Original Research Article

IMR

International Migration Review
1-38

DOI: 10.1177/01979183221107923
journals.sagepub.com/home/mrx
SAGE

Linguistic Barriers to Immigrants' Labor Market Integration in Italy

Daniela Ghio
European Commission Joint Research Centre, Ispra, Italy

Massimiliano Bratti
Università degli Studi di Milano, Milan, Italy

Simona Bignami
Université de Montréal, QC, Canada

Abstract

This article investigates whether and to what extent poor proficiency in Italian impairs immigrants' labor market integration in Italy. Using individual-level survey data, we apply instrumental variables methods to leverage presumably exogenous variations in Italian proficiency induced by immigrants' demo-linguistic characteristics (e.g., age at arrival, linguistic distance between mother tongue and destination language, speaking Italian during childhood) and their interplays. We find that, given the low-skill nature of Italy's immigrant labor market, poor proficiency in communication skills (speaking and understanding Italian) produces larger penalties for immigrants' labor force participation and employment than does the lack of formal skills (reading and writing). In contrast, no effect is found on immigrants' job characteristics like the type of contract and full-time or part-time work. Whereas female immigrants were more penalized than males by poor linguistic proficiency in labor force participation, immigrants in linguistic groups that were more likely to work with (for) co-nationals were less affected by linguistic barriers than other immigrant groups. Yet, when investigating perceived integration outcomes, immigrants working

Corresponding Author:
Daniela Ghio, European Commission Joint Research Centre, Ispra, Italy.
Email: daniela.ghio@ec.europa.eu

with (for) co-nationals fared worse on *feeling at home*, *feeling accepted*, and *overall life satisfaction* in Italy. As our analysis shows, linguistic enclaves in workplaces, while not always representing a hurdle to immigrants' labor market success, can generate trade-offs for other non-labor market integration outcomes. These findings highlight that the development of linguistic skills should be prioritized in migration policy agendas, taking into account heterogeneity in immigrants' demographic and linguistic profiles.

Keywords
labor market integration, linguistic integration

1. Introduction

The literature on immigrant integration has shown how proficiency in the destination language can foster social interactions between immigrants and natives and shape immigrants' incorporation into destination societies (Chiswick and Miller 2001). Immigrants with higher human capital, education, and proficiency in the destination language have greater opportunities for better occupations and earnings (Chiswick and Miller 2003; Van Tubergen and Kalmijn 2009; Heckman and Mosso 2014). The key role played by destination language in immigrant integration explains why, in most European countries, national legislations define linguistic skills as a requirement for immigrants to reside permanently and as a necessary condition for their naturalization.[1] Consequently, understanding how language skills operate among different immigrant groups is essential for national and European policies aiming to enhance immigrants' labor market integration (Zwysen 2019). Yet, findings for countries whose official languages are widely spoken internationally (e.g., the United Kingdom) or where large population groups speak vehicular languages[2] (e.g., English in the Netherlands) are not easily generalizable to countries whose languages are only narrowly spoken abroad (e.g., Italy).

[1] For instance, linguistic proficiency in the destination language is required in Germany, by the Nationality Act of 22 July 1913 (Reich Law Gazette I p. 583 — Federal Law Gazette III 102-1), as last amended by Article 1 of the Act of 12 August 2021 (Federal Law Gazette I p. 3538), in France, by the Loi n. 73-42 du 9 janvier 1973 complétant et modifiant le code de la nationalité française et relative à certaines dispositions concernant la nationalité française, as last amended by the Law of 24 July 2006 and the Law of 16 June 2011, and in Italy, by the Law n.91 of 5 February 1992, as last amended by the D.L. n. 113 of 4 October 2018.

[2] A vehicular language, or *lingua franca*, is a language shared among different native and migrant speakers as a vehicle or bridge for communicating (Coulmas 2005).

This article investigates the effects of immigrants' linguistic proficiency on their labor market integration in Italy. We empirically assess how poor proficiency in Italian affects immigrants' labor market outcomes, using instrumental variable estimation to leverage presumably exogenous variations in linguistic proficiency generated by individual demo-linguistic characteristics (e.g., age at arrival, linguistic distance between immigrants' mother tongues and Italian), and their interplays. We identify Italy as an ideal case for this study for three main reasons. First, Italy reports highly diverse immigrant populations (Vertovec 2007), encompassing more than 195 nationalities (Istat 2019). In contrast to countries with a long history of colonialism (e.g., France), where many immigrants come from former colonies and are exposed to the destination language before migration (Van Tubergen and Kalmijn 2009), the majority of immigrants in Italy are unlikely to know Italian before their arrival: this lack of knowledge can affect their adaptation process into Italian society. Moreover, language proficiency is at the heart of Italian legislation on immigrant integration. Indeed, the *Italian Integration Agreement* establishes that immigrants must demonstrate knowledge of Italian to be eligible for a long-term residence permit, which can be granted only after six years of legal residence in Italy (Italian Presidential Decree no. 179, September 14, 2011).

Second, Italy is characterized by a segmented labor market, where unemployment and labor shortages co-exist (Reyneri 2004). To meet labor market demands, the Italian governments adopt a quota system to recruit economic migrants (Italian Presidential Decree *Decreto Flussi*). However, many economic migrants, trying to enter Italy illegally (Levinson 2005), are attracted by potential earnings in the underground economy (Reyneri 1998). The regularization programs, taking place in Italy since 1986, attest to immigrants' role in the Italian labor market (Levinson 2005). Indeed, with considerable variations across Italian regions, immigrants have become complementary to, and sometimes substitute for, natives as a lower-cost workforce (Reyneri 2004). For instance, female immigrants have replaced Italians in the vast majority of domestic work, particularly in northern and central Italy (Reyneri 2004). Immigrants may also arrive in Italy through channels established by their earlier cohorts of co-nationals, becoming employees in sectors with a favorable ethnic-specific labor profile (Piil Damm 2009). Chiswick and Miller (2005) assert that immigrants may accept lower wages if they find a job in their ethnic enclave, confining immigrants' social interactions to their own origin communities and reducing contacts with natives (Chiswick 1991; Portes 1998). Broadly, according to Bertrand, Luttmer, and Mullainathan (2000), social networks can potentially affect immigrants' labor market behaviors through two main channels: an information channel, which operates through contacts with valuable knowledge (e.g., co-ethnic workers or employers), and a social-norm channel, which influences employment status directly, such as work-ethnic profiling, and indirectly, such as norms imposing female roles (Borjas 1992, 1995, 1998).

The third reason that makes Italy an interesting case study is the availability of rich individual-level survey data on immigrant integration. We use the Italian National Survey on the Social Condition and Integration of Foreign Citizens (SCIF),[3] an ideal data source for studying how proficiency in the Italian language impacts immigrants' labor market integration in Italy. Although the survey was conducted in 2011-2012, SCIF remains the richest and most recent micro-dataset on immigrants' demographic and linguistic characteristics, such as age, sex, age at arrival, birth country, and mother tongue, which are essential to investigate how the effects of individual linguistic skills vary across immigrant mother-tongue groups, leveraging presumably exogenous variations in immigrants' proficiency in Italian (Section 4). The SCIF sample was designed to be representative of immigrant populations living in Italy (Section 3), by birth country and mother tongue. More recent surveys, such as the Italian Labor Force Surveys (LFS), report immigrants' birth country,[4] but not their linguistic proficiency, and the European LFS ad-hoc waves focus on immigrants'[5] proficiency in destination language without detailing immigrants' birth country (Fasani, Frattini, and Minale 2021). Hence, only the SCIF allows us to distinguish between immigrants' proficiency in communication skills (understanding and speaking Italian) and formal language abilities (reading and writing in Italian). Research on language acquisition (Netten and Germain 2012) reveals that knowledge of formal rules and grammar lies in the brain's declarative memory, while skills, such as speaking fluently, reside in procedural memory. This neurolinguistic approach (Netten and Germain 2012), explaining why individuals do not have to acquire formal skills to speak a language, gives the empirical basis to make specific educational plans for classrooms aimed at teaching speaking competencies and at teaching grammar (formal knowledge) required for reading and writing. This distinction in how a foreign language is learned justifies our analytical decision to separately examine the effects of linguistic abilities (understanding and speaking versus reading and writing) on immigrant labor market integration in Italy.

We formulate three hypotheses about the relative importance of communication versus formal linguistic abilities in immigrants' labor market outcomes, and about differences in the impacts of linguistic deficiencies on immigrants' labor market outcomes by linguistic group, and by gender. We put forward that, since in Italy immigrants often fill unskilled jobs (Fullin and Reyneri 2011), communication should be more important than formal language abilities in granting them satisfactory employment outcomes. Yet, immigrants working within their linguistic enclaves

[3] https://www.istat.it/en/archivio/191097

[4] Nevertheless, it should be noted that availability of immigrants' birth country is limited to the restricted-access version of the LFS.

[5] LFS 2014, Labour market situation of migrants and their immediate descendants, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=EU_labour_force_survey_%E2%80%93_data_and_publication#Main_indicators_and_EU-LFS_data

(Carliner 2000), defined here as *workplace-ethnicized linguistic groups* to identify the status of working for a co-national employer or with co-national employees[6] and its related life space,[7] are less likely to be impaired by low proficiency in Italian in terms of labor market outcomes. Female immigrants should be more affected by language proficiency, in terms of their labor market outcomes, compared to men, because they are more likely to work for native employers, typically as housekeepers or caregivers (Barone and Mocetti 2011; Ranci and Arlotti 2019).

These predictions are generally supported by the results of our analysis. We find that the most workplace-ethnicized linguistic group in the Italian labor market is also the least negatively affected, workwise, by poor proficiency in Italian. Although the core of our analysis concerns labor market outcomes, we also present an exploratory analysis of immigrant outcomes that proxy for their perceived integration, or individual feelings of belonging to the host society (e.g., *feeling at home* and *feeling accepted*) and their level of *life satisfaction*. Results show that successful labor market outcomes do not always go hand in hand with other subjective indicators, highlighting how potential trade-offs between economic and social outcomes may exist for immigrants working within linguistic enclaves.

We provide two main contributions to the literature on immigrant integration. First, our analysis makes a clear distinction between the causal effects of communication skills (understanding and speaking Italian) and of formal skills (reading and writing in Italian) on immigrants' labor market outcomes, whereas the majority of existing studies of immigrants' linguistic skills operationalize language proficiency as a unidimensional ability (Van Tubergen and Kalmijn 2009; Heckman and Mosso 2014). By differentiating communication from formal skills, we provide valuable arguments for policy makers, not only in support of investment in educational programs to improve immigrants' employability, but also in highlighting priorities for an effective teaching of destination languages. Although the analysis deals with Italy, our findings are likely to extend to other Southern European countries hosting a majority of low-skilled immigrants, like Greece and Spain (Frattini and Solmone 2022). In these countries, training in the destination language should focus on teaching informal communication skills that are essential for immigrants' labor market integration.

Second, while confirming the important role of immigrants' linguistic skills for their occupational achievements (Chiswick and Miller 2003, 2005), our findings also reveal heterogeneous effects related to linguistic enclaves in workplaces. Although deficiencies in the destination language may not affect successful

[6] Hermansen (2022) has created a visualization tool to capture differences and convergence in workplace dimensions across immigrants, immigrant descendants, and non-immigrant natives.

[7] The life space refers to the locations where and with which individuals interact during their daily life and across their life course (Lelièvre and Robette 2010).

immigrants' labor outcomes, poor language proficiency can lead immigrants toward social marginalization and feelings of exclusion from the hosting societies.

The remainder of this article formulates hypotheses on the role of Italian language proficiency in shaping immigrants' labor market integration in Italy. We, then, describe the data and explain our methodological approach. Subsequent sections discuss results, quantify the effects of Italian proficiency on immigrants' labor market outcomes, and investigate individual perceived outcomes of integration. Finally, we summarize our main findings and draw some conclusions on the role played by the destination language in immigrant incorporation into hosting societies.

2. Research Hypotheses

We define three research hypotheses shaped by certain characteristics of the Italian labor market: two hypotheses are consistent with scholars arguing for the sectoral segmentation of the Italian labor market (Reyneri 2004), whereas the third derives from the feminization of particular occupations, such as those related to home and health care (Ranci and Arlotti 2019).

2.1 Immigrants' Linguistic Abilities and Unskilled Jobs

Using the O*NET dataset and individual US census data, Peri and Sparber (2009) demonstrate that foreign-born workers specialize in occupations that are intensive in terms of manual labor skills, while natives are employed in jobs with communication-intensive tasks. Similar findings are reported for Spain, where Amuedo-Dorantes and de la Rica (2011) demonstrate that proficiency in Spanish plays an important role in shaping immigrants' comparative advantage in manual jobs. Indeed, they also show that Spanish-speaking immigrants tend to be less specialized in manual tasks, due to their comparative advantage in interactive tasks compared to non-Spanish-speaking immigrants. This international evidence could explain the low-skill content of jobs typically done by immigrants in countries such as Italy, whose official language is not widely spoken abroad.

According to Reyneri (2004), in Italy, immigrants tend to concentrate on jobs where working conditions are hard, occupation hazards are high, and no specific formal skills are needed (namely, low-skilled jobs) and are more likely to fill low-skilled jobs that do not require a good command of Italian than natives (Fullin and Reyneri 2011). Based on these considerations, we formulate the following research hypothesis:

> *Hypothesis 1.* In Italy, where there is high labor market demand for low-skilled immigrants, poor communication skills in Italian negatively impact immigrants' labor market outcomes (e.g., labor force status, employment status, job characteristics, and job satisfaction) while poor formal linguistic skills in Italian do not.

2.2 Differences in Immigrants' Labor Market Returns to Linguistic Ability by Ethnic Enclave

According to Portes (1998, 23), "enclaves are dense concentrations of immigrant or ethnic firms that employ a significant proportion of their co-ethnic labor force and develop a distinctive physical presence in urban space." Proficiency in the destination language typically improves immigrants' labor market outcomes (Chiswick 2009; Duleep, Liu, and Regets 2014), whereas working in ethnic enclaves decreases the acquisition of the destination country's human capital (e.g., language) and is associated with limited opportunities in terms of social interactions with natives (Lazear 1999).

The Italian labor market presents several forms of sectoral and ethnic segmentation (Fullin and Reyneri 2011). In Italy, immigrants are mainly employed in specific sectors, such as personal services, agriculture, construction, and transport jobs, where productivity is low and where cheap labor can provide profitable margins for employers (Venturini and Villosio 2006). Using Italian administrative sources, Bratti et al. (2019) show different levels of ethnic concentration by firm: the percentage of workers whose nationality coincides with that of their employer was 89.3 percent for Chinese workers, 32.7 percent for Egyptian workers, and 4.8 percent for Moroccan workers. The existence of these ethnic enclaves can make language skills in the destination language less important for immigrants' labor market outcomes. Liu (2015), for instance, describes Chinatowns around the world to explain the *symbolic diasporic spaces* created by Chinese immigrants in destination countries. Using Liu's metaphor of urban-space aggregation, the largest Chinese community, as a percentage of the resident population in Europe, is hosted by the relatively small town of Prato, located in the Italian region of Tuscany (Raffaetà and Baldassar 2015). Mainly attracted by the textile industry, Chinese immigrants in Prato approached 26,400 registrations in 2020 (+ 1,483 in comparison with 2019), constituting 13.5 percent of the population (Ufficio di Statistica del Comune di Prato 2021). In Prato's industrial district, many Chinese migrants have been able to upscale to the position of manufacturer since the mid-2000s, achieving ownership in the local clothing industry and predominately employing co-nationals (Ceccagno 2015).[8]

We, therefore, put forward a second hypothesis that builds on the ethnicization of some industries and activities in the Italian economy, considering that immigrant entrepreneurs often prefer to hire immigrant workers of their same origin[9] to

[8] There is an extensive literature on Chinese enclaves and labor market outcomes (e.g., Zhou and Logan 1989; Li and Li 2016).

[9] It should be noted that here, the term "origin" is used as generic reference that embraces both immigrants' origin country and mother tongue. For instance, China has more than 50 ethnic minorities, each with its own languages (Dong and Blommaert 2009). Therefore, *Chinese* is used as an umbrella term to define immigrants belonging to the Chinese linguistic group, irrespective of the dialect/language spoken in their birth region.

create labor market enclaves in which proficiency in the destination language may not be important for labor market outcomes. Thus, we argue:

Hypothesis 2. Due to the existence of linguistic enclaves in the Italian labor market, the more ethnicized workplaces are, the less crucial proficiency in Italian is for immigrants' labor market outcomes.

2.3. Differences in Immigrants' Labor Market Returns to Linguistic Ability by Gender

Finally, domestic work in Italy remains strongly characterized by the employment of female immigrants (Reyneri 2004). Two main factors could explain this female dominance: the lack of public services to cover families' needs and the traditionally gendered division of labor which relegates women to household labor (Andall 2000). The domestic sector's importance in Italy clearly emerged in 2002, when more than 300,000 families benefitted from the regularization of previously unregistered immigrant women dedicated to house and elder care (Reyneri 2004). The female characterization of the household/domestic sector (Barone and Mocetti 2011; Ranci and Arlotti 2019) could make proficiency in Italian more relevant for female immigrants than for male immigrants, as the former are more likely to interact with native employers. Thus, our third hypothesis is the following:

Hypothesis 3. Proficiency in the destination language is more important for the labor market integration of immigrant women than immigrant men, since immigrant women are likely to work in jobs that entail more contacts and interactions with natives.

3. Data and Variables

Our empirical analysis is based on the SCIF carried out by ISTAT in 2011-2012, which is the richest and most recent data source on the social conditions of immigrants residing in Italy. As discussed in the Introduction, more recent data sources, such as the National and European LFS, are less detailed than the SCIF, which reports both demographic and linguistic characteristics of immigrants in Italy. For the SCIF, immigrants were defined as individuals with a migration background, meaning those born in Italy or born abroad to non-Italian parents. The SCIF sample included more than 25,000 respondents,[10] of whom 21,030 were immigrants

[10] Sample design is based on families living in Italy with at least one member who was a foreigner or naturalized citizen; for children under 14 years old and temporarily unavailable persons, proxy replies were collected by other family members. Families are classified as de facto and associated with a unique identification or family code in the population registers, when persons had their usual residence in the same household and were linked by marriage, kinship, affinity, adoption, protection, or affection relationship. To classify de facto families, belonging to the same household supersedes administrative association through the population registers.

without Italian citizenship (17,545 first-generation immigrants and 2,834 second-generation immigrants) and 4,010 were born in Italy (to parents born abroad) and had Italian citizenship. Immigrants were interviewed, applying the computer-assisted personal interviewing (CAPI) technique, with the questionnaire translated into a language they mastered.[11]

Along with the scope of our analysis, we selected immigrants aged 15 to 64, as this group represents the economically active age group (6,594 observations are dropped from the sample). This restriction allows us to focus on the populations that mainly benefit from linguistic abilities in terms of future incorporation into host societies' labor markets (Chiswick, Lee, and Miller 2002). Moreover, since students and second-generation immigrants are likely to acquire proficiency in the destination language (Lenneberg 1967), we only retained individuals who declared not being students and who were not born in Italy. This further selection criterion reduced the sample by 3,458 observations. In all empirical models, only observations for which the dependent, independent, and control variables were non-missing are retained in the estimation (the exact number of observations is reported in the tables when results are presented).

The SCIF questions related to linguistic ability were formulated as follows:

Do you have difficulties understanding Italian?

Do you have difficulties speaking in Italian?

Do you have difficulties reading in Italian?

Do you have difficulties writing in Italian?

Respondents had to choose one of four possible answers: no difficulty, few difficulties, quite a lot of difficulty, or a lot of difficulty. We use an indicator for each type of linguistic ability, defined as a dichotomous variable that equaled one for immigrants who reported *a lot* or *quite a lot* of difficulty and zero for immigrants who reported having *few difficulties* or *no difficulty*. We refer to this variable as *poor proficiency* in Italian for brevity. Like the variables used in most previous studies (e.g., Bleakley and Chin 2004, 2010; Miranda and Zhu 2013, 2021; Yao and van Ours 2015; Guven and Islam 2015), this indicator may reflect some individual subjectivity in

[11] The survey questionnaire was translated into 10 languages (English, French, Spanish, German, Romanian, Albanian, Polish, Russian, Arabic, and Chinese) that represent mother tongues spoken by the vast majority of immigrants or vehicular languages that immigrants are likely to understand.

the response and, since it is self-assessed, can be affected by measurement error, as is discussed in Section 4.

We refer to speaking and understanding as communication or informal[12] linguistic skills, in opposition to reading and writing, which are defined as formal skills. Supplemental Table A1 reports the matrix of correlations among linguistic poor proficiency indicators. As expected, although poor proficiency across all linguistic abilities is positively correlated, within-correlation (within communication and formal skills) is higher than between-correlation (between communication and formal skills). For instance, the correlation between speaking and understanding poor proficiency is 0.87, whereas that between reading and writing poor proficiency is 0.76; in contrast, the correlation between understanding and writing poor proficiency is much lower at 0.46. Supplemental Table A2 shows the results of factor analysis using the principal-factor method to analyze the correlation matrix. The factor analysis identifies two factors. In the first, all variables are weighted positively and similarly. We interpret this result as individuals with a good (bad) command of the Italian language to be likely to perform well (poorly) in every language skill. In the second factor, reading and writing enter positively and speaking and understanding negatively. This second factor is more strictly related to our classification of skills into communication and formal skills. Overall, Supplemental Tables A1 and A2 support our choice of modeling poor proficiency in communication and formal linguistic skills separately.

Supplemental Table A3 shows the levels of language proficiency by skill and linguistic group. Distinguishing by language skill, immigrants report having higher communication skills (understanding and speaking) than other linguistic abilities (reading and writing) on average. The poorest linguistic abilities were reported by Chinese-speaking immigrants (68, 72, 50, and 50 percent of respondents in this group reported having reading, writing, speaking, and understanding difficulties, respectively). Linguistic proficiency was higher among Portuguese, Spanish, and Romanian mother-tongue immigrants, languages that exhibit shorter linguistic distances from Italian.[13]

To assess immigrants' labor market outcomes, we use the following indicators: labor force participation (a dichotomous variable where the value of 1 corresponds to being in the labor force); perceived employment status (a dichotomous variable where the value of 1 corresponds to unemployment status); current work status (a

[12] This classification is based on the type of learning through which mastering these skills takes place: formal (e.g., classes or self-study) or informal (e.g., daily interactions with natives).

[13] Linguistic distance is a measure of the difference between the origin and destination languages. An index of linguistic distance was developed by Chiswick and Miller (1996). Hart-Gonzalez and Lindemann (1993) introduced a set of language-learning scores (a low value of the score indicates a high level of difficulty to learn a foreign language), and linguistic distance is the reciprocal of the language score.

dichotomous variable where the value of 1 corresponds to working during the last week and 0 means not working but having searched for a job in the last four weeks); working under an open-ended or fixed-term contract (a dichotomous variable where the value of 1 corresponds to working under an open-ended contract and 0 means working under a fixed-term contract); working full- or part-time (a dichotomous variable where the value of 1 corresponds to working full-time and 0 to working part-time); level of job satisfaction on a scale from 0 to 10, which is the highest level of satisfaction; occupational socio-economic prestige, which is described in Section 5.[14]

Although this article focuses on labor market outcomes, proxies of immigrants' perceived integration, such as feeling part of the host society, and life satisfaction are investigated as additional outcomes. Specifically, the SCIF gathers answers to the following questions:

> How much do you feel accepted in the city where you live?
>
> Some foreigners interviewed for this same survey have stated that they feel at home in Italy. Do you agree with this statement?

Answers are expressed, using a Likert scale: *a lot, enough, little, not at all* for the first question and *totally agree, more agree than disagree, more disagree than agree, totally disagree* for the second. In both cases, we dichotomized the answers by pooling the first two categories.[15] The SCIF also provides a measure of overall life satisfaction that, like job satisfaction, ranges from 0 (no satisfaction) to 10 as the highest level of satisfaction.

4. Causal Effects of Linguistic Skills

Destination-language proficiency is key to successful immigrant incorporation into a society: without sufficient linguistic proficiency, other individual capabilities may be irrelevant (Chiswick and Miller 1992). Learning a language is an investment in human capital connected to demographic characteristics, such as age and gender, but also to personal motivation, effort, and individual ability (Berman, Lang, and Siniver 2000). Following Chiswick and Miller (1996), several economic studies have attempted to tackle the role of language in immigrant integration in an English-speaking country (e.g., Bleakley and Chin 2004, 2010; Miranda and Zhu 2013, 2021). Yao and van Ours (2015) clustered immigrant language proficiency by mother tongue to assess heterogeneity effects on

[14] Details on wages and earnings are not available in the SCIF survey.

[15] Answers are dichotomized because we use instrumental variables estimation (see the next section) and because considering separate categories would imply finding separate instruments for each category, which is unfeasible with the data at hand.

labor market outcomes. Guven and Islam (2015) examined the causal effects of language proficiency on economic integration in Australia. Confirming a positive effect on wages, they also found that proficiency in English had increased health benefits for female immigrants and decreased a male-dominant gender view in marriage relationships.

Several studies adopt an instrumental variables approach based on the interaction between age at arrival and other demographic variables, including knowledge of the destination language or its proximity to immigrants' mother tongue. For instance, Bacolod and Rangel (2017) apply a multi-stage model of childhood to assert that age at arrival and linguistic distance to English interact in the skill accumulation process of child migrants during adulthood. Using Norwegian administrative registries, Hermansen (2017) demonstrates that the timing of childhood immigration affects immigrants' later-life outcomes. Nevertheless, these authors have mainly considered linguistic ability as a unidimensional indicator. For example, Geurts and Lubbers (2019) do not disaggregate their analysis by type of linguistic skill but, rather, construct a Likert scale to measure the mean of reported scores across all skills. Similarly, Raijman, Semyonov, and Gffen (2015) examine immigrant groups' linguistic proficiency by combining all skills to construct an index. In contrast, we model each linguistic ability separately to measure its potentially different impact on labor market integration.

When assessing the effects of destination-language proficiency on immigrants' labor market integration, three key empirical issues should be clarified. First, there might be individual characteristics that simultaneously affect immigrants' integration outcomes and language acquisition (e.g., individuals' cognitive abilities measured by standardized test scores, which are rarely available in datasets). The omission of these variables from econometric models could generate an *endogeneity* problem and bias the Ordinary Least Squares regression (OLS) estimates. Attempts to address this *endogeneity* issue have used Instrumental Variables (IV) estimators. Some scholars have leveraged variation in immigrants' origin countries — namely, between countries where the destination language is an official language (typically former colonies) and countries where the destination language is not an official language (e.g., Guven and Islam 2015; Miranda and Zhu 2021). Other scholars have exploited whether individuals spoke the destination language during childhood while residing in their home country (e.g., Yao and van Ours 2015). There is a rationale behind the use of these instruments. Research in psychology has demonstrated that young children learn languages more easily than older children and adults: maturational changes that start before puberty (*the critical age*) reduce a child's ability to acquire a second language (Lenneberg 1967). When exposure to a second language begins before the critical age, a child is able to acquire a level of proficiency that could be almost the same as a native speaker; in contrast, after the critical age, a child's proficiency in a second language would be significantly lower (Lenneberg 1967).

Second, an important hurdle for identifying the effects of destination-language proficiency on immigrant integration is that in survey data, linguistic skills are typically

self-assessed and self-reported: self-reported measures of linguistic proficiency are likely to be affected by severe *measurement error* that biases OLS estimates toward zero (Bleakley and Chin 2004; Clarke and Isphording 2017). Even worse, measurement error could be correlated with the true language proficiency, causing a non-classical measurement error. For instance, a non-classical measurement error may generate the paradoxical situation in which immigrants who are more aware of their linguistic difficulties are also those who have more contact with natives and, consequently, are relatively more integrated. Those who do not use or rarely use the destination language, perhaps because they work or live within an ethnic enclave, may not be fully aware of their language problems. This lack of awareness, in turn, may create a spurious positive correlation between poor language proficiency and integration outcomes. IV estimation represents a solution to this measurement-error problem (see, Miranda and Zhu 2021 for a thorough discussion of this point).

Third, there could be potential *reverse causality* problems. Indeed, more integrated migrants are likely to live and work closer to natives and, through work and social interactions, acquire a better command of the destination language (Chiswick and Miller 2003). The IV approach helps address this issue through the selection of appropriate instruments based on variables that precede immigrants' arrival in the destination country.

4.1. Method

We follow a three-step estimation strategy suggested by Wooldridge (2002, 623) to deal with the above-described endogeneity, measurement error, and reverse causality issues. As a first step, we estimate a probit model for poor language proficiency, measured as dichotomous indicators for having a lot/quite a lot of difficulty with Italian (separately for reading, writing, speaking, and understanding) on a set of exogenous variables and some *excluded variables* (i.e., variables that we assume explain language proficiency but are not correlated with unobservable individual traits affecting labor market integration). The probit equation reads as follows:

$$\Pr(D_i = 1) = \Phi(\alpha_0 + \alpha_1 Z_i + \alpha_2 X_i), \tag{1}$$

where D_i is the variable for an individual having poor language proficiency in a given domain (writing, reading, speaking, understanding); X_i is a vector of control variables; Z_i is a vector of variables that are excluded from the integration outcome equations; $\Phi(.)$ is the normal distribution function. We include the following variables in the vector Z_i:

- the number of years the individual stayed in Italy after the age of 11. Based on Bleakley and Chin (2004), we use the maximum value obtained when comparing the age at arrival reduced by 11 years and the value 0, hereafter indicated as the *modified age at arrival*;
- the interaction of *modified age at arrival* with *speaking Italian during childhood* (see, Yao and van Ours 2015);

- *linguistic distance* measured through the Automated Similarity Judgment Program (ASJP), as suggested by Isphording and Otten (2014). As they state, "The basic idea behind the ASJP is the automatic comparison of the pronunciation of words across languages. A more similar pronunciation proxies the number of cognates, word pairs between languages with common ancestors, which then again indicates a closer relation between the languages" (p. 34). When the origin language is more similar to the destination language, immigrants should be able to master the latter more easily; thus, linguistic proximity is here computed between immigrants' mother tongue and Italian;
- the interaction of *modified age at arrival* with *linguistic distance*. Clarke and Isphording (2017) give the rationale for using this interaction as a driver of language acquisition: because children learn languages more easily than adults, linguistic distance should limit the attainment of destination-language proficiency for immigrants who arrived at older ages.

Graphical support for the *critical age theory* is given in Figure 1, where we plot the average reported poor language proficiency by age at arrival, a linear trend fitting the age bins, and 95% confidence intervals. The graphs show that although a small discontinuity in poor language proficiency occurs at age 11, the change in slope before versus after age 11 is consistent with the *critical age theory*. Indeed, before age 11, the linear trend in age is almost flat, especially for informal skills (i.e., individuals arriving in Italy between ages 0 and 11 are almost indistinguishable in terms of language proficiency).[16] We refer to equation (1) as the auxiliary probit model or probit model for brevity.

After estimating the probit model, the predicted probability of having language difficulties $\Pr(\widehat{D_i = 1})$ is used as an *excluded instrument* in a two-stage least squares (2SLS) estimator. Thus, the first stage of 2SLS (i.e., the second step of our procedure) reads as

$$D_i = \gamma_0 + \gamma_1 \Pr(\widehat{D_i = 1}) + \gamma_2 X_i + C_i + \epsilon_i, \tag{2}$$

where ε_i is an idiosyncratic error term and C_i represents birth country fixed effects. The second stage of 2SLS (i.e., the third and last step of the procedure described in Wooldridge 2002) reads as

$$Y_i = \beta_0 + \beta_1 \hat{D}_i + \beta_2 X_i + C_i + u_i, \tag{3}$$

where Y_i is the outcome of interest and u_i is an idiosyncratic error term.

The above-defined estimator has a number of desirable properties (Wooldridge 2002). The first-step auxiliary probit model is just used to create an excluded instrument for the 2SLS and does not need to be correctly specified (unlike what happens

[16] The graphs show the average language difficulty by age bin and a linear trend fit with the 95 percent confidence interval obtained using the Stata command *rdplot* (Calonico et al. 2017) with a triangular kernel and sampling weights.

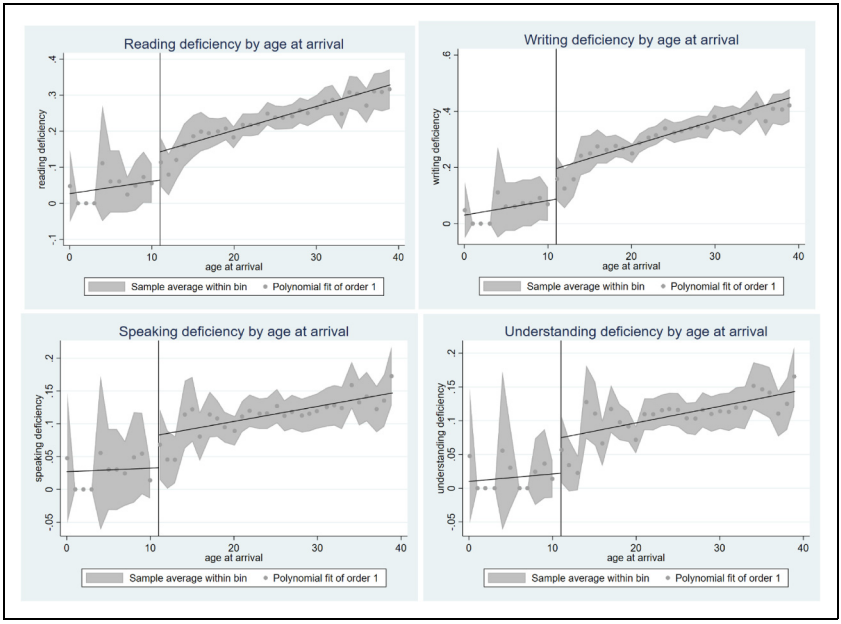

Figure 1. Language Deficiencies and Age at Arrival (Cut-Off at Age 11).
Note: This figure shows the average proportion of immigrants reporting language difficulties by age of arrival bins along with linear regression fits (they are estimated using the Stata package *rdplot*, see, Calonico et al. 2017). The vertical line is set at age 11 in accordance with the critical age theory.

when the integration and language outcomes are jointly modeled with a bivariate probit model). The fact that the instrument $\Pr(\widehat{D_i} = 1)$ is a generated regressor can be neglected in computing the standard errors of the estimates of equation (3), since the 2SLS procedure takes care of it. Given the endogenous variables' dichotomous nature, we define poor linguistic proficiency, using the prediction from the probit model as an instrument, which generally increases the efficiency of the 2SLS estimator compared to using excluded variables from the probit model directly as instruments.[17]

Although the 2SLS estimates are formally identified without exclusion restrictions in the first-step probit model (equation (1)), thanks to the nonlinearity of the probit prediction in the regressors, it is, nonetheless, convenient to have some excluded variables Z_i to avoid the identification solely relying on functional-form assumptions (Wooldridge 2002). The main identification assumptions are similar to those presented by Miranda and Zhu (2021) — namely, that all variables in the vector Z_i, conditional on the

[17] For examples of the application of this procedure, see, Egger, Koethenbuerger, and Smart (2010) and Alcaraz Pribaz, Chiquiar, and Salcedo (2012).

control variables, only affect integration outcomes through poor linguistic proficiency and are not correlated with measurement error in the latter.

4.2. Empirical Strategy

This section details how we operationalize our research hypotheses. The assumption that the excluded instrument Z_i only affects integration outcomes through poor linguistic proficiency, which has been extensively adopted in the literature studying the effects of destination-language proficiency on immigrant integration (Yao and van Ours 2015; Miranda and Zhu 2021), is maintained by including age at arrival, origin country fixed effects, and the non-interacted indicator for speaking Italian during childhood as control variables in equation (3).

As for the assumption that the excluded instrument Z_i is not correlated with measurement error in the survey's measure of poor proficiency in the destination language, since the survey was conducted in a language fully understood by immigrants (Section 3), there is no reason that the measurement error should be correlated with immigrants' linguistic difficulties. Therefore, measurement error is unlikely to be correlated with any excluded variable we adopt (interactions between linguistic distance or speaking Italian during childhood and the critical age). Specifically, the vector X_i consists of the following control variables:

- *Demographic variables:* gender, age, and age squared;
- *Pre-migration variables:* indicators for speaking Italian during childhood and having ever worked in the origin country (as an indicator of long-term labor market attachment);
- *Arrival and current variables:* age at arrival and age at arrival squared, current macro-area of residence in Italy (Northeast, Northwest, Center, South, and Islands), education level, father's education level,[18] an indicator for education completed in Italy, knowledge of Italian at arrival, expected length of stay in Italy (temporary or no plan versus permanent), and an indicator for economic migrants.

Sample summary statistics for the dependent, independent, and control variables are illustrated in Supplemental Table A4.

Furthermore, we set up two indicators of workplace ethnicization, based on two SCIF questions about the nationalities of colleagues (asked to employees) or employees (asked to employerses), respectively. According to whether *all colleagues* (or employees) *were from the same origin country* or *most colleagues* (or employees) *were from the same origin country* as the respondent, we define dichotomous indicators of high workplace ethnicization. We run a regression analysis of these indicators, using the model described in Section 4.1 by linguistic ability. We do not attribute a causal interpretation to these OLS estimates because our aim is to assess differences across linguistic groups that

[18] Mother's education level is not available in SCIF data.

are orthogonal to individuals' demo-linguistic characteristics. As shown by column (1) of Supplemental Table A5, the Chinese linguistic group presents stronger workplace-ethnicization patterns: while, on average, immigrants had a low probability of working in ethnicized workplaces, a striking 80 percent of Chinese-speaking immigrants declared working with co-workers who were mostly or all from China. This result marks a gap between Chinese and other linguistic groups, ranging between −47 percentage points (pp) among Portuguese-speaking immigrants and −61 pp among Romanian-speaking immigrants. Estimates of workplace ethnicization based on employers' answers, in column (2) of Supplemental Table A5, confirm that with the exception of Indo-Iranian employers, foreign employers hired fewer workers from their origin country than did Chinese-speaking employers. On the basis of this evidence, in the analysis estimating differential effects by linguistic group, we group Chinese speakers versus all other linguistic groups.

5. Poor Proficiency in Italian and Labor Market Integration

Table 1 presents the results of the models that test our first research hypothesis concerning the impact of formal linguistic skills on immigrants' labor market integration.

Each panel of Table 1 is denoted by a letter and reports the coefficient on poor proficiency in Italian computed in the second stage of 2SLS for each dependent variable indicated in the first column (with the standard error and level of statistical significance). Panels indicate the Chi^2 statistics for the excluded variables in the auxiliary probit model (used to build the excluded instrument in the first stage of 2SLS), the first stage F-statistics of the excluded instrument (the prediction from the probit model), and the p-value. For brevity, given the high number of models estimated, only model diagnostics are reported in Table 1 (and in later tables), while the first stage of the 2SLS and the auxiliary probit regressions, as well as the coefficients on the control variables, are omitted from the tables (available upon request). Each regression includes the control variables listed in Section 4.1 and those mentioned in each table's footnotes.

Overall, the *excluded variables* used for model identification (e.g., modified age at arrival, the interaction between speaking Italian during childhood and modified age at arrival, linguistic distance, and the interaction between modified age of arrival and linguistic distance) are always statistically significant (at the 1 percent level) in the probit models of poor linguistic proficiency, with Chi^2 statistics between 10 and 25.[19] The excluded instrument used in 2SLS is also highly relevant; the F-statistic is always above 10 and does not show any indication of a weak-instrument problem (Stock and Yogo 2005). In what follows, we comment on the main results.

[19] The auxiliary probit model is estimated on the largest possible sample with non-missing dependent, control, and independent variables. Thus, the Chi^2 statistic does not change across panels and is only reported for the first panel.

Table 1. Poor Proficiency in Italian and Labor Market Integration.

Regressions (Dependent Variable)	Poor Proficiency in							
	Reading		Writing		Speaking		Understanding	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
(a) Labor force participation (LFP)	−0.389***	(0.067)	−0.312***	(0.063)	−0.595***	(0.070)	−0.645***	(0.088)
Chi^2 probit model	11.985		10.489		17.863		25.143	
p-value	0.017		0.000		0.001		0.000	
First-stage *F*-statistic	83.115		130.991		147.600		82.532	
p-value	0.000		0.033		0.000		0.000	
No. of obs.	13,724		13,724		13,724		13,724	
(b) Perceived employment status (employed)	−0.175**	(0.079)	−0.117	(0.092)	−0.349***	(0.109)	−0.398***	(0.121)
First-stage *F*-statistic	68.562		134.380		44.608		25.342	
p-value	0.000		0.033		0.000		0.000	
No. of obs.	11,166		11,166		11,166		11,166	
(c) Current working status (working)	−0.200**	(0.076)	−0.129	(0.083)	−0.391***	(0.107)	−0.438***	(0.118)
First-stage *F*-statistic	75.087		113.343		77.103		37.395	
p-value	0.000		0.033		0.000		0.000	
No. of obs.	11,565		11,565		11,565		11,565	
(d) Open-ended contract versus fixed-term	−0.117	(0.113)	−0.077	(0.096)	−0.238	(0.298)	−0.259	(0.297)
First-stage *F*-statistic	34.600		57.981		31.786		12.424	
p-value	0.000		0.033		0.000		0.000	
No. of obs.	8,394		8,394		8,394		8,394	
(e) Full-time versus part-time work	−0.009	(0.100)	−0.009	(0.095)	−0.149	(0.233)	−0.134	(0.234)
First-stage *F*-statistic	46.848		92.932		32.535		14.728	
p-value	0.000		0.033		0.000		0.000	
No. of obs.	9,672		9,672		9,672		9,672	
(f) Job satisfaction (0–10 scale)	−0.980*	(0.502)	−0.886*	(0.493)	−1.252	(1.043)	−1.761	(1.170)

(continued)

Table 1. (continued)

Regressions (Dependent Variable)	Poor Proficiency in							
	Reading		Writing		Speaking		Understanding	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
First-stage *F*-statistic	51.255		94.734		34.181		15.462	
p-value	0.000		0.033		0.000		0.000	
No. of obs.	9,672		9,672		9,672		9,672	
(g) ISEI	20.527***	(5.161)	19.120***	(5.844)	22.907***	(6.665)	21.820***	(7.920)
First-stage *F*-statistic	46.848		92.932		32.535		14.728	
p-value	0.000		0.000		0.001		0.000	
No. of obs.	9,672		9,672		9,672		9,672	

Notes. Each panel represents a separate regression for the dependent variables in Italics in the first column. The table reports the second-stage 2SLS coefficients for each linguistic skill on the outcome variable, the standard error (s.e.), and significance level. Chi^2 refers to the excluded variables in the auxiliary probit model (modified age at arrival, the interaction between speaking Italian during childhood and modified age at arrival, linguistic distance, and the interaction between modified age at arrival and linguistic distance; see Section 4.1), and it is only reported for the first dependent variable since it does not vary across panels (the probit model is estimated on the largest sample with non-missing dependent and independent variables). The first-stage *F*-statistic refers to the predicted probability of linguistic deficiency used in the first stage of 2SLS as the excluded instrument. All models include the control variables in Section 4.1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors clustered by mother tongue are in parentheses.

Table 2. Poor Proficiency in Italian and Labor Market Integration by Linguistic Group.

Regressions (Dependent Variable)	Poor Proficiency in															
	Reading				Writing				Speaking				Understanding			
	Baseline		Chinese		Baseline		Chinese		Baseline		Chinese		Baseline		Chinese	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
(a) Labor force participation (LFP)	−0.389***	(0.069)	0.019	(0.040)	−0.305***	(0.065)	−0.039	(0.046)	−0.616***	(0.073)	0.274***	(0.052)	−0.681***	(0.092)	0.361***	(0.063)
Chi^2 probit model	11.63				9.404				17.36				25.07			
p-value	0.000				0.000				0.000				0.000			
First-stage *F*-statistic	144.2		23093		216.3		51334		702.7		363156		855.5		573383	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
No. of obs.	13,724				13,724				13,724				13,724			
(b) Perceived employment status	−0.186**	(0.084)	0.129***	(0.040)	−0.113	(0.096)	−0.009	(0.039)	−0.404***	(0.118)	0.459***	(0.078)	−0.480***	(0.129)	0.501***	(0.097)
First-stage *F*-statistic	140.3		4313		149.4		20686		364.5		126390		813.9		290690	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
No. of obs.	11,166				11,166				11,166				11,166			
(c) Current working status	−0.200**	(0.079)	0.022	(0.041)	−0.115	(0.085)	−0.089**	(0.036)	−0.433***	(0.117)	0.388***	(0.078)	−0.504***	(0.128)	0.435***	(0.096)
First-stage *F*-statistic	142.4		16300		155.3		24743		560		141867		1203		261829	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
No. of obs.	11,565				11,565				11,565				11,565			
(d) Open-ended contract versus fixed-term	−0.117	(0.118)	0.041	(0.074)	−0.075	(0.099)	0.042	(0.049)	−0.246	(0.338)	0.139	(0.273)	−0.301	(0.360)	0.229	(0.311)
First-stage *F*-statistic	105.8		90951		93.69		105107		353.5		52930		902		177209	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
No. of obs.	8,394				8,394				8,394				8,394			
(e) Full-time versus part-time work	−0.033	(0.104)	0.200***	(0.073)	−0.032	(0.098)	0.163**	(0.066)	−0.253	(0.260)	0.425**	(0.196)	−0.264	(0.294)	0.411*	(0.238)
First-stage *F*-statistic	116.1		1776		107.2		15058		500.6		29550		891.5		136283	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	

(continued)

Table 2. (continued)

Regressions (Dependent Variable)	Poor Proficiency in															
	Reading				Writing				Speaking				Understanding			
	Baseline		Chinese		Baseline		Chinese		Baseline		Chinese		Baseline		Chinese	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
No. of obs.	9,672				9,672				9,672				9,672			
(f) Job satisfaction (0–10 scale)	−1.294***	(0.479)	2.690***	(0.351)	−1.245***	(0.425)	2.475***	(0.259)	−2.176**	(0.994)	3.876***	(0.797)	−3.056***	(1.045)	4.250***	(0.892)
First-stage *F*-statistic	124.1		2817		115		14435		483.4		14621		884.4		93838	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
No. of obs.	9,589				9,589				9,589				9,589			
(g) ISEI	20.251***	(5.312)	3.175	(2.317)	18.262***	(5.945)	6.712***	(1.704)	24.895***	(7.839)	−7.609	(5.526)	25.009**	(10.013)	−9.757	(7.633)
First-stage *F*-statistic	116.1		1776		107.2		15058		500.6		29550		891.5		136283	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
No. of obs	9,672				9,672				9,672				9,672			

Notes. Each panel represents a separate regression for the dependent variables in Italics in the first column. The table reports the second-stage 2SLS coefficients for each linguistic skill on the outcome variable, the standard error, (s.e.), and significance level. Chi^2 refers to the excluded variables in the auxiliary probit model (modified age at arrival, the interaction between speaking Italian during childhood and modified age at arrival, linguistic distance, and the interaction between modified age at arrival and linguistic distance; see Section 4.1), and it is only reported for the first dependent variable since it does not vary across panels (the probit model is estimated on the largest sample with non-missing dependent and independent variables). The first-stage *F*-statistic refers to the predicted probability of linguistic deficiency used in the first stage of 2SLS as the excluded instrument. Since the model includes both the baseline effect for all linguistic groups and the interaction with the Chinese linguistic group, the model includes two endogenous variables and two first-stage statistics are reported. All models include the control variables in Section 4.1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors clustered by mother tongue are in parentheses.

Table 3. Poor Proficiency in Italian and Labor Market Integration by Gender.

Regressions (Dependent Variable)	Poor Proficiency in															
	Reading				Writing				Speaking				Understanding			
	Baseline		Female		Baseline		Female		Baseline		Female		Baseline		Female	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
(a) Labor force participation (LFP)	−0.156	(0.096)	−0.383***	(0.069)	−0.161**	(0.077)	−0.288***	(0.088)	−0.050	(0.132)	−0.649***	(0.105)	−0.085	(0.148)	−0.687***	(0.111)
Chi^2 probit model	11.631				9.404				17.360				25.065			
p-value	0.020				0.052				0.000				0.000			
First-stage *F*-statistic	103.558		40.806		71.291		86.626		85.212		129.388		47.274		69.091	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
No. of obs.	13,724				13,724				13,724				13,724			
(b) Perceived employment status	−0.194**	(0.088)	0.075	(0.133)	−0.138	(0.099)	0.093	(0.108)	−0.365***	(0.127)	0.068	(0.213)	−0.410***	(0.145)	0.059	(0.243)
First-stage *F*-statistic	69.955		15.209		64.960		63.060		22.976		12.798		13.543		10.687	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
No. of obs.	11,166				11,166				11,166				11,166			
(c) Current working status (working)	−0.171*	(0.096)	−0.081	(0.132)	−0.115	(0.098)	−0.031	(0.115)	−0.269**	(0.130)	−0.216	(0.208)	−0.316**	(0.140)	−0.245	(0.239)
First-stage *F*-statistic	70.394		19.403		54.367		64.799		41.497		26.992		21.952		17.493	
p-value	0.000		0.000		0.052		0.000		0.002		0.000		0.000		0.000	
No. of obs.	11,565				11,565				11,565				11,565			
(d) Open-ended contract versus fixed-term	−0.101	(0.103)	−0.053	(0.089)	−0.070	(0.097)	−0.001	(0.066)	−0.121	(0.257)	−0.332	(0.216)	−0.148	(0.259)	−0.343	(0.227)
First-stage *F*-statistic	31.695		15.559		30.360		74.424		17.978		9.390		6.918		6.727	
p-value	0.000		0.000		0.000		0.000		0.000		0.002		0.002		0.000	
No. of obs.	8,394				8,394				8,394				8,394			

(continued)

Table 3. (continued)

Regressions (Dependent Variable)	Poor Proficiency in															
	Reading				Writing				Speaking				Understanding			
	Baseline		Female		Baseline		Female		Baseline		Female		Baseline		Female	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
(e) Full-time versus part-time work	−0.009	(0.105)	0.005	(0.085)	−0.003	(0.101)	−0.015	(0.068)	−0.103	(0.213)	−0.122	(0.231)	−0.112	(0.217)	−0.064	(0.226)
First-stage *F*-statistic	47.973		10.764		44.531		43.623		20.001		5.210		8.989		4.203	
p-value	0.000		0.000		0.052		0.000		0.008		0.000		0.018		0.000	
No. of obs.	9,672				9,672				9,672				9,672			
(f) Job satisfaction (0–10 scale)	−1.037**	(0.478)	0.255	(0.541)	−0.955*	(0.490)	0.306	(0.415)	23.524***	(6.781)	−1.637	(4.871)	22.255***	(8.109)	−1.323	(4.806)
First-stage *F*-statistic	49.611		10.804		45.619		41.611		20.001		5.210		8.989		4.203	
p-value	0.000		0.000		0.000		0.052		0.000		0.008		0.018		0.000	
No. of obs.	9,589				9,589				9,672				9,672			
(g) ISEI	21.180***	(5.139)	−2.184	(2.874)	20.520***	(6.149)	−5.154*	(2.624)	23.524***	(6.781)	−1.637	(4.871)	22.255***	(8.109)	−1.323	(4.806)
First-stage *F*-statistic	47.973		10.764		44.531		43.623		20.001		5.210		8.989		4.203	
p-value	0.000		0.000		0.052		0.000		0.000		0.008		0.018		0.000	
No. of obs.	9,672				9,672				9,672				9,672			

Notes. Each panel represents a separate regression for the dependent variables in Italics in the first column. The table reports the second-stage 2SLS coefficients for each linguistic skill on the outcome variable, the standard error (s.e.), and significance level. Chi^2 refers to the excluded variables in the auxiliary probit model (modified age at arrival, the interaction between speaking Italian during childhood and modified age at arrival, linguistic distance, and the interaction between modified age at arrival and linguistic distance; see Section 4.1), and it is only reported for the first dependent varaible since it does not vary across panels (the probit model is estimated on the largest sample with non-missing dependent and independent variables). The first-stage *F*-statistic refers to the predicted probability of linguistic deficiency used in the first stage of 2SLS as the excluded instrument. Since the model includes both the baseline effect for both genders and the interaction with the female indicator, the model includes two endogenous variables and two first-stage statistics are reported. All models include the control variables in Section 4.1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors clustered by mother tongue are in parentheses.

Difficulties in reading and writing Italian reduced immigrants' labor force participation (LFP) by 39 and 31 pp, respectively, compared to effects that were approximately twice as large for informal skills, specifically −60 pp and −65 pp for speaking and understanding difficulties, respectively. These effects are remarkable, especially given that in the regressions, we control for a wealth of variables that are likely to affect LFP, such as gender, education level, age, marital status, an economic migrant indicator, and having ever worked in the origin country, which is especially useful as a proxy of women's long-term attachment to the labor force. All in all, linguistic difficulties in Italian represent a hurdle to immigrants' LFP. Having poor proficiency in Italian reduces LFP, irrespective of the type of linguistic ability involved, with larger effects when communication skills are considered.

As for the probability of employment, we consider two variables. The first is immigrants' perceived labor force status (in panel (b) of Table 1), while the second is based on whether the immigrant worked in the last week (*current work*, hereafter) or did not work but searched for a job in the last four weeks (unemployed) in panel (c). Both panels convey a similar picture. Consistent with the low-skill profile of the jobs filled by immigrants in Italy, formal linguistic skills were less important than informal ones for finding employment in the Italian labor market. Writing difficulties are never statistically significant, while reading difficulties generated employment gaps of 18 pp and 20 pp on self-perceived employment status and current work, respectively. The effects are much larger for informal skills, with employment penalties of −35 pp and −40 pp for speaking and understanding difficulties, respectively, when perceived employment is considered. The corresponding effects on current work are very similar, −39 pp and −44 pp for speaking and understanding difficulties, respectively.[20]

[20] Our estimated effects of poor linguistic proficiency on integration outcomes are larger compared to the effects documented for other countries. In some cases, differences could be explained by the index of linguistic difficulties used. Yao and van Ours (2015), for instance, report a 22 pp employment penalty for women in the Netherlands and about a 10 pp penalty (statistically non-significant) for men, using a language deficiency indicator that pools deficiencies in both reading and speaking. Thus, the effects they estimate represent an average of the effect of formal and informal skills. Smaller effects compared to Italy could also be explained by the larger diffusion of vehicular languages (e.g., English) in other countries, such as the Netherlands, which could partly compensate for poor proficiency in the destination language. As for Southern Mediterranean countries, sizeable employment penalties on the order of –15 pp for men and –22 pp for women are estimated for Spain by Budria, Colino, and de Ibarreta (2019), who compare individuals with linguistic deficiencies with those who spoke Spanish very well. Conversely, studies of English-speaking countries generally do not find a strong selection of immigrants into employment according to language proficiency (e.g., Guven and Islam 2015 for Australia and Miranda and Zhu 2021 for the United Kingdom).

Shifting the focus to the characteristics of immigrants' jobs, panels (d)–(f) of Table 1 report the results regarding the probability of having open-ended versus fixed-term contracts (panel d), working full-time versus part-time (panel e), and immigrant job satisfaction levels (panel f). All outcomes are analyzed as conditional on employment. Although poor linguistic proficiency was generally associated with a lower probability of having an open-ended contract, with larger penalties for informal skills, the estimates are never statistically significant. Negative coefficients are also estimated for the probability of working full time but are much smaller in magnitude than for the type of contract, and in this case, as well, they are never statistically significant. Only formal skills seemed to impact the level of job satisfaction: poor proficiency in reading and writing led to around a one-point reduction in the satisfaction scale (with an average job satisfaction of 7.6).

SCIF provides a detailed five-digit occupational code of immigrants' occupations (International Standard Classification of Occupations 2008, ISCO-08 hereafter). Although this variable may be affected by substantial measurement errors compared to those we have already considered, we ran further regression models, using as a dependent variable the International Socio-Economic Index of Occupational Status (i.e., ISEI) matched from ISCO-08. ISEI is a proxy of the socio-economic prestige of occupations (Ganzeboom, De Graaf, and Treiman 1992). Curiously, we find in panel (g) that language difficulties were, on average, associated with a *better* ISEI (with estimated coefficients ranging from 19 for writing difficulties to 23 for speaking difficulties). The existence of ethnic (spatial or sectoral) enclaves in immigrants' labor markets could partly explain this apparent contradiction. Linguistic enclaves, especially in the workplace, may shelter immigrants with poor linguistic proficiency from labor market discrimination and allow them to reach better employment positions compared to immigrants working for native employers or with native employees. Yet, the measurement error remains a concern, and this finding deserves further investigation, preferably using administrative data for which the measurement error for occupations is smaller.

6. Heterogeneity Across Linguistic Groups

In *Hypothesis 2,* we put forward that linguistic barriers are less likely to negatively impact the labor market outcomes of more workplace-ethnicized language groups, and in Section 4.2, we identified Chinese-speaking immigrants as the most workplace-ethnicized linguistic group in the Italian immigrant labor market.

Table 2 reports the effect of linguistic difficulties on labor market integration, split between Chinese speakers and other linguistic groups. The coefficients refer to the second stage of 2SLS, involving two endogenous variables (and, therefore, two first-stage equations) — namely, poor proficiency in Italian and poor proficiency interacted with the Chinese linguistic group indicator — which are in turn instrumented with the prediction from the probit model and its interaction with the same indication. The non-interacted coefficient represents the baseline for all linguistic groups, while the coefficient on the

Table 4. Poor Proficiency in Italian and Other Individual Outcomes.

Regressions (Dependent Variable)	Poor Proficiency in							
	Reading		Writing		Speaking		Understanding	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
(a) Feeling accepted	−0.147***	(0.042)	−0.182***	(0.043)	−0.111***	(0.041)	−0.137***	(0.046)
First-stage *F*-statistic	83.172		131.024		144.858		81.837	
p-value	0.000		0.033		0.000		0.000	
No. of obs.	13,787		13,787		13,787		13,787	
(b) Feeling at home	−0.234***	(0.068)	−0.279***	(0.081)	−0.164**	(0.072)	−0.199***	(0.074)
First-stage *F*-statistic	83.172		131.024		144.858		81.837	
p-value	0.000		0.000		0.001		0.000	
No. of obs.	13,787		13,787		13,787		13,787	
(c) Life satisfaction (0–10 scale)	−0.283	(0.227)	−0.365	(0.282)	−0.049	(0.235)	−0.099	(0.241)
First-stage *F*-statistic	83.172		131.024		144.858		81.837	
p-value	0.000		0.000		0.000		0.000	
No. of obs.	13,787		13,787		13,787		13,787	

Notes. Each panel represents a separate regression for the dependent variables in Italics in the first column. The table reports the second-stage 2SLS coefficients for each linguistic skill on the outcome variable, the standard error (s.e.), and significance level. Chi^2 refers to the excluded variables in the auxiliary probit model (modified age at arrival, the interaction between speaking Italian during childhood and modified age at arrival, linguistic distance, and the interaction between modified age at arrival and linguistic distance; see Section 4.1), and it is only reported for the first dependent variable since it does not vary across panels (the probit model is estimated on the largest sample with non-missing dependent and independent variables). The first-stage *F*-statistic refers to the predicted probability of linguistic deficiency used in the first stage of 2SLS as the excluded instrument. All models include the control variables in Section 4.1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors clustered by mother tongue are in parentheses.

Table 5. Poor Proficiency in Italian and Other Individual Outcomes by Linguistic Group.

Regressions (Dependent Variable)	Poor Proficiency in															
	Reading				Writing				Speaking				Understanding			
	Baseline		Chinese		Baseline		Chinese		Baseline		Chinese		Baseline		Chinese	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
(a) Feeling accepted	−0.127***	(0.032)	−0.265***	(0.026)	−0.159***	(0.035)	−0.215***	(0.020)	−0.082***	(0.031)	−0.320***	(0.027)	−0.106***	(0.035)	−0.297***	(0.031)
Chi^2 probit model	11.63				9.404				17.36				25.07			
p-value	0.020				0.000				0.000				0.000			
First-stage *F*-statistic	146.7		16932		232.8		46865		694.8		341125		842.3		538966	
p-value	0.000		0.000		0.052		0.000		0.002		0.000		0.000		0.000	
No. of obs.	13,787				13,787				13,787				13,787			
(b) Feeling at home	−0.214***	(0.067)	−0.276***	(0.068)	−0.252***	(0.082)	−0.254***	(0.061)	−0.133*	(0.071)	−0.345***	(0.055)	−0.168**	(0.075)	−0.290***	(0.066)
First-stage *F*-statistic	146.7		16932		232.8		46865		694.8		341125		842.3		538966	
p-value	0.000		0.000		0.000		0.000		0.000		0.002		0.000		0.000	
No. of obs.	13,787				13,787				13,787				13,787			
(c) Life satisfaction (0–10)	−0.228	(0.233)	−0.614**	(0.306)	−0.319	(0.298)	−0.363	(0.259)	0.022	(0.240)	−0.861***	(0.296)	−0.030	(0.256)	−0.686**	(0.303)
First-stage *F*-statistic	146.7		16932		232.8		46865		694.8		341125		842.3		538966	
p-value	0.000		0.000		0.000		0.052		0.002		0.000		0.000		0.000	
No. of obs.	13,787				13,787				13,787				13,787			

Notes. Each panel represents a separate regression for the dependent variables in Italics in the first column. The table reports the second-stage 2SLS coefficients for each linguistic skill on the outcome variable, the standard error (s.e.), and significance level. Chi^2 refers to the excluded variables in the auxiliary probit model (modified age at arrival, the interaction between speaking Italian during childhood and modified age at arrival, linguistic distance, and the interaction between modified age at arrival and linguistic distance; see Section 4.1), and it is only reported for the first dependent variable since it does not vary across panels (the probit model is estimated on the largest sample with non-missing dependent and independent variables). The first-stage *F*-statistic refers to the predicted probability of linguistic deficiency used in the first stage of 2SLS as the excluded instrument. Since the model includes both the baseline effect for all linguistic groups and the interaction with the Chinese linguistic group, the model includes two endogenous variables and two first-stage statistics are reported. All models include the control variables in Section 4.1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors clustered by mother tongue are in parentheses.

interaction with the Chinese indicator measures differences between this group and the baseline. Given that both the interacted endogenous variable and the instrument are zero for non-Chinese immigrants, the instrument predicts well the value of the endogenous variable for all non-Chinese immigrants, representing the majority of the sample (pooling all other language groups): the *F*-statistics for the excluded instrument in the first-stage equation shown in Table 2 is very high.

LFP was not differentially affected by language difficulties in formal skills among Chinese-speaking immigrants compared to other language groups (panels (a) and (b) of Table 2). However, Chinese-speaking immigrants were less harmed by poor communication skills compared to other groups: the LFP penalty was 27 pp lower for poor speaking proficiency and 36 pp for poor understanding proficiency. As for results on the perceived and current employment status, differences between Chinese speakers and immigrants of other linguistic groups are even more striking. Panel (b) shows that the sizeable average employment penalty associated with poor speaking proficiency and understanding was not observed for Chinese speakers (i.e., the sum of the baseline and the interaction coefficients is close to zero). Panel (c) confirms that this result is robust to considering immigrants' current working status. This result reveals that a lack of speaking and understanding abilities in Italian is less crucial to finding a job for Chinese-speaking immigrants, compared to other immigrants.

Interestingly, we do not find any effect of language skills on the probability of working under an open-ended contract, regardless of the linguistic group. In contrast, language difficulties impacted the probability of full-time work, for which we record positive effects only for Chinese-speaking immigrants. Since the latter are more likely to work for same-country employers, Chinese employers may demand, on average, a longer working-time engagement than their co-nationals. To further elaborate the point, we estimated additional regressions, using weekly working hours as the dependent variable (results available upon request). We find that Chinese-speaking immigrants with speaking or understanding difficulties worked, on average, 11 and nine more hours per week, respectively. A possible explanation is that Chinese employers might exploit the lack of labor market opportunities of co-nationals without adequate language skills to extract economic advantages from them.

All of these elements contribute to explaining why poor linguistic proficiency did not negatively affect job satisfaction among Chinese-speaking workers compared to immigrants from other linguistic groups, who suffered job satisfaction penalties of between 1.2 points (on a 0–10 scale) when lacking sufficient writing skills and 3 points when having a poor understanding of the destination language. Finally, in the case of writing difficulties, we find significantly different ISEI premia for Chinese-speaking immigrants compared to other linguistic groups. However, with Chinese-speaking immigrants being the most workplace-ethnicized linguistic group, we would have expected larger positive

premia for this group, as they would be the least labor-market discriminated. This positive correlation, therefore, remains a caveat and should be addressed in further investigations.

7. Gender Differences

In *Hypothesis 3*, we argued that since immigrant women tend to work in sectors in which the employer is more likely to be a native (i.e., as housekeepers or caregivers), they are likely to suffer larger labor market integration penalties associated with poor language proficiency compared to immigrant men.

Table 3 reports the results of the models exploring gender differences. As in the previous section, we have two endogenous variables in the 2SLS estimation — the baseline poor proficiency indicator, both for men and women, and the interaction with the female indicator — which are instrumented with the predicted probability of poor proficiency in Italian from the probit model and the interaction with the female indicator.

We find that linguistic difficulties were much more important in determining female LFP than male LFP. The smallest gap of −29 pp is observed for writing, but the gap reached −69 pp for understanding, whereas the lack of speaking or understanding skills did not negatively affect men's LFP. Thus, female immigrants' LFP was influenced by linguistic skills, whereas male immigrants were more likely to participate in the labor force, irrespective of their language proficiency.

After having overcome this LFP hurdle, however, women did not suffer higher penalties for poor language proficiency compared to men in terms of employability. In reality, the absence of a gender gap is more clear-cut for perceived employment status, for which poor language proficiency–female interactions were both small and statistically non-significant, than for current working status, for which the interactions for speaking and understanding, albeit statistically non-significant, were negative and quite large in magnitude.

Gender gaps did not emerge in the probability of working under an open-ended versus a fixed-term contract, full-time versus part-time, the level of job satisfaction, or ISEI, irrespective of the linguistic ability domain considered. It must be noted that except in panel (a) of Table 3, the F-statistics for poor informal linguistic proficiency–female interactions were a bit low (below 10), and we might lack sufficient statistical power to detect significant gender differences (weak-instrument problem).[21]

[21] Although there might also be differences by both gender and linguistic group (i.e., the combination of the two characteristics), we are not able to investigate these patterns, mainly because our empirical strategy would require simultaneously instrumenting the baseline (linguistic deficits) and the double (linguistic deficits * gender, linguistic deficits * linguistic group) and triple interactions (linguistic deficits * gender * linguistic group) between endogenous variables, making the instruments weak.

Table 6. Poor Proficiency in Italian and Other Individual Outcomes by Gender.

Regressions (Dependent Variable)	Poor Proficiency in															
	Reading				Writing				Speaking				Understanding			
	Baseline		Female		Baseline		Female		Baseline		Female		Baseline		Female	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
(a) Feeling accepted	−0.187***	(0.051)	0.068**	(0.029)	−0.225***	(0.050)	0.085***	(0.028)	−0.180**	(0.071)	0.084*	(0.049)	−0.222***	(0.075)	0.105*	(0.055)
First-stage *F*-statistic	101.452		41.029		71.077		86.986		83.406		127.360		47.145		68.194	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
No. of obs.	13,787				13,787				13,787				13,787			
(b) Feeling at home	−0.307***	(0.085)	0.120**	(0.048)	−0.342***	(0.097)	0.125**	(0.048)	−0.286***	(0.106)	0.148*	(0.079)	−0.329***	(0.112)	0.161*	(0.092)
First-stage *F*-statistic	101.452		41.029		71.077		86.986		83.406		127.360		47.145		68.194	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
No. of obs.	13,787				13,787				13,787				13,787			
(c) Life satisfaction (0–10 scale)	−0.357	(0.273)	0.138	(0.193)	−0.446	(0.304)	0.174	(0.161)	−0.244	(0.395)	0.230	(0.279)	−0.334	(0.379)	0.286	(0.275)
First-stage *F*-statistic	101.452		41.029		71.077		86.986		83.406		127.360		47.145		68.194	
p-value	0.000		0.000		0.000		0.000		0.000		0.002		0.000		0.000	
No. of obs.	13,787				13,787				13,787				13,787			

Notes. Each panel represents a separate regression for the dependent variables in Italics in the first column. The table reports the second-stage 2SLS coefficients for each linguistic skill on the outcome variable, the standard error (s.e.), and significance level. Chi^2 refers to the excluded variables in the auxiliary probit model (modified age at arrival, the interaction between speaking Italian during childhood and the modified age at arrival, linguistic distance, and the interaction between modified age at arrival and linguistic distance; see Section 4.1), and it is only reported for the first dependent variable since it does not vary across panels (the probit model is estimated on the largest sample with non-missing dependent and independent variables). The first-stage *F*-statistic refers to the predicted probability of linguistic deficiency used in the first stage of 2SLS as the excluded instrument. Since the model includes both the baseline effect for both genders and the interaction with the female indicator, the model includes two endogenous variables and two first-stage statistics are reported. All models include the control variables in Section 4.1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors clustered by mother tongue are in parentheses.

8. Other Individual Outcomes

Labor market outcomes are only a part of the multidimensional experience of immigrant integration (Harder et al. 2018). In particular, it is interesting to assess whether economic outcomes go hand in hand with other proxies of immigrant perceptions of their integration into the host society and their level of life satisfaction. In this section, we focus on three additional variables provided by SCIF: *feeling at home where they live*, *feeling accepted in Italy,* and overall *life satisfaction,* with a scale from 0, indicating the poorest level of perceived fulfillment, to 10 indicating the highest.

Panel (a) of Table 4 explores whether immigrants felt accepted where they lived. Linguistic difficulties significantly reduced the likelihood of feeling accepted, with slightly larger effects for formal skills. The largest effect (–18 pp) was observed for writing and the smallest (–11 pp) for speaking skills. In panel (b), we investigate whether immigrants with linguistic difficulties were unlikely to feel at home. As a result, we find statistically significant negative effects in all linguistic domains, with larger effects when poor proficiency concerned formal skills (e.g., −28 pp for writing and −16 pp for speaking). Broadly, we notice that linguistic barriers negatively impacted immigrants' perceptions of being part of Italian society, with larger effects linked to a lack of formal skills.

The last panel of Table 4 illustrates the effects of linguistic barriers on immigrants' overall life satisfaction, which encompasses satisfaction with multidimensional individual life aspects. The estimated coefficients are always negative and statistically non-significant (or very small for informal skills); therefore, we conclude that poor proficiency in Italian did not strongly reduce immigrants' life satisfaction levels.

Table 5 reports the analysis of the additional outcomes by linguistic group. We observe a sort of *reversal of fortune* with respect to labor market outcomes. Indeed, Chinese-speaking immigrants suffered significantly more negative effects from poor language proficiency compared to other linguistic groups. For indicators of both *feeling accepted* and *feeling at home*, the largest gap was observed with regard to speaking skills (–32 pp and −35 pp, respectively). This result is mirrored in significantly lower life satisfaction among Chinese-speaking immigrants with poor proficiency in Italian, when compared to other linguistic groups (–0.61, −0.86, and −0.69 lower satisfaction -on a 0–10 scale- for reading, speaking, and understanding difficulties, respectively).

When allowing the effect of linguistic difficulties to be heterogeneous across genders (Table 6), women's perceptions of their integration in Italy appear to be less dependent on their linguistic proficiency compared to men. The probability of *feeling accepted* was reduced by 7 pp for poor proficiency in reading, 9 pp for writing, 8 pp for speaking, and 11 pp for understanding. A similar picture emerges for the probability of *feeling at home*, with even larger gender gaps in favor of women and with differences in the range of 12 pp to 16 pp (i.e., the negative baseline effects were reduced by the positive interaction effects). When *overall life satisfaction*

is considered, poor proficiency in Italian did not seem to play an important role, irrespective of gender.

9. Conclusion

This article examined the impact of poor proficiency in Italian on immigrants' labor market outcomes, using instrumental variables estimators to deal with the endogeneity of linguistic proficiency, measurement error in self-reported language proficiency, and reverse causality and to investigate differences between informal and formal linguistic abilities. Despite previous studies having mainly used a unidimensional indicator of immigrants' language proficiency (e.g., limiting the focus on speaking ability, like Bleakley and Chin 2010), our analysis distinguishes between immigrants' communication skills (understanding and speaking the destination language) and more formal linguistic skills (reading and writing in the destination language), while also investigating heterogeneity across immigrant linguistic groups and genders.

We find that poor communication skills in Italian were much more crucial than deficiencies in formal skills in determining immigrants' labor force participation and their working status. Conditional on immigrants' employment, we do not find evidence of large or even statistically significant negative effects of poor proficiency in Italian on the type of contract, working-time arrangements, level of job satisfaction, or socio-economic occupational prestige (ISEI).

Our analysis uncovers heterogeneity by immigrant linguistic enclave. Except for some caveats concerning the analysis of ISEI, the results show that, in terms of labor market opportunities, Chinese-speaking immigrants, as the most workplace-ethnicized linguistic group, were less penalized by poor proficiency in Italian than other linguistic groups. Although poor proficiency in Italian generally did not represent an obstacle to their successful labor market outcomes, Chinese immigrants with poor proficiency in Italian scored much worse on measures of perceived integration and level of life satisfaction than other linguistic groups. We also identify a significant gender differential in the role of Italian as the destination language for immigrants' labor market integration. Our results reveal that poor proficiency in Italian hindered labor market outcomes more for immigrant women than for immigrant men and that women suffered from larger linguistic obstacles to their labor force participation, especially when speaking and understanding skills are examined.

Three main policy implications are derived from our results. First, in assessing the role of Italian proficiency in shaping immigrants' labor market outcomes, our analysis offers support for public investments in educational programs targeting adult immigrants who arrive in Italy without previous knowledge of Italian. Programs aimed at teaching Italian as a second language offer immigrants the opportunity to improve their labor market participation and employment. As the COVID-19 pandemic has clearly demonstrated, immigrants play an important role in the Italian economy, not only substituting for natives but also complementing them in performing essential activities in agriculture and food distribution (Fasani and Mazza 2020), as well as in health and elderly care

sectors. Second, our analysis suggests that immigrants' language training in the destination language should focus on teaching informal communication skills (speaking and understanding Italian), which seem to be more important than formal skills (reading and writing in Italian) for immigrants' labor market integration. Finally, this article raises issues regarding ethnic linguistic enclaves, identifying Chinese-speaking immigrants as the most workplace-ethnicized linguistic group in Italy's labor market. Our results show that although higher workplace ethnicization increases labor force participation and employment among immigrants, it can negatively impact immigrants' feelings of belonging to the destination society and their perception of life satisfaction. This last result raises concerns regarding the potential trade-offs produced by linguistic enclaves at workplaces: although deficiencies in Italian may not prevent successful labor outcomes, poor language proficiency can lead immigrants toward social marginalization and feelings of exclusion from the hosting societies. Nevertheless, due to the lack of systematic data collection on the social conditions of immigrants living in Italy, our analysis is limited to one point in time, while a better understanding of linguistic integration dynamics requires a constant monitoring during the entire immigrant life course to appreciate dynamics along with a longitudinal perspective. Until new individual-level survey datasets are available, the SCIF remains the richest and most recent data source to investigate immigrant integration in Italy adopting a demographic and linguistic perspective.

Our contribution to the study of international migration is twofold. First, we show that linguistic skills in the destination language are not all the same and that, for low-skilled jobs, communication (speaking and understanding) skills are more important than formal skills (reading and writing) for successful immigrant labor market outcomes. Second, this article demonstrates that linguistic enclaves in the labor market do not always represent a hurdle to immigrants' labor market success: they can enhance immigrants' labor force participation and employment status yet also generate trade-offs for other non-labor market integration outcomes.

Despite the analytic focus on Italy, the findings presented here are likely to be relevant for other Southern European countries with similar migration-transition pathways and immigrant population profiles. As an avenue for future research, then, the proposed approach could be extended to assess how the interplay between labor market integration and other outcomes of social integration vary according to immigrants' demographic and linguistic patterns, contextual labor market characteristics, and the language spoken in the destination society.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Daniela Ghio https://orcid.org/0000-0002-2687-7033

Supplemental Material

Supplemental material for this article is available online.

References

Alcaraz Pribaz, C., D. Chiquiar, and A. Salcedo. 2012. "Remittances, Schooling, and Child Labor in Mexico." *Journal of Development Economics* 97 (1): 156–65. doi:10.1016/j.jdeveco.2010.11.004

Amuedo-Dorantes, C., and S. de la Rica. 2011. "Complements or Substitutes? Task Specialization by Gender and Nativity in Spain." *Labour Economics* 18 (5): 697–707. doi:10.1016/j.labeco.2011.02.002

Andall, J. 2000. *Gender, Migration and Domestic Service. The Politics of Black Women in Italy*. Abingdon, UK: Routledge.

Bacolod, M., and M. A. Rangel. 2017. "Economic Assimilation and Skill Acquisition: Evidence from Occupation Sorting of Childhood Immigrants." *Demography* 54 (2): 571–602. doi:10.1007/s13524-017-0558-2

Barone, G., and S. Mocetti. 2011. "With a Little Help from Abroad: The Effect of Low-Skilled Immigration on the Female Labour Supply." *Labour Economics* 18 (5): 664–75. doi:10.1016/j.labeco.2011.01.010

Barry, Chiswick R. 2009. "Ten Top Myths and Fallacies Regarding Immigration." *Writing Immigration: Scholars and Journalists in Dialogue*. Marcelo Suarez-Orozco, ISBN-13: 9780520267176. doi:10.1525/california/9780520267176.003.0008

Berman, E., K. Lang, and E. Siniver. 2000. "Language-skill Complementarity: Returns to Immigrant Language Acquisition." Department of Economics, Boston University, Working paper.

Bertrand, M., E. F. P. Luttmer, and M. S. Mullainathan. 2000. "Network Effects and Welfare Cultures." *Quarterly Journal of Economics* 115 (3): 1019–55. doi:10.1162/003355300554971

Bleakley, H., and A. Chin. 2004. "Language Skills and Earnings: Evidence from Childhood Immigrants." *Review of Economics and Statistics* 86 (2): 481–96. doi:10.1162/003465304323031067

Bleakley, H., and A. Chin. 2010. "Age at Arrival, English Proficiency, and Social Assimilation among U.S. Immigrants." *American Economic Review: Applied Economics* 2 (1): 165–92. doi:10.1257/app.2.1.165

Borjas, G. J. 1992. "Ethnic Capital and Intergenerational Mobility." *Quarterly Journal of Economics* 107 (1): 123–50. doi:10.2307/2118325

——— 1995. "Ethnicity, Neighborhoods, and Human-Capital Externalities." *American Economic Review* 85 (3): 365–90.

——— 1998. "To Ghetto or not to Ghetto: Ethnicity and Residential Segregation." *Journal of Urban Economics* 44 (2): 228–53. doi:10.1006/juec.1997.2068

Bratti, M., P. Cella, L. De Benedictis, and G. Santoni. 2019. "Imprenditoria Immigrata ed Esportazioni." In *L'Italia nell'Economia Internazionale In Rapporto ICE 2018-2019*, edited by S. Camerano, G. Corò, L. De Benedictis, V. De Luca, A. Dell'Acqua, S. De Nardis, G. Di Vaio, S. Fabiani, P. Giordani, G. Giovannetti, A. Giunta, L. Iapadre, A. Lanza, G. Maglione, S. Menghinello, R. Monducci, F. Onida, F. Rullani, L.Tajoli, A. Terzulli, F. Tilli, and G. Viesti. 232–236. Roma: Istituto per il Commercio Estero (ICE). https://www.ice.it/it/sites/default/files/inline-files/Rapporto%20ICE%202018-2019_completo%20per%20web_1.pdf.

Budria, S., A. Colino, and C. M. de Ibarreta. 2019. "The Impact of Host Language Proficiency on Employment Outcomes among Immigrants in Spain." *Empirica* 46 (4): 625–52. doi:10.1007/s10663-018-9414-x

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2017. "Rdrobust: Software for Regression-Discontinuity Designs." *Stata Journal* 17 (2): 372–404. doi:10.1177/1536867X1701700208

Carliner, G. 2000. "The Language Ability of U.S. Immigrants: Assimilation and Cohort Effects." *International Migration Review* 34 (1): 158–82.

Ceccagno, A. 2015. "The Mobile Emplacement: Chinese Migrants in Italian Industrial Districts." *Journal of Ethnic and Migration Studies* 41 (7): 1111–30. doi:10.1080/1369183X.2014.967755

Chiswick, B. R. 1991. "Speaking, Reading, and Earnings among Low-Skilled Immigrants." *Journal of Labor Economics* 9 (2): 149–70. doi:10.1086/298263

———, Y. L. Lee, and P. W. Miller. 2002. "A Longitudinal Analysis of Immigrant Occupational Mobility: A Test of the Immigrant Assimilation Hypothesis." *International Migration Review* 39 (2): 332–53. doi:10.1111/j.1747-7379.2005.tb00269.x

———, and P. W. Miller. 1992. "Language in the Immigrant Labor Market." In *Immigration, Language and Ethnicity: Canada and the United States*, edited by B. R. Chiswick, 229–96. Washington: American Enterprise Institute.

———. 1996. "Ethnic Networks and Language Proficiency among Immigrants." *Journal of Population Economics* 9 (1): 19–35. doi:10.1007/PL00013277

———. 2001. "A Model of Destination-Language Acquisition: Application to Male Immigrants in Canada." *Demography* 38 (3): 391–409. doi:10.1353/dem.2001.0025

———. 2003. "The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada." *Economics of Education Review* 22 (5): 469–80. doi:10.1016/S0272-7757(03)00037-2

———. 2005. "Do Enclaves Matter in Immigrant Adjustment?" *City and Community* 4 (1): 5–35. doi:10.1111/j.1535-6841.2005.00101.x

Clarke, A., and I. E. Isphording. 2017. "Language Barriers and Immigrant Health." *Health Economics* 26 (6): 765–78. doi:10.1002/hec.3358

Coulmas, F. 2005. *Sociolinguistics: The Study of Speaker's Choices*. Cambridge: Cambridge University Press.

Dong, J. K., and J. Blommaert. 2009. "Space, Scale and Accents: Constructing Migrant Identity in Beijing." *Multilingual Journal of Cross Cultural and Interlanguage Communication* 28 (1): 45–78. doi:10.1515/mult.2009.003

Duleep, H. O., X. Liu, and M. Regets. 2014. "Country of Origin and Immigrant Earnings, 1960–2000: A Human Capital Investment Perspective." IZA Discussion Paper N. 8628. Bonn: IZA Institute of Labor Economics.

Egger, P., M. Koethenbuerger, and M. Smart. 2010. "Do Fiscal Transfers Alleviate Business Tax Competition? Evidence from Germany." *Journal of Public Economics* 94 (3-4): 235–46. doi:10.1016/j.jpubeco.2009.10.002

Fasani, F., T. Frattini, and L. Minale. 2021. "Lift the Ban? Initial Employment Restrictions and Refugee Labour Market Outcomes." *Journal of the European Economic Association* 19 (5): 2803–54. doi:10.1093/jeea/jvab021

———, and J. Mazza. 2020. "Immigrant Key Workers: Their Contribution to Europe's COVID-19 Response." IZA Policy Paper No. 155. Bonn: IZA Institute of Labor Economics.

Frattini, T., and I. Solmone. 2022. *6th Migration Observatory Report: "Immigrant Integration in Europe"*. Milan: Centro Studi Luca D'Agliano (LdA). ISBN 978-88-946015-1-0.

Fullin, G., and E. Reyneri. 2011. "Low Unemployment and Bad Jobs for New Immigrants in Italy." *International Migration* 49 (1): 118–47. doi:10.1111/j.1468-2435.2009.00594.x

Ganzeboom, H. B. G., P. M. De Graaf, and D. J. Treiman. 1992. "A Standard International Socio-Economic Index of Occupational Status." *Social Science Research* 21 (1): 1–56. doi:10.1016/0049-089X(92)90017-B

Geurts, N., and M. Lubbers. 2019. "The Role of Country of Origin Engagement in Second-Language Proficiency of Recent Migrants." *Ethnic and Racial Studies* 42 (16): 120–40. doi:10.1080/01419870.2019.1606434

Guven, C., and A. Islam. 2015. "Age at Migration, Language Proficiency, and Socioeconomic Outcomes: Evidence from Australia." *Demography* 52 (2): 513–42. doi:10.1007/s13524-015-0373-6

Harder, N., L. Figueroa, R. M. Gillum, D. Hangartner, D. D. Laitin, and J. Hainmueller. 2018. "Multidimensional Measure of Immigrant Integration." *Proceedings of the National Academy of Sciences* 115 (45): 11483–8. doi:10.1073/pnas.1808793115

Hart-Gonzalez, L., and S. Lindemann. 1993. "Expected Achievement in Speaking Proficiency." School of Language Studies, Foreign Service Institute, Department of State, mimeo.

Heckman, J. J., and S. Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annual Review of Economics* 6: 689–733. doi:10.1146/annurev-economics-080213-040753.

Hermansen, A. S. 2017. "Age at Arrival and Life Chances Among Childhood Immigrants." *Demography* 54 (1): 201–29. doi:10.1007/s13524-016-0535-1

——— 2022. "Visualizing Intergenerational Immigrant Assimilation at Work." *Sociological Research for a Dynamic World*, January. https://doi.org/10.1177/23780231211072590.

Isphording, I., and S. Otten. 2014. "Linguistic Barriers in the Destination Language Acquisition of Immigrants." *Journal of Economic Behavior & Organization* 105 (C): 30–50. doi:10.1016/j.jebo.2014.03.027

Istat. 2019. "Welcome to Immigrants.Stat." http://stra-dati.istat.it/Index.aspx?DataSetCode=DCIS_POPSTRCIT1.

Lazear, E. P. 1999. "Culture and Language." *Journal of Political Economy* 107 (6): 95–126. doi:10.1086/250105

Lelièvre, E., and N. Robette. 2010. "A Life Space Perspective to Approach Individual Demographic Processes." *Canadian Studies in Population* 37 (1-2): 207–44. doi:10.25336/P69918

Lenneberg, E. H. 1967. *Biological Foundation of Language*. New York: Wiley.

Levinson, A. 2005. *The Regularisation of Unauthorized Migrants: Literature Survey and Country Case Studies*. Oxford, UK: University of Oxford Centre on Migration, Policy and Society.

Li, E. X., and P. S. Li. 2016. "Immigrant Enclave Thesis Reconsidered: Case of Chinese Immigrants in the Enclave and Mainstream Economy in Canada." *Journal of International Migration and Integration* 17 (1): 131–51. doi:10.1007/s12134-014-0391-z

Liu, S. 2015. *Identity, Hybridity, and Cultural Home: Chinese Migrants and Diaspora in Multicultural Societies*. London and New York: Rowman & Littlefield International.

Miranda, A., and Y. Zhu. 2013. "English Deficiency and the Native Immigrant Wage Gap." *Economics Letters* 118 (1): 38–41. doi:10.1016/j.econlet.2012.09.007

———. 2021. "The Effect of Deficiency at English on Female Immigrants' Wage in the UK: Correcting for Measurement Error, Endogenous Treatment, and Sample Selection Bias." *Applied Economics Letters* 28 (5): 349–53. doi:10.1080/13504851.2020.1753873

Netten, J., and C. Germain. 2012. "A New Paradigm for the Learning of a Second or Foreign Language: The Neurolinguistic Approach." *Neuroeducation Journal* 1 (1): 85–114. doi:10.24046/neuroed.20120101.85

Peri, G., and C. Sparber. 2009. "Task Specialization, Immigration, and Wages." *American Economic Journal: Applied Economics* 1 (3): 135–69. doi:10.1257/app.1.3.135

Piil Damm, A. 2009. "Ethnic Enclaves of Immigrant Labor Market Outcomes: Quasi Experimental Evidence." *Journal of Labor Economics* 27 (2): 281–314. doi:10.1086/599336

Portes, A. 1998. "Social Capital: Its Origins and Applications in Modern Sociology." *Annual Review of Sociology* 24: 1–24. doi:10.1146/annurev.soc.24.1.1

Raffaetà, R., and L. Baldassar. 2015. "Spaces Speak Louder Than Words: Contesting Social Inclusion Through Conflicting Rhetoric About Prato's Chinatown." In *Chinese Migration to Europe*, edited by L. Baldassar, G. Johanson, N. McAuliffe, and M. Bressan, 119–37. London: Palgrave Macmillan.

Raijman, R., M. Semyonov, and R. Gffen. 2015. "Language Proficiency among Post-1990 Immigrants in Israel." *Journal of Ethnic and Migration Studies* 41 (8): 1347–71. doi:10.1080/1369183X.2014.982523

Ranci, C., and M. Arlotti. 2019. "Resistance to Change, the Problem if High Non-Take Up in Implementing Policy Innovations in the Italian Long-Term Care System." *Policy and Society* 38 (4): 572–88. doi:10.1080/14494035.2019.1619995

Reyneri, E. 1998. "The Role of The Underground Economy in Irregular Migration to Italy: Cause or Effect?" *Journal of Ethnic and Migration Studies* 24 (2): 313–31. doi:10.1080/1369183X.1998.9976635

——— 2004. "Immigrants in a Segmented and Often Undeclared Labour Market." *Journal of Modern Italian Studies* 9 (1): 71–93. doi:10.1080/1354571042000179191

Stock, J., and M. Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by D. Andrews and J. Stock, 80–108. Cambridge: Cambridge University Press.

Ufficio di Statistica Comune di Prato. 2021. "Analisi demografica della popolazione residente al 31/12/2020 a cura dell'Ufficio di Statistica." http://allegatistatistica.comune.prato.it/dl/20210302115944079/relazione_demo2020.pdf.

Van Tubergen, F., and M. Kalmijn. 2009. "A Dynamic Approach to the Determinants of Immigrants' Language Proficiency: The United States, 1980–2000." *International Migration Review* 43 (3): 519–43. doi:10.1111/j.1747-7379.2009.00776.x

Venturini, A., and C. Villosio. 2006. "Labour Market Effects of Immigration into Italy: An Empirical Analysis." *International Labour Review* 145 (1-2): 91–118. doi:10.1111/j.1564-913X.2006.tb00011.x

Vertovec, S. 2007. "Super-diversity and its Implication." *Ethnic and Racial Studies* 30 (6): 1024–54. doi:10.1080/01419870701599465

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Boston, MA: The MIT Press.

Yao, Y., and J. C. van Ours. 2015. "Language Skills and Labor Market Performance of Immigrants in the Netherlands." *Labour Economics* 34: 76–85. doi:10.1016/j.labeco.2015.03.005

Zhou, M., and J. Logan. 1989. "Returns on Human Capital in Ethnic Enclaves: New York City's Chinatown." *American Sociological Review* 54 (5): 809–20. doi:10.2307/2117755

Zwysen, W. 2019. "Different Patterns of Labor Market Integration by Migration Motivation in Europe: The Role of Host Country Human Capital." *International Migration Review* 53 (1): 59–89. doi:10.1177/0197918318767929