage of respondents that report no problems in speaking Dutch (variable 'speak') is 92. For the migrants, this percentage is much lower, 57. There is considerable heterogeneity depending on the region of origin: people from former Dutch colonies report relatively often to have no problems with speaking. For Suriname, Indonesia, and the Dutch Antilles the percentages are 85, 78, and 68. Among migrants in the German/Nordic group, 78 per cent reports not to experience any speaking problems. For respondents from Anglo-American and Latin countries, the speaking performance is still a little above the average for migrants (59 per cent without any speaking problems for both groups). People from Asia experience speaking problems most often: only 22 per cent reports no problems. Below average is also the speaking performance of the Turkish, Moroccan, people from countries in which English is a second language, Eastern Europe, and Africa, with respective percentages of 44, 42, 41, 44, and 50 for the absence of speaking problems. People from the Middle East report in only 31 per cent of the cases to have no speaking problems.

Table 6 Descriptives of fluency by origin

Origin	Sample selection 22<age<65		'Attached to' labour market"	
	Speak (%)	Age (mean)	Speak (%)	Age (mean P)
Dutch/Belgium	92	48	93	45
Foreign	57	45	59	44
Turkish	46	39	42	38
Moroccan	39	42	50	39
Dutch Antilles	66	45	63	45
Suriname	86	49	89	48
Indonesia	78	53	75	50
German/Nordic	77	46	76	44
Anglo-American	59	46	63	45
Latin language	61	46	63	43
English 2nd lang	43	44	47	40
Asia	22	43	24	45
Africa	48	40	43	42
Eastern Europe	43	43	51	42
The Middle East	33	48	35	48

Table 7 Probit regressions speaking fluency for sample of migrants, $22 < \text{age} < 65$, NT=1303, parameter estimates of coefficients

Variable	Coef.	Std.	Coef.	Std.	Coef.	Std.
Linguistic distance	−0.63**	0.18			−0.10	0.30
Ling. Dist. x Age at Migr.			−0.31**	0.08	−0.27**	0.14
Geographic distance	−0.02	0.03	−0.02	0.03	−0.02	0.03
Genetic distance	−2.73	2.23	−2.93	2.20	−2.91	2.21
Prim. Education	−0.56**	0.23	−0.57**	0.24	−0.57**	0.24
Lower voc./prof.	−0.53**	0.18	−0.55**	0.18	−0.54**	0.18
Higher sec./middle voc.	−0.16	0.15	−0.18	0.15	−0.18	0.15
Age at migration	−0.93**	0.18	−0.89**	0.18	−0.89**	0.18
Sqr. Of age at migr.	0.07*	0.04	0.09**	0.04	0.09**	0.04
Age	0.16**	0.07	0.19**	0.07	0.18**	0.07
Female	0.01	0.13	0.02	0.13	0.02	0.13
# Children	0.06	0.09	0.07	0.09	0.07	0.09
couple without children	−0.17	0.20	−0.18	0.20	−0.18	0.20
Couple with children	−0.22	0.24	−0.24	0.24	−0.24	0.24
Lone parent	−0.55*	0.31	−0.59**	0.30	−0.59*	0.30
Other household	−0.70	0.62	−0.73	0.59	−0.73	0.60
Speak Dutch at home	0.55**	0.13	0.55**	0.13	0.54**	0.13
Region-of-origin fixed effects:						

Table 7 continued

Variable	Coef.	Std.	Coef.	Std.	Coef.	Std.
Turkey	0.63*	0.36	0.57	0.36	0.58	0.36
Moroccan	0.35	0.40	0.32	0.41	0.33	0.41
Dutch Antilles	1.58**	0.34	1.51**	0.34	1.51**	0.34
Suriname	1.37**	0.38	1.29**	0.39	1.29**	0.39
Indonesia	0.56	0.38	0.60	0.38	0.59	0.38
German/Nordic	1.01**	0.46	1.02**	0.46	1.01**	0.46
Anglo-American	0.38	0.41	0.39	0.41	0.39	0.41
Latin, western	0.88*	0.45	0.88*	0.46	0.89*	0.46
Latin, nonwestern	1.01**	0.34	1.00**	0.34	1.01**	0.34
English 2nd lang.	0.35	0.37	0.31	0.38	0.32	0.38
Africa	1.03*	0.57	1.12*	0.58	1.11*	0.58
Eastern Europe	0.66	0.42	0.70	0.43	0.70	0.43
The Middle East	0.66	0.45	0.67	0.46	0.68	0.46
Intercept	0.73	0.56	0.50	0.55	0.53	0.56

Reference categories dummy variables: Asia; Higher professional/vocational training, university; Single; **/*Significant at 5/10 % level;
Standard errors are robust to correlation in unobserved errors across time for the same individual (clustering)

Appendix B: Analysis of nonresponse

Survey data can be subject to selective response. While cross-section data do not offer the opportunity to analyse nonresponse, panel data allow us to exploit the information contained in observing a household responding in one wave, but not in another. Our base sample consists of $NT = 1303$ observations on $N = 549$ different individuals. The average number of waves in which an individual is observed is 2.4, while we use data on 4 waves. Observing fewer waves can be attributed to attrition, wave nonresponse, or individuals were drawn into the survey in a later year.

Nonresponse or attrition in panel data is not a problem if it is nonrandom with respect to the outcome of interest and can be explained on basis of observable regressors (Hausman and Wise 1979). In the present application, a particular concern is that nonresponse may be related to fluency. This section is to provide some background information on the nature of nonresponse in our sample.

The data collection in the LISS survey consists of a monthly household box and yearly questionnaires. The household box contains basic information, such as age, education level, and household composition, and is updated on a monthly basis. This information is mostly available for respondents once they have been selected in the LISS survey and have not left it. The information about ethnic background (including country of origin) and language proficiency is obtained from a questionnaire that is sent to the survey respondents in January/February of each year. If survey respondents do not fill out this questionnaire in a given year, information on language proficiency is missing for that year. For respondents who never filled out this questionnaire, we therefore not only miss information on language proficiency, but also on their country

Table 8 Probit regressions speaking fluency for sample of migrants, 22 < age < 65, NT=1303, estimates expressed as average marginal effects. Include culture distance

Variable	Coef.	Std.	Coef.	Std.	Coef.	Std.
Linguistic distance	−0.153**	0.041			−0.023	0.073
Ling. Dist. x Age at Migr.			−0.075**	0.018	−0.066**	0.033
Geographic distance	−0.006	0.007	−0.004	0.007	−0.004	0.007
Genetic distance	−0.688	0.537	−0.741	0.528	−0.735	0.529
Cultural distance	0.033	0.478	0.034	0.477	0.030	0.475
Prim. Education	−0.136**	0.055	−0.137**	0.056	−0.137**	0.056
Lower voc./prof.	−0.128**	0.043	−0.131**	0.042	−0.131**	0.042
Higher sec./middle voc.	−0.039	0.037	−0.043	0.037	−0.043	0.037
Age at migration	−0.226**	0.042	−0.216**	0.043	−0.214**	0.043
Sqr. of age at migr.	0.018*	0.009	0.023**	0.010	0.022**	0.010
Age	0.040**	0.016	0.045**	0.016	0.044**	0.015
Female	0.001	0.031	0.003	0.031	0.004	0.031
# Children	0.014	0.023	0.017	0.022	0.017	0.022
couple without children	−0.042	0.047	−0.043	0.048	−0.043	0.048
Couple with children	−0.052	0.060	−0.057	0.059	−0.056	0.059
Lone parent	−0.133*	0.076	−0.142*	0.075	−0.141*	0.075
Other household	−0.171	0.153	−0.176	0.145	−0.175	0.145
Speak Dutch at home	0.134**	0.030	0.131**	0.030	0.131**	0.030
Region-of-origin fixed effects:						
Turkey	0.149	0.096	0.132	0.095	0.136	0.096
Moroccan	0.079	0.116	0.071	0.117	0.073	0.117
Dutch Antilles	0.396**	0.099	0.382**	0.100	0.381**	0.100
Suriname	0.342**	0.108	0.325**	0.109	0.324**	0.109
Indonesia	0.132	0.113	0.141	0.112	0.139	0.112
German/Nordic	0.250**	0.125	0.248*	0.123	0.246**	0.123
Anglo-American	0.094	0.110	0.094	0.109	0.094	0.109
Latin, western	0.214*	0.111	0.214*	0.111	0.215*	0.111
Latin, nonwestern	0.245**	0.083	0.243**	0.082	0.244**	0.083
English 2nd lang.	0.087	0.095	0.078	0.099	0.078	0.098
Africa	0.258*	0.154	0.280*	0.156	0.279*	0.156
Eastern Europe	0.164	0.104	0.174	0.104	0.174	0.104
The Middle east	0.159	0.117	0.163	0.119	0.164	0.119
No value	−0.006	0.103	−0.012	0.104	−0.011	0.103
Likelihood value	−562.7		−559.0		−558.9	
Pseudo-R2	0.368		0.372		0.372	
LR test statistic for (joint) significance ling. dist. and Ling. Dist. x Age at Migr.	23.7		31.1		31.3	

Table 8 continued

Variable	Coef.	Std.	Coef.	Std.	Coef.	Std.
F-test lin. prob. model						
instr. strength ling. dist.	32.3		49.9		25.0	
and Ling. Dist. x Age at Migr.						

Reference categories dummy variables: Asia; Higher professional/vocational training, university; Single;
**/*Significant at 5/10 % level;
Standard errors are robust to correlation in unobserved errors across time for the same individual (clustering)

of origin (e.g. we do not even observe whether they are migrants). But for individuals who respond in at least one wave, we can assign the time-invariant variables, including country of origin, age at migration, and linguistic distance, to the missing year. We thus can provide insight in wave nonresponse. The nature of nonresponse goes beyond the classical attrition case, where individuals are in the survey in year t but not anymore from some later year, say $t+j$, on. We can have individuals in the survey not responding to the specific questionnaire in year t, who at a later year $t+j$ do respond.

We define respondents with information in the household box in January/February of a given year as 'in the survey': they could have answered the questionnaire on ethnicity and language proficiency. If the number of years in which they answered this questionnaire is lower than the number of years in the survey, there is wave nonresponse. We define an indicator $s_{it} = 1$ if individual i is in the survey in year t and did fill out the questionnaire, zero otherwise, conditional on answering the questionnaire for at least one wave (otherwise we do not observe ethnicity). We first compare the sample means of the subsample with $s_{it} = 1$ with those of $s_{it} = 0$. Note that the subsample with observations with $s_{it} = 1$ is equal to our base sample of $NT = 1303$ observations. The observations with $s_{it} = 0$ have been drawn if their information on the household box is available. These are $NT = 370$ observations that are added to the base sample. The left pane of Table 18 shows sample means of the different subsamples and the differences between those. Age attributes to a higher response, so does being single. The nonrespondents have a higher number of children on average. Linguistic distance and age at migration were strong predictors of language proficiency, but they show no relation to wave response. Being single and the number of children do show a relation with nonresponse, but we saw that they do not have much impact on language proficiency.

We have now been comparing individual-years (it) observations, but in terms of individuals the subsample of observations with nonresponse is a subset of the observations with response, albeit in different years. We do a different comparison defining a variable $k_i = 1$ if an individual answers the questionnaire in all waves in which s/he is in the survey, and is zero if there is at least one wave in which s/he does not respond. This gives us different individuals in the two comparison groups, and therefore we compare individuals, rather than individual-years observations. To compare sample means, we assign individual characteristics observed in the first wave in which the individual is in the survey for the time-variant regressors. We observe 549 different individuals, but for the analysis we dropped 39 individuals who are 'in the survey' for

Table 9 Response to job satisfaction and job suitability questions (migrants, 22<age<65, with paid job)

Response category	Fit education and work	Fit skills work	Job performance	Satisfaction wage	Satisfaction type of work	Satisfaction hours	Satisfaction career
Both men and women							
0	12.4	5.6	0.6	4.5	2.9	0.7	2.1
1	3.6	1.9	1.4	1.2	0.6	0.4	1.1
2	5.6	1.8	1.0	4.5	1.3	2.3	2.0
3	2.8	3.0	1.8	6.1	2.6	1.7	2.3
4	6.6	4.7	3.2	6.3	2.0	2.4	3.3
5	7.4	6.6	4.6	13.3	8.0	8.1	9.0
6	11.7	10.2	5.5	13.8	9.3	11.3	14.3
7	15.7	18.1	12.4	20.8	19.5	21.3	23.9
8	17.9	26.4	20.4	19.5	28.1	27.6	26.3
9	8.2	13.9	21.6	7.2	16.1	13.7	10.4
10	8.2	7.8	27.7	3.0	9.7	10.4	5.3
Nobs	728	728	726	694	701	700	700
Men only							
0	9.3	4.4	0.3	3.2	2.0	0.3	1.4
1	4.4	1.1	1.6	0.9	0.3	0.0	0.9
2	4.1	1.6	1.4	3.5	1.1	2.3	2.3
3	3.3	3.6	2.5	5.8	2.3	0.9	1.7
4	6.3	3.6	2.7	6.3	2.3	3.4	4.0
5	7.7	7.1	3.8	15.0	8.8	9.7	10.2
6	13.7	8.7	5.5	13.0	10.2	10.0	14.7
7	18.3	18.9	12.6	18.4	21.0	20.2	22.4
8	18.3	29.0	17.5	19.6	24.9	28.8	24.4
9	7.7	15.0	22.5	10.1	17.6	13.1	11.9
10	7.1	7.1	29.6	4.3	9.6	11.4	6.2
Nobs	366	366	365	347	353	351	353
Women only							
0	15.5	6.9	0.8	5.8	3.7	1.2	2.9
1	2.8	2.8	1.1	1.4	0.9	0.9	1.4
2	7.2	1.9	0.6	5.5	1.4	2.3	1.7
3	2.2	2.5	1.1	6.3	2.9	2.6	2.9
4	6.9	5.8	3.6	6.3	1.7	1.4	2.6
5	7.2	6.1	5.3	11.5	7.2	6.6	7.8
6	9.7	11.6	5.5	14.7	8.3	12.6	13.8
7	13.0	17.4	12.2	23.1	18.1	22.4	25.4

Table 9 continued

Response category	Fit education and work	Fit skills work	Job performance	Satisfaction wage	Satisfaction type of work	Satisfaction hours	Satisfaction career
8	17.4	23.8	23.3	19.3	31.3	26.4	28.2
9	8.8	12.7	20.8	4.3	14.7	14.3	8.9
10	9.4	8.6	25.8	1.7	9.8	9.5	4.3
Nobs	362	362	361	347	348	349	347

one wave only, for the definition of k_i makes no sense for them. From the resulting 510 observations 295 (57.8%) appear in all waves in which they are 'in the survey', while for 215 (42.2%) missing waves occur. The right pane of Table 18 shows the results of this comparison. We again find an effect of age: older respondents are more likely to respond in all possible waves. The other variables do no show much significance anymore, although they still point in the same direction. Linguistic distance and age at migration do not show any relation with the nonresponse pattern. To get an impression of the more classical panel attrition, we checked how many individuals of the before-mentioned 510 were observed to be in the survey in year 2011, the final year of observation: they are 421 (82.6%), showing panel attrition has a lower incidence than wave nonresponse.

The analysis in Table 18 is univariate. We did a probit analysis with s_{it} and k_i as dependent variables and the variables in Table 18 as regressors to show the combined effect of the regressors. In both cases, we find that age is the only significant variable, showing a positive relation with response (Table 19).

A more formal analysis of nonresponse bias requires a joint analysis of nonresponse and the outcome variable (Hausman and Wise 1979). It also needs exclusion restrictions: variables that do affect response, but not the outcome. Such variables are not available. But the above analysis shows that we cannot detect a relationship between nonresponse and some strong predictors of language proficiency. We hope that this is sufficiently indicative that nonresponse is not a potential major source of bias in our results.

Appendix C: A model of job match satisfaction and language proficiency

The analysis in this study is empirical and data driven. Nevertheless it may be good to capture the underlying concepts in a theoretical model. It should be clear, though, that these theoretical considerations are driven by the empirical framework, and not the other way around. Thus, the aim of the empirical analysis is not to recover every aspect of the theoretical model.

The data concern job satisfaction which we model as an outcome of the match quality of the migrant's job. Suppose that a migrant's optimal (or preferred) job is described by value p^*_M. (More general, p^*_M may represent a vector of job characteristics). It represents the job level the migrant aims for, matching his/her education and

Table 10 Regressions for satisfaction of migrants with job characteristics and fit between education/skills and jobs, including fluency ('speak'). Parameter estimates of coefficients of the simultaneous equations model with random effects

Variable	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Speak (latent)	0.13	0.12	0.32**	0.13	0.15	0.11	0.38**	0.13	0.35**	0.12	0.27**	0.12	−0.12	0.08
Turkey	0.09	0.38	−0.47	0.46	−0.24	0.38	−0.73	0.48	0.07	0.49	−0.22	0.41	0.04	0.29
Moroccan	0.75	0.47	0.63	0.51	0.01	0.42	0.05	0.52	0.43	0.61	0.47	0.45	0.22	0.33
Dutch Antilles	0.35	0.63	−0.12	0.60	−0.41	0.52	−0.76	0.66	−0.43	0.67	−0.18	0.56	0.59	0.39
Suriname	0.05	0.61	−0.76	0.79	−0.44	0.65	−1.20	0.74	−0.51	0.79	−0.39	0.72	1.17**	0.47
Indonesia	−0.35	0.53	−1.06	0.67	−0.32	0.50	−1.15	0.72	−0.21	0.64	−0.25	0.60	0.23	0.39
German/Nordic	0.30	0.53	0.08	0.66	−0.01	0.51	−0.51	0.64	0.34	0.68	0.12	0.58	0.75*	0.39
Anglo-American	0.48	0.44	0.30	0.59	−0.49	0.44	−0.24	0.57	0.49	0.64	0.69	0.49	0.35	0.35
Latin	−0.04	0.42	−0.33	0.52	−0.45	0.43	−1.00*	0.58	−0.21	0.56	−0.38	0.50	0.29	0.32
Africa	−0.41	0.69	−0.71	0.67	−1.04*	0.55	−1.05	0.72	−0.30	0.71	−0.75	0.66	0.75*	0.45
Eastern Europe	0.45	0.39	0.60	0.57	0.11	0.40	0.20	0.54	−0.02	0.58	0.37	0.44	0.13	0.33
Geographic distance	0.02	0.03	0.06	0.04	0.03	0.03	0.05	0.05	0.03	0.05	0.02	0.04	−0.05*	0.03
Genetic distance	0.35	2.55	1.19	3.58	−0.78	2.49	1.22	3.69	−0.32	3.84	3.15	2.96	−0.45	2.03
Prim. Education	−0.26	0.45	0.94	0.48	0.15	0.40	0.70	0.45	−0.04	0.46	0.21	0.37	−0.51*	0.28
Lower voc./prof.	0.05	0.34	0.43	0.41	0.28	0.31	0.39	0.39	−0.30	0.37	−0.09	0.31	−0.20	0.22
Higher sec./middle voc.	0.05	0.22	0.28	0.26	0.05	0.20	0.11	0.26	0.10	0.26	0.06	0.22	−0.28*	0.16
Age at migration/10	0.34	0.36	1.07**	0.44	0.37	0.36	1.26**	0.47	0.92**	0.48	0.97**	0.44	−0.46*	0.28
Age at migr./10 squared	−0.08	0.06	−0.16**	0.07	−0.06	0.06	−0.19**	0.08	−0.11	0.09	−0.15**	0.07	0.06	0.05

Table 10 continued

Variable	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Age/10	0.11	0.12	0.01	0.16	−0.09	0.13	−0.08	0.18	−0.18	0.17	−0.16	0.15	0.11	0.10
Female	−0.46*	0.19	−0.20	0.21	−0.13	0.15	−0.17	0.23	−0.05	0.21	−0.15	0.18	−0.08	0.13
# Children	−0.76	0.90	1.12	1.03	1.32	0.85	1.58	1.09	2.65**	1.12	0.83	0.88	−1.04	0.68
Couple	0.07	0.21	−0.10	0.23	−0.13	0.19	−0.12	0.25	−0.57**	0.24	−0.16	0.21	0.08	0.16
Intercept1	−1.89**	0.61	−1.07	0.74	−3.29**	0.63	−1.77**	0.81	−1.29*	0.78	−1.43**	0.65	−3.39**	0.50
Intercept2	−1.71**	0.61	−0.94	0.74	−3.06**	0.61	−1.49*	0.80	−0.99	0.78	−1.19*	0.64	−2.85**	0.47
Intercept3	−1.22**	0.61	−0.70	0.74	−2.44**	0.59	−1.15	0.79	−0.61	0.77	−1.01	0.64	−2.65**	0.47
Intercept4	−0.76	0.60	−0.32	0.73	−2.18**	0.59	−0.85	0.79	−0.44	0.77	−0.75	0.64	−2.40**	0.46
Intercept5	−0.39	0.60	−0.09	0.73	−1.90**	0.59	−0.52	0.79	−0.10	0.77	−0.44	0.63	−2.09**	0.46
Intercept6	0.25	0.60	0.57	0.73	−1.30**	0.58	0.12	0.78	0.25	0.77	−0.09	0.63	−1.78**	0.46
Intercept7	0.86	0.60	1.08	0.73	−0.75	0.58	0.81	0.78	0.78	0.77	0.35	0.63	−1.48**	0.46
Intercept8	1.80**	0.60	1.94**	0.73	0.05	0.57	1.75**	0.78	1.47*	0.77	1.00	0.63	−0.98**	0.45
Intercept9	3.02**	0.60	3.12**	0.74	1.12*	0.57	2.97**	0.79	2.42**	0.77	1.98**	0.63	−0.32	0.45
Intercept10	4.00**	0.62	4.10**	0.75	1.93**	0.58	3.90**	0.81	3.08**	0.78	2.83**	0.64	0.39	0.45
$\rho_{\epsilon v}$	−0.21	0.14	−0.22	0.15	−0.14	0.15	−0.29*	0.16	−0.26	0.17	−0.30*	0.16	0.09	0.14
σ_m	2.35**	0.47	2.45**	0.50	2.14**	0.36	2.42**	0.49	2.29**	0.43	2.52**	0.49	2.14**	0.36
σ_θ	1.26**	0.10	1.41**	0.22	1.07**	0.13	1.46**	0.21	1.45**	0.17	1.22**	0.19	0.77**	0.13
$\rho_{\theta m}$	−0.12	0.27	−0.58**	0.17	−0.42**	0.21	−0.64**	0.14	−0.53**	0.17	−0.63**	0.16	0.63**	0.16
Nobs =	749		750		750		750		735		735		735	
Likelihood value	−1623.8		−1519.7		−1542.2		−1525.2		−1732.6		−1681.9		−1568.3	

Exclusion restriction in language proficiency equation: Cross effect linguistic distance and age at migration
Coefficient estimates of language proficiency equation not shown in table

Table 11 Regressions for satisfaction of migrants with job characteristics and fit between education/skills and jobs, including fluency ('speak'), males only. Parameter estimates of coefficients of the simultaneous equations model with random effects

Variable	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Speak (latent)	0.29	0.19	0.54**	0.16	0.25*	0.14	0.51**	0.15	0.37**	0.13	0.32**	0.12	−0.19	0.12
Turkey	0.11	0.55	−0.15	0.67	−0.29	0.46	−0.51	0.67	−0.06	0.52	−0.75*	0.45	0.11	0.38
Moroccan	0.28	0.64	0.27	0.73	−0.51	0.52	−0.29	0.72	−0.31	0.57	−0.12	0.51	0.33	0.42
Dutch Antilles	−0.37	0.94	−1.62*	0.83	−1.55**	0.78	−2.12**	0.89	−1.21	0.83	−1.70**	0.71	0.62	0.62
Suriname	−1.08	1.09	−2.28**	1.10	−1.24	0.84	−2.13**	0.99	−1.11	0.89	−1.20	0.75	1.17*	0.68
Indonesia	−0.07	0.87	−1.31	0.99	−0.21	0.67	−0.90	1.01	−0.30	0.78	−0.66	0.67	0.08	0.55
German/Nordic	−0.43	0.90	−0.37	0.93	−0.14	0.65	−1.06	0.90	1.06	0.74	−0.12	0.62	1.08**	0.54
Anglo-American	0.57	0.68	−0.33	0.80	−0.63	0.53	−0.71	0.75	0.62	0.60	0.32	0.52	0.42	0.43
Latin	−0.38	0.63	−0.53	0.71	−0.47	0.52	−0.86	0.72	−0.53	0.57	−0.57	0.51	0.50	0.43
Africa	0.02	1.06	−2.17**	0.99	−0.94	0.78	−1.53	0.95	−2.00**	0.89	−2.24**	0.77	0.98	0.67
Eastern Europe	0.05	0.61	0.38	0.80	0.10	0.55	−0.27	0.75	0.45	0.66	0.59	0.52	0.20	0.44
Geographic distance	0.04	0.06	0.12**	0.06	0.04	0.04	0.06	0.06	0.09	0.05	0.07	0.04	0.00	0.04
Genetic distance	−1.27	4.17	0.33	4.55	−3.46	3.17	−1.06	4.55	1.21	4.08	3.72	3.40	−1.98	2.60
Prim. Education	−0.59	0.63	1.38**	0.63	0.50	0.52	1.13	0.73	0.29	0.49	0.70*	0.41	−0.75*	0.40
Lower voc./prof.	−0.16	0.61	0.95*	0.56	0.63	0.41	0.88*	0.53	−0.23	0.42	0.48	0.35	−0.31	0.32
Higher sec./middle voc.	0.11	0.35	0.48	0.37	0.33	0.28	0.38	0.36	0.21	0.30	0.28	0.26	−0.34	0.22
Age at migration/10	1.44**	0.67	1.39**	0.55	0.51	0.46	1.44**	0.53	0.94*	0.51	0.92**	0.44	−0.69*	0.39
Age at migr./10 squared	−0.28**	0.11	−0.14	0.11	−0.06	0.08	−0.20*	0.11	−0.12	0.09	−0.14*	0.08	0.09	0.07
Age/10	−0.04	0.22	−0.28	0.21	−0.11	0.17	−0.25	0.22	−0.15	0.18	−0.23	0.16	0.14	0.14
# Children	0.73	1.53	1.72	1.46	2.34**	1.16	2.11	1.50	3.11**	1.25	1.56	1.07	−1.41	0.95

Table 11 continued

Variable	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Couple	−0.33	0.40	−0.56	0.35	−0.66**	0.28	−0.47	0.36	−0.49	0.31	−0.21	0.28	0.13	0.24
Intercept1	−1.83*	1.00	−1.34	1.07			−2.23**	1.12	−0.77	0.89	−1.55**	0.78	−3.45**	0.71
Intercept2	−1.65*	1.00	−1.27	1.06			−2.02*	1.11	−0.38	0.88	−1.39*	0.78	−2.74**	0.64
Intercept3	−1.14	0.98	−1.05	1.06			−1.64	1.10	−0.08	0.88	−1.19	0.77	−2.51**	0.64
Intercept4	−0.60	0.97	−0.74	1.05	−2.43**	0.82	−1.43	1.10	0.12	0.88	−0.87	0.77	−2.23**	0.63
Intercept5	−0.20	0.96	−0.50	1.04	−1.97**	0.81	−1.04	1.09	0.44	0.87	−0.64	0.77	−2.01**	0.63
Intercept6	0.54	0.95	0.11	1.04	−1.27	0.79	−0.39	1.08	0.79	0.87	−0.26	0.76	−1.77**	0.63
Intercept7	1.13	0.94	0.60	1.03	−0.82	0.78	0.25	1.07	1.36	0.87	0.11	0.76	−1.49**	0.62
Intercept8	2.02**	0.93	1.37	1.03	−0.07	0.77	1.09	1.07	2.09**	0.87	0.74	0.75	−1.01	0.62
Intercept9	3.19**	0.94	2.26**	1.05	1.00	0.76	2.16**	1.07	3.00**	0.89	1.70**	0.75	−0.46	0.61
Intercept10	4.31**	0.96	3.15**	1.07	1.71**	0.77	3.06**	1.09	3.59**	0.90	2.55**	0.76	0.22	0.61
$\rho_{\epsilon v}$	−0.17	0.27	−0.51**	0.16	−0.21	0.22	−0.42	0.18	−0.29	0.21	−0.31*	0.18	0.26	0.18
σ_m	2.33**	0.44	2.28**	0.41	2.20**	0.39	2.30**	0.43	2.37**	0.44	2.35**	0.42	2.11**	0.34
σ_θ	1.48**	0.29	1.73**	0.34	1.05**	0.23	1.67**	0.32	1.23**	0.26	1.06**	0.24	0.77**	0.23
$\rho_{\theta m}$	−0.40	0.31	−0.87**	0.07	−0.65**	0.20	−0.82**	0.09	−0.72**	0.14	−0.79**	0.12	0.74**	0.20
Nobs	749		750		750		750		735		735		735	
Likelihood value	−927.5		−884.6		−867.9		−895.6		−973.0		−940.8		−910.3	

Exclusion restriction in language proficiency equation: Cross-effect linguistic distance and age at migration
Coefficient estimates of language proficiency equation not shown in table

Table 12 Regressions for satisfaction of migrants with job characteristics and fit between education/skills and jobs, including fluency ('speak'), females only. Parameter estimates of coefficients of the simultaneous equations model with random effects

Variable	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Speak (latent)	0.17	0.15	0.23	0.18	0.19	0.17	0.36**	0.18	0.32*	0.18	0.08	0.15	−0.05	0.11
Turkey	−0.09	0.59	−1.02	0.72	0.15	0.65	−1.09	0.74	0.29	0.76	0.37	0.60	−0.31	0.46
Moroccan	1.00	0.68	0.59	0.84	0.88	0.72	0.11	0.77	1.62**	0.80	1.21*	0.69	−0.02	0.53
Dutch Antilles	0.21	0.71	0.59	0.86	−0.06	0.91	−0.40	0.93	−0.21	0.90	0.96	0.68	0.34	0.50
Suriname	−0.35	1.04	−0.18	1.25	−0.29	1.15	−1.05	1.20	0.54	1.29	1.67	1.14	1.32*	0.80
Indonesia	−0.48	0.66	−0.72	0.82	−0.92	0.84	−1.14	0.89	−0.31	0.95	0.54	0.74	0.30	0.55
German/Nordic	−0.01	0.83	−0.13	0.98	−0.01	0.91	−0.43	1.00	0.87	1.08	0.93	0.90	0.21	0.63
Anglo-American	−0.10	0.75	0.20	0.95	−0.27	0.86	−0.34	0.98	0.92	1.26	1.55*	0.87	−0.10	0.62
Latin	−0.10	0.70	−0.28	0.83	−0.31	0.79	−0.92	0.89	0.50	0.88	0.83	0.79	0.13	0.54
Africa	−0.57	0.77	0.42	1.04	−1.14	0.83	−0.54	0.90	0.93	1.03	0.99	0.87	0.29	0.62
Eastern Europe	0.15	0.62	0.31	0.74	0.33	0.65	0.31	0.79	0.40	0.89	0.55	0.66	−0.18	0.49
Geographic distance	−0.03	0.05	−0.01	0.06	0.04	0.06	0.05	0.06	−0.04	0.08	−0.09	0.06	−0.16**	0.05
Genetic distance	2.22	4.11	3.23	4.81	4.52	4.45	1.93	4.79	7.51	6.97	7.76*	4.32	4.51	3.42
Prim. Education	−0.14	0.57	0.70	0.63	0.26	0.59	0.85	0.62	−0.52	0.63	−0.17	0.49	−0.07	0.41
Lower voc./prof.	0.48	0.43	0.09	0.52	0.04	0.51	0.45	0.46	−0.20	0.51	−0.29	0.37	−0.01	0.29
Higher sec./middle voc.	−0.05	0.27	0.06	0.32	−0.26	0.30	0.06	0.30	0.16	0.36	−0.01	0.27	−0.23	0.21
Age at migration/10	0.53	0.55	1.19*	0.67	0.88	0.60	1.26**	0.61	1.12*	0.68	0.58	0.56	−0.20	0.41
Age at migr./10 squared	−0.08	0.10	−0.23*	0.12	−0.15	0.10	−0.19*	0.11	−0.16	0.12	−0.12	0.10	0.03	0.07

Table 12 continued

Variable	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Age/10	0.05	0.17	0.21	0.20	−0.14	0.18	−0.12	0.21	−0.09	0.23	0.05	0.17	0.13	0.13
# Children	−0.73	1.19	0.73	1.34	−0.14	1.27	0.99	1.50	1.06	1.99	−1.30	1.26	−1.12	0.97
Couple	0.12	0.24	0.13	0.28	0.23	0.27	−0.10	0.30	−0.64**	0.32	−0.16	0.24	0.09	0.20
Intercept1	−1.32	0.92	−0.72	1.00	−2.33**	0.96	−1.91*	1.04	−0.49	1.08	−1.19	0.85	−3.39**	0.72
Intercept2	−1.15	0.91	−0.52	1.00	−2.03**	0.94	−1.56	1.03	−0.26	1.08	−0.85	0.85	−2.95**	0.69
Intercept3	−0.70	0.90	−0.25	0.99	−1.55*	0.92	−1.24	1.01	0.27	1.09	−0.65	0.84	−2.82**	0.68
Intercept4	−0.33	0.89	0.19	0.98	−1.22	0.91	−0.85	1.01	0.42	1.09	−0.42	0.84	−2.60**	0.67
Intercept5	−0.01	0.88	0.41	0.98	−1.07	0.91	−0.57	1.00	0.85	1.09	−0.01	0.84	−2.15**	0.66
Intercept6	0.52	0.87	1.06	0.98	−0.54	0.90	0.03	0.99	1.25	1.10	0.35	0.84	−1.75**	0.65
Intercept7	1.13	0.87	1.58	0.98	0.15	0.90	0.75	0.99	1.77	1.11	0.91	0.84	−1.43**	0.65
Intercept8	2.11**	0.87	2.45**	0.98	1.04	0.90	1.80*	0.99	2.47**	1.12	1.64*	0.85	−0.90	0.65
Intercept9	3.40**	0.88	3.83**	1.00	2.14**	0.91	3.20**	1.00	3.57**	1.14	2.70**	0.86	−0.09	0.64
Intercept10	4.19**	0.91	4.80**	1.02	3.06**	0.92	4.16**	1.02	4.36**	1.16	3.59**	0.87	0.65	0.64
$\rho_{\epsilon v}$	−0.43**	0.18	−0.16	0.22	−0.23	0.21	−0.17	0.23	0.00	0.27	−0.18	0.21	−0.12	0.21
σ_m	2.24**	0.40	2.29**	0.40	2.28**	0.40	2.30**	0.43	2.32**	0.43	2.33**	0.43	2.22**	0.40
σ_θ	1.07**	0.13	1.35**	0.20	1.22**	0.18	1.45**	0.24	1.70**	0.27	1.26**	0.14	0.78**	0.16
$\rho_{\theta m}$	−0.11	0.39	−0.27	0.34	−0.35	0.32	−0.56**	0.20	−0.34	0.23	−0.21	0.29	0.47*	0.28
Nobs	749		750		750		750		735		735		735	
Likelihood value	−932.2		−868.2		−891.8		−870.3		−973.5		−957.7		−887.2	

Exclusion restriction in language proficiency equation: Cross effect linguistic distance and age at migration
Coefficient estimates of language proficiency equation not shown in table

Table 13 Coefficients of the cross-effects with age at migration in the equation for language proficiency

Coefficients	Coef.	Std.
Age at migration × linguistic distance	−0.25**	0.08
Age at migration × geographic distance	−0.04	0.03
Age at migration × genetic distance	1.03	1.91
Age at migration × western country	−0.22	0.14
Age at migration × Dutch colony	0.27	0.20

Age at migration normalized by dividing by 10

skills. This value thus depends on the migrant's observed and unobserved characteristics x (among which can be educational attainment) and possibly also on the migrant's destination country language proficiency l, characterizing the optimal job as $p^*_M(l, x)$. The attainable job (say, the best available job) has value $p_{JM}(l, x)$, also depending on language proficiency l and characteristics x, with

$$p_{JM}(l, x) \leq p^*_M(l, x) \tag{4}$$

Suppose that l can be expressed at a scale from 0 to 1, with 1 indicating perfect language proficiency and 0 no proficiency at all. In general, the value of an attainable job can be lower than the preferred value due to labour market frictions, but focussing on the role of language proficiency in determining job match quality we assume

$$p_{JM}(1, x) = p^*_M(1, x) \tag{5}$$

meaning that the migrant can reach the preferred job (given x) if proficiency is at the highest level. Let $D(.)$ be a distance metric, with $D(.) \geq 0$, $D(0) = 0$ and $D'(.) > 0$. Then, the match quality depends on the distance between the actual job and the optimal job, while job satisfaction, JS, is a decreasing function of this distance:

$$\begin{aligned} &D\left(p_{JM}(l, x) - p^*_M(l, x)\right) \text{ with } D'(.) > 0, \\ &JS\left[D\left(p_{JM}(l, x) - p^*_M(l, x)\right)\right] \text{ with } JS'[.] < 0 \end{aligned} \tag{6}$$

According to (4) and (5), the highest level of job satisfaction is obtained if $l = 1$. Language proficiency affects job satisfaction if it creates a difference between the optimal job and the attainable job. If language proficiency reduces this distance we have

$$\frac{\partial D\left(p_{JM}(l, x) - p^*_M(l, x)\right)}{\partial l} < 0 \tag{7}$$

leading to an increase in job satisfaction upon an increase in l. To interpret things further, starting from $l = 1$ in (5), a lower language proficiency than $l = 1$ will have no impact (that is, the derivative in (7) equals zero) in: (i) the case where language proficiency neither affects the attainable job, nor the optimal job; (ii) the case where a lower language proficiency affects the attainable job in the same way as the optimal

Table 14 Coefficients of the cross-effects with age at migration in the sensitivity analysis. Results from the simultaneous equations model with random effects

Coefficients labour market outcomes Eq. (2)	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Results for males														
Age at migration × geogr. distance	0.05	0.05	0.08	0.06	0.04	0.04	0.09	0.06	0.03	0.04	0.06	0.04	−0.02	0.03
Age at migration × genetic distance	2.76	3.63	−1.74	3.14	1.12	2.29	1.73	3.25	2.54	2.54	1.45	2.43	−2.37	2.01
Age at migration × western country	0.45	0.29	0.58*	0.30	0.31	0.22	0.82**	0.32	0.68**	0.23	0.37*	0.21	−0.24	0.20
Age at migration × Dutch colony	−0.24	0.50	−1.01**	0.45	−0.46	0.34	−1.05**	0.45	−0.24	0.38	−0.58	0.36	0.70**	0.28
Results for females														
Age at migration × geogr. distance	−0.03	0.04	−0.04	0.05	0.03	0.05	0.03	0.05	0.04	0.05	−0.05	0.04	0.01	0.03
Age at migration × genetic distance	0.06	2.17	−1.14	2.32	−0.83	2.09	1.65	2.57	0.52	2.79	0.53	2.02	−2.19	1.65
Age at migration × western country	−0.03	0.25	0.11	0.27	−0.03	0.29	0.37	0.32	0.40	0.27	0.15	0.25	−0.05	0.20
Age at migration × Dutch colony	−0.05	0.33	0.28	0.41	0.16	0.36	0.02	0.45	0.38	0.52	0.56	0.38	0.07	0.27

Age at migration normalized by dividing by 10

Table 15 Regressions for satisfaction of migrants with job characteristics and fit between education/skills and jobs. Parameter estimates of model with professional levels among the explanatory variables

Variable	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Turkey	0.19	0.25	0.00	0.23	0.03	0.24	−0.10	0.24	0.16	0.24	0.03	0.22	0.12	0.22
Moroccan	0.13	0.29	0.24	0.28	−0.16	0.32	−0.14	0.28	0.03	0.30	0.28	0.26	0.25	0.30
Dutch Antilles	0.30	0.26	0.41*	0.23	−0.11	0.24	0.21	0.25	0.26	0.24	0.29	0.23	0.17	0.21
Suriname	0.12	0.26	0.27	0.23	0.00	0.25	0.15	0.24	0.53**	0.23	0.41**	0.20	0.43**	0.21
Indonesia	0.09	0.31	−0.10	0.25	0.00	0.26	−0.24	0.27	0.29	0.28	0.18	0.28	0.00	0.25
German/Nordic	0.13	0.27	0.35	0.24	−0.04	0.27	−0.04	0.28	0.71**	0.29	0.43*	0.25	0.46*	0.25
Anglo-American	0.00	0.27	0.18	0.22	−0.52**	0.26	−0.23	0.24	0.34	0.23	0.31*	0.19	−0.12	0.24
Latin	0.13	0.26	0.22	0.22	−0.07	0.23	−0.07	0.25	0.21	0.25	0.21	0.23	0.11	0.24
Africa	0.02	0.38	0.19	0.29	−0.19	0.37	0.22	0.28	0.01	0.42	0.19	0.38	0.36	0.29
Eastern Europe	0.36	0.28	0.53*	0.28	0.05	0.27	0.38	0.28	0.36	0.30	0.52**	0.25	0.21	0.24
Geographic distance	0.00	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.01	0.02	−0.01	0.02
Genetic distance	−0.72	1.54	−0.84	1.64	−2.12	1.50	−1.61	1.59	−0.77	1.81	0.01	1.51	−1.58	1.72
Prim. Education	0.04	0.28	0.68**	0.26	0.06	0.26	0.37	0.29	0.19	0.31	0.55*	0.28	−0.22	0.23
Lower voc./prof.	0.35*	0.21	0.24	0.17	0.13	0.16	0.23	0.17	−0.11	0.19	0.27	0.17	0.02	0.15
Higher sec./middle voc.	0.25*	0.15	0.14	0.15	−0.04	0.13	0.07	0.15	0.24*	0.13	0.23*	0.13	−0.13	0.12
Age at migration/10	0.04	0.16	0.18	0.14	−0.02	0.15	0.11	0.14	−0.02	0.15	0.07	0.14	−0.23*	0.13
Age at migr./10 squared	−0.04	0.04	−0.06	0.04	−0.02	0.04	−0.06	0.04	−0.01	0.04	−0.03	0.04	0.05	0.03
Age/10	0.10	0.07	0.13**	0.06	0.03	0.06	0.17**	0.06	0.12*	0.06	0.14**	0.06	0.06	0.05

Table 15 continued

Variable	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Female	−0.31**	0.12	−0.18	0.12	−0.18	0.11	−0.14	0.12	−0.22*	0.12	−0.35**	0.11	−0.18*	0.10
#Children	0.09	0.59	0.80	0.57	0.95*	0.56	0.54	0.56	1.89**	0.51	0.61	0.54	−0.45	0.52
Couple	0.01	0.13	−0.10	0.13	−0.19	0.13	−0.04	0.13	−0.28**	0.12	−0.09	0.12	0.04	0.11
Higher academic	1.23**	0.25	1.22**	0.21	0.35	0.25	1.12**	0.24	1.31**	0.26	1.24**	0.25	−0.44*	0.22
Higher supervisory	1.02**	0.30	1.41**	0.25	0.42*	0.25	1.22**	0.27	0.90**	0.28	1.02**	0.26	0.09	0.23
Intermediate academic	0.83**	0.20	1.32**	0.20	0.68**	0.19	1.05**	0.21	1.27**	0.20	1.31**	0.19	0.18	0.18
Intermediate supervisory	0.46**	0.23	0.86**	0.21	0.53**	0.20	0.82**	0.23	0.69**	0.21	0.84**	0.22	0.01	0.19
Other mental work	0.41**	0.18	0.94**	0.20	0.52**	0.18	0.72**	0.19	0.76**	0.19	0.88**	0.18	−0.16	0.16
Skilled manual work	0.29	0.30	0.59**	0.25	0.29	0.21	0.66**	0.25	0.82**	0.23	0.42**	0.19	−0.59**	0.21
Intercept1	−0.85*	0.39	−0.41	0.35	−2.34**	0.45	−0.87**	0.37	0.04	0.39	−0.22	0.37	−2.86**	0.38
Intercept2	−0.74	0.38	−0.31	0.35	−2.20**	0.42	−0.65	0.38	0.24	0.40	−0.03	0.37	−2.30**	0.36
Intercept3	−0.41	0.38	−0.13	0.36	−1.71**	0.41	−0.40	0.38	0.51	0.40	0.11	0.38	−2.14**	0.36
Intercept4	−0.10	0.38	0.14	0.36	−1.50**	0.41	−0.18	0.37	0.62	0.40	0.29	0.38	−1.92**	0.37
Intercept5	0.15	0.38	0.31	0.36	−1.31**	0.40	0.06	0.38	0.87**	0.41	0.54	0.38	−1.65**	0.36
Intercept6	0.56	0.38	0.77**	0.35	−0.83**	0.40	0.51	0.37	1.09**	0.41	0.81**	0.38	−1.39**	0.36
Intercept7	0.95**	0.38	1.16**	0.35	−0.42	0.40	1.01**	0.37	1.45**	0.41	1.17**	0.38	−1.13**	0.36
Intercept8	1.55**	0.38	1.76**	0.36	0.17	0.40	1.68**	0.37	1.92**	0.42	1.71**	0.38	−0.68*	0.36
Intercept9	2.35**	0.39	2.60**	0.36	0.96**	0.40	2.55**	0.38	2.57**	0.42	2.52**	0.39	−0.10	0.35
Intercept10	3.04**	0.40	3.29**	0.37	1.54**	0.41	3.19**	0.39	3.04**	0.44	3.20**	0.40	0.52	0.35
Nobs	681		688		687		687		699		699		699	
Likelihood Value	−1413.5		−1279.7		−1301.9		−1300.7		−1481.9		−1385.5		−1271.9	

Table 16 Regressions for satisfaction of migrants with job characteristics and fit between education/skills and jobs. Parameter estimates of model with professional levels among the explanatory variables, males only

Variable	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Turkey	0.25	0.35	0.17	0.29	0.04	0.32	0.02	0.34	0.10	0.34	−0.44*	0.25	0.09	0.26
Moroccan	0.08	0.37	0.24	0.37	−0.37	0.48	−0.07	0.40	−0.32	0.36	0.05	0.31	0.39	0.33
Dutch Antilles	0.43	0.38	0.17	0.34	−0.30	0.38	0.34	0.38	0.37	0.39	0.00	0.40	−0.14	0.32
Suriname	−0.01	0.36	0.11	0.31	−0.09	0.30	0.04	0.32	0.54*	0.32	0.11	0.27	0.15	0.27
Indonesia	0.30	0.43	−0.10	0.33	0.23	0.37	0.03	0.38	0.48	0.41	0.18	0.38	−0.22	0.40
German/Nordic	−0.28	0.35	0.41	0.37	0.24	0.37	−0.19	0.36	1.26**	0.28	0.19	0.26	0.63**	0.32
Anglo-American	0.15	0.31	0.14	0.29	−0.32	0.28	−0.16	0.32	0.60**	0.28	0.27	0.22	−0.19	0.29
Latin	0.23	0.34	0.40	0.29	0.10	0.29	0.02	0.34	0.28	0.35	0.08	0.30	0.17	0.30
Africa	0.54	0.50	0.01	0.43	−0.02	0.41	0.45	0.45	−0.87*	0.48	−0.67	0.67	0.29	0.39
Eastern Europe	0.26	0.49	0.61	0.46	0.17	0.33	0.27	0.48	0.51	0.41	0.62*	0.32	0.26	0.29
Geographic distance	0.01	0.02	0.05*	0.02	0.01	0.02	0.01	0.02	0.03	0.03	0.02	0.02	0.03	0.03
Genetic distance	−2.26	1.85	−1.88	1.56	−3.26**	1.46	−2.68	1.94	−0.76	2.01	−0.17	1.65	−2.44	1.96
Prim. Education	−0.08	0.43	0.63*	0.32	0.04	0.31	0.35	0.42	0.44	0.40	0.80**	0.37	−0.13	0.29
Lower voc./prof.	0.17	0.34	0.34	0.28	0.04	0.27	0.41	0.29	−0.24	0.25	0.59**	0.23	0.19	0.23
Higher sec./middle voc.	0.32	0.28	0.12	0.23	−0.03	0.21	0.15	0.25	0.32	0.21	0.32*	0.19	0.01	0.18
Age at migration/10	0.18	0.24	−0.02	0.19	−0.04	0.23	0.05	0.21	−0.15	0.20	0.08	0.19	−0.28	0.20
Age at migr./10 squared	−0.08	0.05	−0.03	0.05	−0.02	0.06	−0.06	0.05	0.01	0.05	−0.06	0.05	0.06	0.05
Age/10	0.12	0.10	0.13	0.09	0.16*	0.09	0.19**	0.09	0.13	0.09	0.07	0.09	0.04	0.08

Table 16 continued

Variable	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
# Children	0.66	0.85	1.06	0.89	1.59*	0.94	1.01	0.87	2.33**	0.81	1.07	0.76	−0.23	0.75
Couple	−0.20	0.21	−0.17	0.22	−0.45**	0.22	−0.03	0.21	−0.20	0.19	0.03	0.20	−0.08	0.19
Higher academic	1.24**	0.33	1.21**	0.29	0.17	0.35	1.40**	0.33	1.14**	0.33	1.11**	0.28	−0.26	0.34
Higher supervisory	0.84**	0.35	1.38**	0.29	0.19	0.31	1.33**	0.34	0.76**	0.35	0.91**	0.32	0.05	0.28
Intermediate academic	0.81**	0.36	1.03**	0.33	0.55*	0.33	1.10**	0.36	0.97**	0.32	0.95**	0.29	0.36	0.28
Intermediate supervisory	0.66*	0.36	0.89**	0.32	0.47	0.34	1.05**	0.37	0.52	0.31	0.91**	0.30	0.03	0.26
Other mental work	0.40	0.27	1.22**	0.25	0.65**	0.26	1.01**	0.25	0.62**	0.28	0.90**	0.26	−0.17	0.23
Skilled manual work	0.27	0.32	0.51*	0.29	0.21	0.25	0.74**	0.28	0.72**	0.27	0.18	0.22	−0.64**	0.24
Intercept1	−0.96*	0.51	−0.84*	0.48			−0.88	0.50	−0.16	0.56	−0.68	0.51	−2.97**	0.56
Intercept2	−0.88*	0.50	−0.77	0.48	−2.43**	0.64	−0.66	0.51	0.16	0.57	−0.53	0.52	−2.22**	0.48
Intercept3	−0.53	0.51	−0.55	0.49	−1.64**	0.53	−0.34	0.51	0.41	0.58	−0.37	0.51	−2.03**	0.49
Intercept4	−0.17	0.50	−0.25	0.49	−1.50**	0.53	−0.15	0.50	0.57	0.58	−0.09	0.52	−1.78**	0.51
Intercept5	0.10	0.50	−0.03	0.48	−1.16**	0.52	0.17	0.50	0.84	0.58	0.14	0.52	−1.56**	0.50
Intercept6	0.61	0.51	0.52	0.48	−0.57	0.51	0.70	0.51	1.08*	0.58	0.45	0.52	−1.35**	0.51
Intercept7	0.99*	0.51	0.95*	0.48	−0.18	0.51	1.22**	0.51	1.53**	0.58	0.78	0.52	−1.10**	0.50
Intercept8	1.52**	0.50	1.59**	0.47	0.43	0.52	1.89**	0.51	2.10**	0.59	1.37**	0.52	−0.62	0.50
Intercept9	2.22**	0.51	2.33**	0.48	1.27**	0.52	2.67**	0.52	2.82**	0.59	2.25**	0.53	−0.13	0.49
Intercept10	2.92**	0.51	3.08**	0.49	1.82**	0.54	3.35**	0.52	3.28**	0.62	2.96**	0.56	0.51	0.49
Nobs	342		348		346		348		351		351		351	
Likelihood Value	−704.1		−641.6		−629.9		−654.7		−717.9		−669.8		−634.2	

Table 17 Regressions for satisfaction of migrants with job characteristics and fit between education/skills and jobs. Parameter estimates of model with professional levels among the explanatory variables, females only

Variable	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
Turkey	0.04	0.42	−0.49	0.39	0.23	0.37	−0.38	0.42	0.15	0.37	0.36	0.40	−0.16	0.37
Moroccan	0.18	0.46	0.34	0.49	0.24	0.54	−0.10	0.47	0.54	0.50	0.70	0.48	−0.02	0.54
Dutch Antilles	0.17	0.38	0.48	0.31	−0.01	0.29	0.13	0.37	−0.06	0.33	0.41	0.31	0.36	0.26
Suriname	0.14	0.43	0.53	0.39	0.19	0.44	0.34	0.44	1.02**	0.39	0.99**	0.38	0.65*	0.37
Indonesia	−0.13	0.44	−0.25	0.34	−0.35	0.39	−0.37	0.37	−0.06	0.41	0.04	0.38	0.20	0.29
German/Nordic	0.17	0.48	0.20	0.37	−0.01	0.45	0.03	0.51	0.64	0.49	0.74	0.48	0.01	0.39
Anglo-American	−0.40	0.52	0.24	0.41	−0.46	0.48	−0.38	0.49	0.48	0.43	0.68	0.42	−0.51	0.40
Latin	−0.04	0.43	0.03	0.31	−0.11	0.42	−0.12	0.43	0.36	0.36	0.46	0.40	−0.07	0.35
Africa	−0.49	0.53	0.31	0.43	−0.26	0.53	0.26	0.42	0.54	0.55	0.77*	0.43	0.32	0.46
Eastern Europe	0.17	0.45	0.35	0.36	0.10	0.44	0.42	0.44	0.32	0.44	0.62	0.44	−0.18	0.38
Geographic distance	−0.03	0.04	−0.02	0.04	0.03	0.03	0.02	0.04	−0.05	0.04	−0.03	0.04	−0.10**	0.04
Genetic distance	1.52	3.42	3.02	3.03	0.52	3.04	0.09	3.46	4.93	3.33	4.60	3.18	0.72	3.46
Prim. Education	0.26	0.39	0.93**	0.43	0.11	0.49	0.67	0.42	0.03	0.42	0.39	0.33	−0.19	0.38
Lower voc./prof.	0.51*	0.28	0.18	0.23	0.12	0.22	0.19	0.24	0.14	0.28	0.22	0.25	−0.12	0.23
Higher sec./middle voc.	0.14	0.18	0.08	0.18	−0.18	0.17	0.05	0.19	0.21	0.18	0.22	0.18	−0.26	0.16
Age at migration/10	0.13	0.22	0.51**	0.20	0.15	0.22	0.29	0.22	0.25	0.22	0.25	0.21	−0.24	0.18
Age at migr./10 squared	−0.04	0.06	−0.12**	0.05	−0.05	0.05	−0.08	0.06	−0.06	0.06	−0.06	0.05	0.05	0.05
Age/10	0.11	0.10	0.16	0.08	−0.04	0.08	0.12	0.09	0.10	0.09	0.18**	0.09	0.14	0.08

Table 17 continued

Variable	Satisfaction wage		Satisfaction type of work		Satisfaction work time		Satisfaction career		Fit education and work		Fit abilities work		Job performance	
	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.	Coef.	Std.
# Children	−0.36	0.95	−0.41	0.94	−0.18	0.87	−0.33	0.93	0.64	0.83	−0.82	0.86	−0.81	0.84
Couple	0.06	0.18	−0.03	0.16	−0.03	0.17	−0.10	0.16	−0.36**	0.16	−0.13	0.15	0.16	0.15
Higher academic	1.24**	0.43	1.14**	0.31	0.51	0.38	0.80**	0.38	1.25**	0.42	1.35**	0.38	−0.79**	0.32
Higher supervisory	1.71**	0.67	1.86**	0.54	1.36*	0.75	1.19**	0.43	0.65	0.68	1.25**	0.63	0.71	0.46
Intermediate academic	0.96**	0.27	1.48**	0.28	0.88**	0.25	1.06**	0.30	1.37**	0.28	1.50**	0.28	0.00	0.26
Intermediate supervisory	0.27	0.30	0.62**	0.30	0.54**	0.25	0.61*	0.31	0.63*	0.34	0.64**	0.37	−0.13	0.30
Other mental work	0.44*	0.26	0.74**	0.29	0.56	0.25	0.56*	0.29	0.82**	0.28	0.89**	0.26	−0.31	0.25
Skilled manual work	1.25**	0.42	1.52**	0.35	0.73**	0.37	0.83**	0.35	1.11*	0.59	0.92*	0.52	−0.47	0.41
Intercept1	−0.47	0.64	−0.03	0.49	−1.82**	0.67	−0.80	0.61	0.58	0.58	0.70	0.58	−2.98**	0.52
Intercept2	−0.34	0.64	0.10	0.50	−1.65**	0.61	−0.56	0.63	0.72	0.59	0.92	0.59	−2.54**	0.52
Intercept3	0.00	0.64	0.27	0.51	−1.27**	0.62	−0.36	0.63	1.02	0.59	1.05	0.59	−2.42**	0.53
Intercept4	0.28	0.64	0.55	0.52	−1.01	0.61	−0.11	0.63	1.10*	0.59	1.18*	0.60	−2.24**	0.52
Intercept5	0.50	0.65	0.68	0.52	−0.90	0.61	0.07	0.62	1.36**	0.60	1.47**	0.60	−1.90**	0.51
Intercept6	0.85	0.65	1.11**	0.52	−0.52	0.62	0.47	0.62	1.59**	0.61	1.72**	0.60	−1.55**	0.50
Intercept7	1.27*	0.65	1.49**	0.52	−0.06	0.61	0.96	0.62	1.89**	0.62	2.14**	0.62	−1.27**	0.49
Intercept8	1.97**	0.65	2.10**	0.54	0.56	0.61	1.67**	0.63	2.30**	0.62	2.68**	0.62	−0.83	0.49
Intercept9	3.01**	0.66	3.10**	0.56	1.34**	0.60	2.66**	0.64	2.92**	0.62	3.48**	0.63	−0.14	0.49
Intercept10	3.75**	0.69	3.76**	0.57	1.99**	0.61	3.28**	0.65	3.42**	0.63	4.16**	0.63	0.49	0.49
Nobs	339		340		341		339		348		348		348	
Likelihood Value	−685.1		−609.7		−646.7		−629.2		−729.0		−688.2		−619.9	

Table 18 Analysis of nonresponse: comparison of sample means

Sample	Observed Waves NT=1303	Missing Waves NT=370			All waves Observed N=295	At least 1 Missing N=215		
	Mean	Mean	Difference	SD diff.	Mean	Mean	Difference	SD diff.
Age	43.0	40.5	2.5**	0.6	42.5	38.7	3.8**	0.9
Female	0.58	0.58	0.002	0.03	0.58	0.60	−0.02	0.04
Single	0.16	0.11	0.05**	0.02	0.17	0.13	0.04	0.03
Couple without children	0.24	0.24	−0.002	0.03	0.24	0.24	0.00	0.04
Couple with children	0.47	0.51	−0.03	0.03	0.48	0.52	−0.04	0.04
Single parent	0.11	0.14	−0.03	0.02	0.09	0.11	−0.02	0.03
Other household	0.02	0.01	0.01	0.01	0.02	0.00	0.02*	0.01
Number of children	1.09	1.28	−0.19**	0.07	1.00	1.17	−0.17*	0.10
Prim. Education	0.12	0.15	−0.03	0.02	0.14	0.14	0.00	0.03
Lower voc./prof.	0.20	0.20	0.00	0.02	0.18	0.20	−0.02	0.04
Higher sec./middle voc.	0.34	0.35	−0.01	0.03	0.35	0.33	0.02	0.04
Higher voc./Univ.	0.34	0.31	0.03	0.03	0.33	0.33	0.00	0.04
Linguistic distance	0.48	0.53	−0.05	0.03	0.48	0.49	−0.02	0.04
Age at migration	18.65	18.41	0.25	0.70	18.47	18.24	0.23	1.0
Genetic distance	0.06	0.06	0.0001	0.003	0.06	0.06	0.001	0.01
Geographic distance	4.85	5.12	−0.26	0.23	4.83	5.02	−0.19	0.35

Table 19 Analysis of nonresponse: comparison of sample means

Dependent variable	s_{it} (Observed Wave)		k_i (All waves observed)	
Variable	Coef.	Std.	Coef.	Std.
Age	0.015**	0.005	0.027**	0.006
Prim. Education	−0.17	0.14	−0.04	0.19
Lower voc./prof	−0.10	0.12	−0.16	0.17
Higher sec./middle voc.o	−0.06	0.11	0.06	0.14
Female	0.04	0.09	−0.01	0.12
Single	−0.38	0.40	−0.77	0.62
Couple without children	−0.65*	0.39	−0.99	0.61
Couple with children	−0.47	0.41	−0.77	0.63
Single parent	−0.55	0.41	−0.90	0.65
Number of children	−0.09	0.06	−0.12	0.09
Linguistic distance	−0.15	0.11	−0.01	0.15
Age at migration	0.001	0.004	−0.003	0.006
Genetic distance	1.17	0.91	1.36	1.14
Geographic distance	−0.03	0.01	−0.03	0.02
Intercept	0.95	0.44	0.19	0.66

Reference categories dummy variables:
Asia; Higher professional/vocational training, university; Single
**/*Significant at 5/10 % level;
Standard errors are robust to correlation in unobserved errors across time for the same individual (clustering)

job, meaning that the migrant fully incorporates language proficiency in forming job satisfaction.

As a simplifying example, suppose that the migrant is educated for a job as an engineer, and in case of full proficiency, $p^*_M(1, x)$ corresponds to a job as an engineer. Then, case (i) refers to the case where the worker can get a job as an engineer, no matter the proficiency level l. Language proficiency thus will not affect job satisfaction. Case (ii) reflects the situation where a lower proficiency level decreases the attainable job level. Suppose that at $l < 1$ the highest attainable job for the engineer is a job as a warehouse employee with job level $p_{JM}(l, x)$. If the migrant fully acknowledges and accepts that due to insufficient proficiency no better job than warehouse employee is attainable, the preferred job level $p^*_J(l, x)$ is adjusted downwards accordingly. Language proficiency thus affects job level but not job satisfaction. Therefore, it is good that in our empirical analysis we both consider job satisfaction and job level as outcomes.

Starting from $l = 1$, lowering language proficiency will have a negative impact on job satisfaction (that is, (7) is satisfied) if the value of the attainable job decreases faster than the value of the optimal job, meaning that language proficiency actually has more impact on the attainable job than is considered ‘fair’ by the migrant. Alternatively, the migrant does not fully perceive the actual impact of language proficiency on job quality. In both cases, the migrant places low emphasis on proficiency l, relative to characteristics x (e.g. educational attainment) in shaping the preferred job. The migrant feels restricted in bridging the gap between the attainable job and the preferred job, leading to a lower level of satisfaction.

Applying the earlier example, where at $l < 1$ the highest attainable job is that of warehouse employee. If the worker does not acknowledge and accept that a decreased level of proficiency leads to a job as a warehouse employee rather than a job as an engineer, at $l < 1$ the preferred job $p^*_M(l, x)$ is still set at the level for an engineer. This leads to lower job satisfaction.

Destination country language proficiency is based on a set of variables z that, among other things, includes the linguistic distance between the destination language l and the language of origin l_o. Age at migration can be part of z, as well as unobserved variables that determine l. There can be variables in x (affecting the job match) that also enter z. For identification, there need to be variables in z that are not in x. These excluded variables affect job match quality only via l. In the empirical section, we discuss in detail which variables will serve as instrumental variables and the issue of confounders.

Unobservables that affect the job match quality (part of x) can be correlated with unobservables affecting l (included in z). The econometric model presented in the Methods section incorporates such correlation.

References

Adsera A, Pytlikova M (2015) The role of language in shaping international migration. Econ J 125:F49–F81

Ashraf Q, Galor O (2013) The Out of Africa hypothesis, human genetic diversity, and comparative Economic development. Am Econ Rev 103(1):1–46

Bakker B, Brown CH, Brown P, Egorov D, Grant A, Holman EW, Mailhammer R, Müller A, Velupillai V, Wichman S (2009) Adding typology to lexicostatistics: a combined approach to language classification. Ling Typol 13(1):169–181

Bleakley H, Chin A (2004) Language skills and earnings: evidence from childhood immigrants. Rev Econ Stat 82(2):481–496

Bound D, Jaeger JA, Baker RM (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. J Am Stat Assoc 90(430):443–450

Chevalier A (2003) Measuring overeducation. Economica 70:509–531

Chiswick BR (2007) The Economics of Language for Immigrants: An Introduction and Overview. Julian Simon Lecture Series, IZA, Bonn

Chiswick BR, Miller PW (1995) The Endogeneity between Language and Earnings: international analyses. J Labor Econ 123(2):246–288

Chiswick BR, Miller PW (2001) A model of Destination-Language acquisition: application to male immigrants in Canada. Demography 38(3):391–409

Chiswick BR, Miller PW (2010) Occupational language requirements and the value of English in the US labor market. J Popul Econ 23:353–372

Dougherty C (2016) Introduction to econometrics, 5th edn. Oxford University Press, Oxford

Dustmann C (1994) Speaking fluency, writing fluency and earnings of migrants. J Popul Econ 7:133–156

Dustmann C, Fabbri F (2003) Language proficiency and labour market performance of immigrants in the UK. Econ J 113(489):695–717

Dustmann C, van Soest A (2001) Language Fluency and Earnings: Estimation with Misclassified Language Indicators. Rev Econ Stat 83(4):663–674

Foged M, Hasager L, Peri G, Arendt JN, Bolvig I (2022) Forthcoming, Language training and refugees' integration. Rev Econ Stat. https://doi.org/10.1162/rest_a_01216

Foged, M, van der Werf C (2022) Access to language training and local integration of refugees, Centre for Research and Analysis of Migration (Cream), Discussion Paper Series CDP 10/22

Gonzalez PS (2010) The impact of education and host language skills on the labor market outcomes of immigrants in Spain. Ivestigaciones de Economia de la Educacion 5:798–824

Hausman JA, Wise DA (1979) Attrition bias in experimental and panel data: the Gary income maintenance experiment. Econometrica 47(2):455–474

Heller BH, Slungaard Mumma K (2022) Immigrant integration in the United States: the role of adult English language training. Economic Policy, American Economic Journal ((**forthcoming**))

Holman EW (2011) Programs for calculating ASJP distance matrices (version 2.1). http://email.eva.mpg.de/~wichmann/software.htm, http://email.eva.mpg.de/~wichmann/listss15.zip

Imai S, Stacey D, Warman C (2016) From engineer to taxi driver? Language proficiency and the occupational skills of immigrants, Working Paper, Ryerson Univesity http://economics.ryerson.ca/workingpapers/wp040.pdf

Isphording IE, Otten S (2014) Linguistic barriers in the destination language acquisition of immigrants. J Econ Behav Organiz 105:30–50

Isphording IE, Sinning M (2013) The returns to language skills in the US labour market. IZA, Bonn

Lochman A, Rapoport H, Speciale B (2019) The effect of language training on immigrants' economic integration: Empirical evidence from France. Eur Econ Rev 113:265–296

Mavromaras K, McGuinnes S, O'Leary N, Sloane P, Wei Z (2013) Job mismatches and labour market outcomes: panel evidence on Australian university graduates. Econ Record 89:382–395

Muthukrishna M, Bell AV, Henrich J, Curtin CM, Gedranovich A, McInerey J, Thue B (2020) Beyond Western, educated, industrial, rich, and democratic (WEIRD) psychology: measuring and mapping scales of cultural and psychological distance. Psychol Sci 31(6):678–701

Scherpenzeel A (2009) Start of the LISS panel: sample and recruitment of probability-based internet panel. CentERdata, Tilburg

Schmid L Forthcoming, the impact of host language proficiency on migrants' employment Outcomes, American Economic Review: Insights

Spolaore E, Wacziarg R (2009) The diffusion of development. Quart J Econ 124(2):469–529

Yao, Y. and J.C.van Ours (2015) Language Skills and Labor Market Performance of Immigrants in the Netherlands, Discussion Paper 2015-010, Labour Economics, 34(C):76-85