DEMOGRAPHIC RESEARCH

Research Article

Local dialect proficiency and migrants' identity integration: A case of Shanghai

Simin Zhang

Muzhi Zhou

Contents

Local dialect proficiency and migrants' identity integration: A case of Shanghai

Simin Zhang[1]

Muzhi Zhou[2]

Abstract

BACKGROUND
Language proficiency is a crucial factor in migrant integration; however, few studies have examined the relationship between local dialect acquisition and migrant integration in China, a country with numerous dialects. The significance of dialects in shaping regional identity, nonetheless, may have diminished as Putonghua becomes increasingly popular.

METHODS
Based on the Shanghai Urban Neighborhood Survey 2017 data, we employ instrumental variable methods with OLS regressions to investigate the relationship between the Shanghai local dialect acquisition and identity integration, one of the most fundamental parts of integration and 'citizenization,' among migrants in Shanghai.

RESULTS
Higher Shanghaihua proficiency is positively linked to migrants' self-identification as Shanghainese. Moreover, the importance of Shanghaihua proficiency in integration into the local identity increases with age, indicating the changing significance of Shanghaihua proficiency concerning identity integration over time. This close relationship between Shanghaihua proficiency and identity recognition also tends to maintain even for those who have already acquired the local *hukou* status (the household registration system).

CONTRIBUTIONS
As migration patterns in China change, integrating migrants into the local population has become a key focus of social welfare and population policy. This paper provides new insights into a society marked by a dual-language system during the transitional period of hukou reform. Considering the growing prevalence of Putonghua and the ongoing inflow of migrants into this megacity, this study also explores the cultural and identity implications of a more diverse urban population.

---

[1] Urban Governance and Design, Hong Kong University of Science and Technology (Guangzhou), China.
[2] Urban Governance and Design, Hong Kong University of Science and Technology (Guangzhou), China. Email: mzzhou@hkust-gz.edu.cn.

1. Introduction

The integration of migrants into their host society is a central topic in migration studies. From the perspective of the overall well-being of migrants, the disparity between migrants' behavior and local cultural norms can lead to stress, exerting adverse effects on their well-being, such as life satisfaction, all-cause mortality, morbidity of physical diseases, and affective balance (Appau, Churchill, and Farrell 2019; Berkman 1995; Bian et al. 2015; Cohen 1988; House, Landis, and Umberson 1988; Wei and Gao 2017). Numerous studies have explored the factors that determine migrants' or mostly immigrants' multidimensional integration into their host society (Ager and Stranger 2004; Alba 1999; Alba and Nee 1997; European Commission: Directorate-General for Migration and Home Affairs 2013). Among these factors, language proficiency has been identified as one of the most influential determinants. Research has shown that a higher level of local language proficiency is often associated with higher levels of integration in terms of income and education (Ager and Stranger 2004; Delander et al. 2005), health (Holmes 2012; Sze et al. 2015), and social attitudes (Alexander, Edwards, and Temple 2007; Larin 2020; Nakhaie 2020).

Although there is extensive literature on the relationship between host language proficiency and migrant integration, few studies have examined this connection within the Chinese context. The role of local language in influencing integration is particularly intriguing as China features a dual-language system that includes the dominant Putonghua (standard Mandarin) alongside various distinct local dialects. China also has a large floating population, reaching almost 38 million reported by the latest 2020 census, a 69.73% increase compared to the number in 2010 (Office of the Leading Group of the State Council for the Seventh National Population Census 2021).[3] Previous research on factors influencing social integration of domestic migrants in China has mainly focused on policy-related structural factors, such as *hukou* policy (Chen and Wang 2015; Wang and Fan 2012) and its changes (Chen, Wu, and Xu 2023) in the historical context of massive rural to urban migration. However, with over 60% of the population now living in urban areas since 2019, there have been substantial changes in hukou policies and social insurance subscription in China, such as the removal of the differentiation between rural and urban hukou, the slowdown of the massive urbanization process, and the migration of highly educated and urban people to tier 1 cities (Wang et al. 2023), especially to metropolitan cities like Beijing or Shanghai. Improving the social integration of migrants has been a growing concern of the government and academia as a vital part of China's new urbanization strategy (Huang et al. 2018; Huang, Zhang, and Xue 2018; Liu, Huang, and Zhang 2018). Since 2014, the government has shifted its

---

[3] A floating population refers to those who live in places other than their household registration area for more than half a year, excluding those who are moving inside a city of a province or a district of a municipality.

focus to encourage the integration of migrants or mostly migrant workers into local 'citizens' who will settle permanently in cities with equal access to public services (L. Li 2024).

The impact of language, particularly dialects, could be a crucial factor in the integration of migrants in the Chinese context, yet it is often overlooked in discussions about domestic migration in China. The clear division between the rural and urban hukou makes most previous studies focus on the disparities in economic status or access to welfare between rural-to-urban migrants and urban residents. With recent hukou reforms – especially the 2014–2025 initiatives easing restrictions and expanding service access – the rigid rural–urban registration divide may have become less pivotal (Chen, Wu, and Xu 2023), and factors such as language, cultural norms, and social networks may increasingly influence how migrants integrate into major cities. In this paper, we focus on language. Although most Chinese, especially the younger, more educated generation, can speak Putonghua, 86.38% of Chinese use dialects in their daily life (Du 2015). Theoretically, language is among the top factors determining one's identity. The difference between home and host dialects in China still represents social distance between migrants and locals and further affects how appealing different migration destinations are (Li and Meng 2014; Liu et al. 2015; Lu and Zhang 2019). In Shanghai, for example, the ability to use local dialects is fundamental in shaping how one could be perceived by Shanghainese as locals or others (Kang 2017). Therefore, the affinity between the home and host dialect could affect the cultural and identity aspects of migrants' integration.

In our study, we selected Shanghai, a city with a population of 24.75 million in 2022 and is one of the largest migrant cities in China. Migrants account for 42.1% of the city's population. Unlike previous literature that focuses on immigrants' mother tongues' defining power of community (Alexander, Edwards, and Temple 2007; Gómez Cervantes 2023; Tannenbaum 2009), in Shanghai, the local dialect, Shanghaihua, is a powerful indicator of an individual's locality (Xu 2021). It is widely used among local residents in daily scenarios, such as grocery shopping and communicating with government officers, but is extremely difficult to understand for people from other regions of China. Our study aims to investigate one of the rarely studied channels of integration – the local dialect – in shaping immigrants' integration in terms of identity, using 2017 data from the Shanghai Urban Neighborhood Survey (SUNS). Meanwhile, with the designation of Putonghua as the official language used in education and the continuous inflow of migrants into Shanghai, the significance of local dialect is diminishing (Shanghai Bureau of Statistics 2014), leaving the role of Shanghaihua in shaping domestic migrants' integration uncertain.

This research contributes to the existing literature on the social integration of migrants by examining a unique region characterized by its local dialect, significant

internal migration, and the predominance of Putonghua at the national level. Since 2014, the Chinese government has recognized the integration of migrants into cities – referred to as their 'citizenization' – as a key objective of new urbanization. This highlights the importance of improving citizenship during a period of mass migration to major metropolitan areas. In this policy context, this study offers new insights into the role of dialects as both cultural bridges and boundaries, as well as the debates surrounding the preservation of linguistic heritage, local identity, and how the relationships among these factors can adapt to a more diverse urban population.

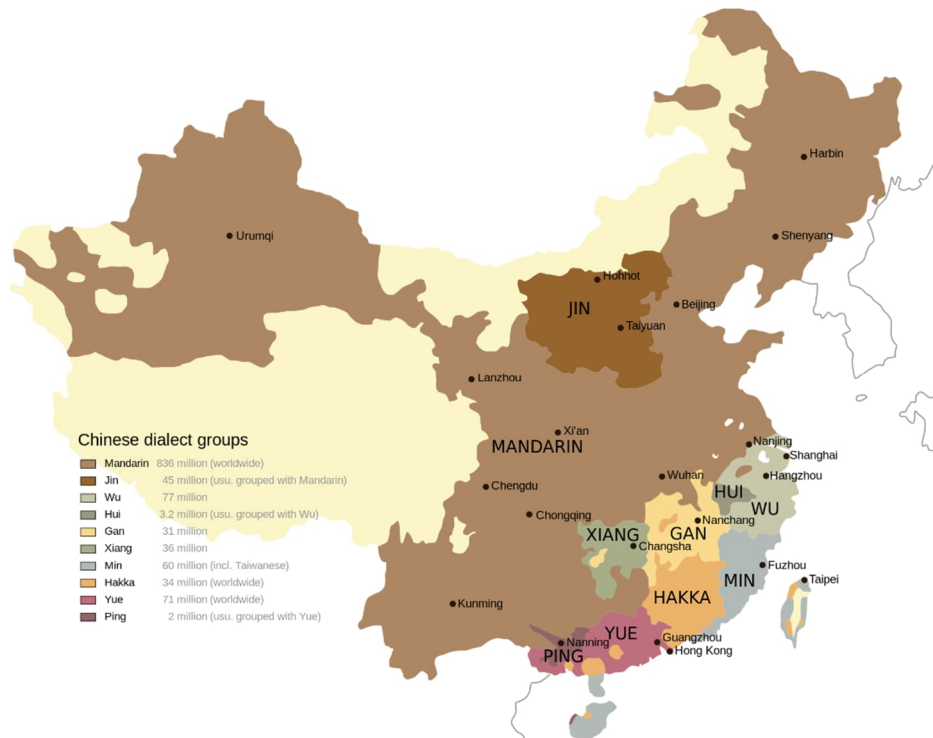## 2. Language systems in China, theories, and previous studies

The significance of language in the context of integration is related to its multiple functions (Tian 1995; Yang 2015; Zhu 2002). Language functions as a means of communication for individuals and possesses the potential to act as a valuable tool for accessing other resources. It can also be considered an investment in human capital. Furthermore, it conveys social norms and values and fosters a sense of belonging and identity. Accordingly, languages as well as accents work as cultural symbols for belonging or foreignness, directly triggering differentiation and discrimination. It has been reflected in multiple important integration theories that take migrants' host language acquisition and fluency and the status of their mother tongue as strong determinants of their integration into the destination society (Alba 1999; Alexander, Edwards, and Temple 2007; Delander et al. 2005; Gómez Cervantes 2023; Kanno 2000; Spolsky and Shohami 1999; Stevens 1992; Tannenbaum 2009). In the following literature review, we will briefly introduce the dual-language system in China. We will then explore the relationship between language and locality, highlighting the similarities between the role of language in international migration and the role of Chinese dialects in shaping the social integration of migrants. We will then focus specifically on China or mostly Shanghai to discuss the differences in the role of languages in shaping integration processes of international migrants and Chinese domestic migrants.

2.1 The dual-language system in China

The language system in China is a nonphonetic unified writing system with a large variety of spoken dialects, which are genetically generated but mutually unintelligible. Figure 1 shows the dialect atlas in China. There are eight major groups of dialects in China, the majority of which concentrate in the southeastern area. Northern dialects are easier for most Chinese people to understand because of their proximity to the official language,

Putonghua, which originated from these dialects. However, the diversity of dialects among other regions remains prevalent. Min and Wu dialects could cause communication barriers even between neighboring towns (Pong 2009). This highlights the linguistic gap that exists between different dialect regions, which is similar to the effect of language barriers on communication. Even in linguistic literature, Chinese dialects are often referred to as 'the Chinese language' (Norman 1988) or are identified as a different language when compared to Putonghua (Francis 2016).

Figure 1:   Language atlas of China

Source: Wurm, S. A., Li, R., and Fu, M. C. (1987). Language atlas of China (in Chinese: 中国语言地图集).

While Putonghua offers most Chinese a way to communicate, it is regional dialects that signify one's origin and identity. Chinese dialects reflect local culture and history. While most parts of China belong to the long tradition of Han Chinese culture with a homogenous staple of folklore, mythologies, and historical identities, it is not accurate to assume that China's dialects do not possess unique cultural significance, as languages

might. Notably, local culture has a strong role, which is reflected in local variations of customs and religious practices (Feuchtwang 2003; Sutton 2007). These local traditions saw a strong revival during the period of economic reforms, with local governments reviving local identities in many forms to attract tourists, such as local historiography or the reconstruction of local sites of historical and often religious significance (Goossaert and Palmer 2011; Kang 2009; Siu 1995).

2.2 Language, locality, and a sense of belonging

The significance of local language proficiency has been verified in sociocultural integration and, eventually, identity transformation. Sociocultural integration refers to what level at which migrants adopt customs, social norms, social relations, and practices of the mainstream in the host society. It results in a decline in social distance between groups and more homogenous social values and practices in society. Studies are predominately about cross-border migrants. For example, Remennick's work (2004) suggested that in Israel, as Russian migrants' proficiency in Hebrew increased, they could cross the social boundaries set by their natural sense of belonging to the previous Russian community and include more local people as their friends. Their cultural and media consumption also diverged from exclusively Russian to more Hebrew-based, suggesting cultural integration. Another study identified English proficiency as the most important predictor of sociocultural integration in Canada (Nakhaie 2020), which is measured by questions such as "I am familiar with Canadian culture and customs" and "I am aware of volunteer opportunities available in the community." Accordingly, language training courses have also been identified as foundations of refugee integration in Scandinavian countries (Valenta and Bunar 2010), whose positive effect on integration again confirms the facilitator role of language in integration proposed by Ager and Stranger (2004). Overall, being proficient in the local language in the host society can generate a stronger sense of belonging, making migrants more similar to locals in their cultural habits and preferences.

We have previously discussed how some of the Chinese local dialects could be treated as different languages. Relatedly, many dialects reflect cultural disparity as they exhibit significant heterogeneity and correspond with local cultures. Taking Shanghai as an example, Shanghai is unique with its modernity being acknowledged to be a mixture of urban and Western (Lu 2002). Shanghaihua is a witness to Shanghai's long migrant and openness history, and the language is a combination of Chinese, English, Hindi, and other languages (Xu 2021). Shanghaihua serves as its metropolitan culture representation with words such as *yangpan* ('foreign tray,' meaning 'fool') and *xianshuimen* (transliteration of 'handsome maid'). In Shanghai, many locals still perceive their dialect

as a marker of their identity and even cultural superiority, and the failure to speak Shanghaihua lead to new migrants being treated as outsiders (Xu 2021).

Some may question the importance of Shanghaihua as the advancement of Putonghua after 1949 because of the great homogenization of language and the diminishing presence of local dialects. However, the existence of a nationwide common language set makes the local dialect even more unique and enables the oral preservation of local dialects (Potter and Potter 2009). Thus, Shanghaihua still enjoys its significance in the identification of locals and nonlocals (Kang 2017; Xu 2021).

2.3 Language and migrants' identity integration

Identity itself, as a sense of self-conception, could enter the social world through only some kind of discourse – in other words, language (Johnstone 1996). Also, the theory of 'doing' identity through language is prevalent across various fields within sociocultural linguistics, highlighting the interplay between language, social practices, and individual agency (Aronsson 1998; Garfinkel 2023; Sidnell 2003; West and Fenstermaker 2002; West and Zimmerman 1987). From a linguistic anthropological perspective, language is a way to signify membership in a given speech community, where language similarities by their nature reflect cultural similarities (Bucholtz and Hall 2005). When people try to speak a language, as Gardner and Lambert identify in their paper (1972), they are displaying a potential desire for integration of that community and further complete identification with the target one.

Identity integration refers to migrants' sense of themselves in relation to the locals. Gordon (1964) suggests identity integration as the gradually developed perception of peoplehood exclusively for the host society. Local identity recognition is often considered a psychological aspect of social integration for migrants and represents a higher stage of integration. Multiple studies have verified the significance of language in identity building (Bucholtz and Hall 2005; Gardner and Lambert 1972) and in triggering identity tension experienced by multilingual speakers or people learning a new language (Block 2006; Norton 2013; Ushioda 2009, 2011; Wang, Zhou, and Wang 2021).

Migrants with higher proficiency in the language of the host society are more likely to display identity integration as they show lower levels of connectedness with their previous ethnic community (Kalbach and Pigott 2005) and higher intention to invest in their cultural capital that would enable their integration into the new society (Darvin and Norton 2015; Norton and Toohey 2011; Peirce 1995). Take Hong Kong as an example. Hong Kong has three main language varieties: Cantonese, Putonghua, and English (Li 2017), where Cantonese is listed by about 89% of the local population as their usual language (Census and Statistics Department 2016). Due to British colonization, before

1997, Cantonese and English were the dominant languages, while after 1997, the end of British governance, Putonghua was formally introduced. People who identify themselves as Hong Kongers are significantly inclined toward Cantonese and less positive toward Putonghua when compared to those who identify themselves as Chinese due to recent political movements (Chan and Chan 2008; Fong et al. 2018; Gu 2011; Gu and Tong 2012; Lai 2011; Li 2017; Y.-T. Li 2024; Oktavianus 2025; Sung 2020). In short, locals speaking in Hong Kong Cantonese reflect the prevailing localist culture, fostering negative sentiments toward the mainland and creating a strong boundary that distances new migrants from the local community. Migrants' Cantonese accents and different demeanors make them direct targets of this bias (Chan and Chan 2008).

We have identified three studies on the role of dialects in shaping migrants' integration into their host society in China. Wang and Fan (2012)'s study finds that understanding the Wuhan dialect doubled the possibility for a migrant to identify themselves as locals, demonstrated by a logit regression model. However, this study suffers from substantial omitted variable issues, as many confounders are correlated with Wuhan dialect proficiency and identity. Lu and Chen (2019) further utilize the instrumental variable (IV) method to study how speaking the Wuhan dialect could influence one's social integration; they find that the command of the local dialect improves social integration in the aspects of intention to integrate, subjectively perceived integration, and objective integration. Chen et al. (2014) study Shanghai and also used the IV strategy. Their study shows that Shanghaihua proficiency significantly increases one's income in the service industry but did not investigate further into identity integration. For our study, we focus on this identity integration to which language might be more connected in the Shanghai context because Shanghaihua is not a strict requirement in the local job market, and there is a declining usage of Shanghaihua among younger generations due to the mandatory use of the country-uniformed Putonghua in education (Xu 2021).

Therefore, when examining regional migration in China, dialects play a significant role as they are similar to languages in terms of communicative efficacy and cultural representativeness. It is reasonable to proxy the influence of language on identity integration to the influence of dialects in the context of domestic migration in China. In recent studies related to dialects in China, the short dialect distance (the linguistic distance between the dialect of the home society and that of the host society) has been argued to give rise to one's labor migration intention in China (Liu, Fang, and Fan 2020) and migrants' social integration in Wuhan, a megacity in the central part of China (Lu and Chen 2019). Accordingly, we expect that migrants who are more proficient in Shanghaihua are more inclined to identify themselves as Shanghainese (Hypothesis 1).

2.4 China's internal migrant integration: Identity, rural–urban disparity, and hukou

A distinct feature of Chinese migrants is the cultural fact that, in traditional Chinese society, individuals typically refer to their ancestral home to denote their place of origin, and they develop a sense of regional identity based on their ancestral home or birthplace beyond simply language proficiency. This regional identity essentially reflects an individual's recognition of their primary affiliation with either a blood-based community or a geographically defined community (Chen, Wu, and Xu 2023). China's traditional conception of home is deeply rooted in Confucian ethics, ancestral rites, and the foundational principle of patrilineality (Zhu 2018). Central to the family structure in traditional Chinese culture is the expectation that family members reside together, collectively contributing to the family's welfare under the authority of a patriarch and adhering to the values of the patrilineal clan (Fei, Hamilton, and Zheng 1992). Consequently, in contrast to the Western emphasis on familiarity, security, and individual value, Chinese culture places greater significance on stability, harmonious relationships among family members, and shared familial aspirations in the construction of the notion of 'home,' often suggesting the birthplace for migrants or even the birthplace of the parents of the migrants as their identity. Some migrants from Ningbo, Zhejiang, would consider Shanghai as a second hometown for 'Ningbo people,' despite their easy integration into Shanghai due to the large number of fellow townsmen and the short cultural distance (Li 2000). Therefore, the integration of migrant workers with the host city is not only affected by their economic gains but more often determined by whether they can establish stable, harmonious relationships with those locals in the host city so that they can construct the notion of home there.

Adding to the complexity of the discourse on migrant identity is the great rural–urban development disparity and its related household registration system. Research on the integration of migrants in China has largely focused on the household registration system, also known as the hukou system, which has functioned as a socioeconomic institution that stratifies the Chinese population along an urban–rural line since its implementation in the late 1950s (Connelly, Roberts, and Zheng 2011; Solinger 1999). Rural migrants in cities could not access social support, such as most state-provided opportunities and state- or employer-subsidized benefits. Their access to jobs, housing, education, and health services is severely restricted compared to those with urban household registration. Although the Chinese government has introduced employment-based social insurance programs to migrant workers since the late 1990s, the participation rates have remained low due to institutional and structural constraints, particularly for those who are not in formal employment (Zhang and Wang 2010). Consequently, the extent to which migrants have integrated into the destination society largely depends on their household registration status (Wang and Fan 2012; Zhang et al. 2020). Especially

for some rural migrants compared to urban ones, working in cities is primarily a means of earning a living, and when they reach old age, they often return to their hometowns due to reasons such as family demands and economic returns (Connelly, Roberts, and Zheng 2011; Wang and Fan 2012; Yuan and Wen 2017; Zhao 2002).

Relatedly, obtaining a Shanghai hukou is extremely competitive. The two main pathways are through talent programs (such as for special talents working in large international or state companies and international top university graduates) or the points-based system, which requires living in Shanghai for at least seven years with stable employment and consistent payment of urban social insurance, along with restrictions on education, job type, and housing. As a result, individuals who have migrated to Shanghai and worked there for decades may have acquired proficient Shanghaihua, though they do not have a Shanghai hukou. On the other hand, some talented individuals could have received their Shanghai hukou upon arrival. Considering the strong connection that many Chinese people feel to their ancestral homes and the competitive nature of obtaining a local hukou in Shanghai, the relationship between Shanghaihua and the establishment of a Shanghainese identity may not be straightforward as in many other cross-border migration cases.

In our study, we further investigated how the significance of Shanghaihua in shaping identity integration depends on whether the migrants have acquired Shanghai hukou. It is possible that for individuals without a Shanghai hukou, their proficiency in Shanghaihua is essential for establishing their sense of belonging and validity in staying in Shanghai. In contrast, those who have obtained a Shanghai hukou do not require additional validation to justify their identity in Shanghai. In short, migrants who have already acquired the local Shanghai hukou, and thus can enjoy the same political or economic rights as locals, would experience an easier time in identity integration compared to their counterparts without local hukou. Accordingly, we expect that the link between Shanghaihua proficiency and identity integration is stronger for migrants without a local hukou (Hypothesis 2).

2.5 The diminishing role of Shanghaihua in identity integration

Most literature on language proficiency and its influences on migrants' social integration has hardly addressed one question: Could the role of language proficiency in shaping social integration change? In the context of most societies, it may be challenging to test this hypothesis because of the small proportion of migrants and because of the great stability of local languages in education, employment, and written forms. However, in Shanghai, the proportion of migrants accounts for 40% of the city's population, among whom approximately 77% are from non-Wuyu areas, and this number is growing

(Shanghai Statistical Bureau 2025). Moreover, the use of Putonghua in education was made mandatory before the mid-1990s in big and coastal cities and before 2000 nationwide (Wang 2016). The large influx of migrants and the improved proficiency in Putonghua among Shanghai residents have made it less common to use Shanghaihua in everyday life with Putonghua being increasingly used.

Putonghua has increasingly substituted Shanghaihua as the medium for communication and has become more popular among younger generations. In fact, 100% of people aged 13 to 20 self-reported being able to speak Mandarin, while only 78.8% reported similar abilities regarding Shanghaihua (Shanghai Bureau of Statistics 2014). This decline in the importance of local dialects may lead more people, especially younger generations – whether they are migrants or local residents – to view Shanghaihua as less significant for shaping local culture and identity. It could also be that younger migrants do not consider Shanghaihua to be important for their establishment of local identity because they are more isolated from Shanghai locals than their more experienced, older migrants, as integration takes time. Younger Shanghainese could also hold less discriminatory views toward those who do not speak Shanghaihua since they grow up in a society with more migrants with fewer possible prejudices and misconceptions due to more interactions (Allport, Clark, and Pettigrew 1954; Pettigrew and Tropp 2006) so that these two groups get along well. Accordingly, we expect that the impact of Shanghaihua proficiency on one's report of being Shanghainese is weaker for the younger generations (Hypothesis 3).

3. Data and methods

This study uses the 2017 wave of data from the Shanghai Urban Neighborhood Survey (SUNS). SUNS covers all districts in Shanghai and constructed a citywide representative sample of the population in Shanghai by asking respondents who have or will stay in Shanghai for more than six months. It employed multistage probability proportional to size sampling. All the subsamples were obtained through three stages: the primary sampling unit was a subdistrict or a town, the second-stage sampling unit was an urban neighborhood community or a village, and the third-stage sampling unit was the household. The sampling units were selected using a systematic sampling method. The SUNS selected 5 to 13 subdistricts/towns from each district, from which two neighborhoods were selected. Then 30 households were selected from each neighborhood, and all eligible members who were aged 15 and above and were Chinese interviewed through face-to-face household visits. The first wave of the survey was completed in 2017 and included 5,102 households and 8,631 adults aged 15 or above.

In our analysis, we select people who were not born with the hukou in Shanghai, have lived in Shanghai or intended to live in Shanghai for more than six months, and define them as 'migrants.' We use a broad definition of migrants due to the long-term process of language acquisition, cultural adaption, and identity formation. Given the significant challenges in obtaining a Shanghai hukou, we are also interested in whether proficiency in Shanghaihua remains important once an individual has acquired local hukou. In our sample, among those who have obtained a local hukou, 38% of respondents reported their proficiency in Shanghaihua as not proficient, 33% rated it as fair, and only 30% considered it to be proficient. Among the whole sample of the survey, there are 3,820 migrants. After dropping 485 samples with missing values in the variables of use, we get the final sample of 3,335 observations.

3.1 Variables

3.1.1 Dependent variables

Shanghainese identity. To capture migrants' integration in the aspect of identity recognition, we choose a question designed specifically for such evaluation. The question is "Do you agree or disagree that you are a Shanghainese?" The answer is measured by a seven-point Likert scale ranging from 1 "Strongly disagree" to 7 refers to "Strongly agree," indicating a higher level of identity integration with higher scores.

3.1.2 Key independent variables

Shanghaihua proficiency. This linguistic capability is measured through a self-reported question in SUNS, asking, "Your Shanghai dialect proficiency is?" The answer is a five-point Likert scale ranging from 1 "Not proficient at all," 2 "Not proficient," 3 "Fair," 4 "Proficient," and 5 "Very proficient."

Local hukou status. For interaction analysis, we included whether the respondent is currently holding a local hukou and interact it with Shanghaihua proficiency. We do not include the local hukou status in our main model, which aims to establish the relationship between Shanghaihua proficiency and Shanghai identity, for two reasons. First, local hukou status could be influenced by both proficiency in Shanghaihua and local identity, making it a collider. Second, even if one's sense of being Shanghainese would be only enhanced after acquiring local hukou, including this variable would mediate the relationship between Shanghaihua proficiency and self-identity as a Shanghainese. Therefore, we include local hukou status in only the interaction models to examine

whether the effects of Shanghaihua differ for individuals with and without a Shanghai hukou.

Age. The age variable is an automatically generated value based on respondents' answers of their birth year and month. Age will additionally be used for interaction analysis.

3.1.3 Control variables

We also control some individual characters that are correlated with both their Shanghaihua proficiency and integration. We adopted the birth hukou type as a control. It has two types (0 = nonagricultural (reference group, including the residential hukou that is partially launched in several cities for removing the previous differentiating categories), 1 = agricultural). Other variables include basic demographic factors: gender (0 = female, 1 = male), educational level (below high school (reference group), high school, associate degree, bachelor's degree and above), and marital status (whether married with 1 indicating currently married). We also control for migrants' living experience with Shanghai, using whether one came to Shanghai for more than half a year before the age of 14 as a proxy. Besides, we include personal income and one's occupational type. The total income is calculated as the sum of one's wage income, economic income, and other sources of income. The occupation type refers to the current occupation type if the individual was employed when surveyed and the type of the last job if one was not.

3.2 Analytical strategy

Aside from the simple OLS regression predicting the identity recognition variable, we employ instrumental variables to further explore the causal relationship between Shanghaihua proficiency and the interested outcome.

Establishing the link between language proficiency and integration may suffer from serious confounding issues. For example, while language skills might imply characteristics associated with productivity, such as quality of education, both are correlated with their integration (Christofides and Swidinsky 2010; Dustmann and Soest 2001; Pendakur and Pendakur 2002). To address the endogeneity problems referring to the unobserved variables associated with language skills and the dependent variables, such as personalities or lifestyle preferences, this research relies on the interaction term (COMP_Diadis) of the respondent's comprehension ability (COMP), a score given by the interviewer based on their interviewing experiences with the respondent and the

dialect distance between one's original province and Shanghai as the instrumental variable (IV). To better capture the relationship between IV and Shanghaihua proficiency, we reverse the original score of comprehension ability so that a lower value of IV would suggest a higher possibility of mastering Shanghaihua. An effective IV needs to satisfy two requirements – the instrument relevance and exclusion restriction.

The relevance condition is met by the intuitively close linkage among comprehension ability, the language environment that the migrant used to live in, and language acquisition. Comprehension, as a vital procedure in learning a language and also a part of cognitive abilities, has proved its significance in language acquisition with multiple related studies (Clark 2011; Clark and Hecht 1983; Ellis 2004; Goldman-Eisler 1951; Krashen 1985; Segalowitz 2010). To specify the importance of one's comprehension abilities, comprehension is often considered to be "at the heart of the language acquisition process" (Ellis 2004: 103) and the process that precedes production (Clark 2011; Ellis 2004). Despite the relatively recent discussion on the asymmetries between comprehension and production, where production might take the lead, the problem is mostly limited to children acquiring a new language instead of teenagers or adults (Hendriks and Koster 2010; Ünal and Papafragou 2016). Thus, the direct relationship between one's comprehension ability and one's Shanghaihua proficiency is intuitive to validate.

The other part of the IV, the dialect distance between the sending province and Shanghai, is applied as an indicator of the language barrier that captures the heterogeneity of migrants' Shanghaihua acquisition ability. Studies have proved that linguistic distance is among the most important determinants of one's acquisition of a new language (Cenoz 2001; Isphording and Otten 2014; Lindgren and Muñoz 2013). In this study, we adopted the formula proposed by Liu et al. (2020) to calculate this dialect distance between one's sending province and Shanghai, which was also used by Wu and Yu (2023) to account for dialect distance. (For detailed calculations of each province, please see Appendix A.)

The exclusion restriction is met if the effect of our IVs on identity integration work only through Shanghaihua proficiency. Comprehension ability, as a part of cognitive abilities, might contribute to one's educational or occupational achievement, enabling the migrants to socialize more with the local Shanghainese people at work and thus gain a higher level of identity integration. In this case, to avoid possible channels that our IV is related to our interested outcome, we control the individual's total income and occupation in addition to the educational level to deal with this possible connection between comprehension abilities and one's socioeconomic integration and consequently identity integration.

The following equations demonstrate the identification using the IV–2SLS (two-stage least squares) model:

$$\text{SHNESE}_i = \alpha_0 + \theta_1 \text{COMP\_Diadis}_i + \theta_2 C_i + \varepsilon_i \tag{1}$$

$$\text{SHID}_i = \alpha_1 + \beta_1 \text{SHNESE}_i + \beta_2 C_i + \varepsilon_{i,} \tag{2}$$

where Equation 1 denotes the first stage of the IV–2SLS model, estimating the coefficient of Shanghaihua proficiency ($SHNESE_i$) is exogenously explained using the interaction result ($COMP\_Diadis_i$) of one's rated comprehension ability ($COMP_i$) and the dialect distance between the sending province and Shanghai. Equation 2 signifies the second-stage regression, where Shanghainese identity recognition ($SHID_i$) is explained by estimated values of the Shanghaihua proficiency ($SHNESE_i$) in the first stage. C denotes a vector of control variables. $COMP\_Diadis_i$ is excluded from Equation 2, alternatively denoted as the excluded instrument. Variables α_0, α_1, β_1, and β_k are the unknown parameters to be estimated, and ε_i is the error term.

4. Results

4.1 Descriptive results

Table 1 shows the descriptive statistics of our selected sample. The mean value of Shanghainese identity reported by the migrants is about 3, suggesting mild disagreement with the statement "I identify myself as Shanghainese." The average score of Shanghaihua proficiency is around 1.94, which indicates that most migrants identify themselves as poor at Shanghaihua and corresponds with the fact that dialects construct communication barriers. The average age of our sample is around 40, and 49% of the sample are males. Most of them are married, taking up 81% of the whole sample. Within this migrant group, 24.4% were with nonagricultural hukou when they were born, while most of them (75.6%) possessed an agricultural hukou, and mostly had never been to Shanghai for more than half a year before the age of 14. Regarding the education level, 53.8% obtained an educational degree below high school and about 31% have an associate degree or higher.

Table 1: **Descriptive statistics (*N* = 3,335)**

Variable	Mean	(SD)
Shanghainese identity	3.027	(1.943)
Shanghaihua proficiency	1.942	(1.037)
Comprehension ability	5.59	(1.393)
Dialect distance	2.614	(0.671)
Age	39.963	(14.781)
Hukou status at birth		
Nonagricultural	24.4%	
Agricultural	75.6%	
Local hukou		
Yes	19.6%	
No	80.4%	
Been to Shanghai before 14		
Yes	10.6%	
No	89.4%	
Male	48.9%	
Marital status		
Single	18.9%	
Married	81.1%	
Educational level		
Below high school	53.8%	
High school	15.1%	
Associate degree	10.6%	
Bachelor's degree and above	20.4%	
Occupation type		
Managers of national organizations	3.9%	
Professionals	21%	
Administrative, security, and emergency personnels	5.8%	
Business and service	31.4%	
Agriculture, forestry, husbandry, fishery, and water-related industry	3.3%	
Production, transportation, equipment operators, and related personnel	20.9%	
Military officers	0.2%	
Others	13.5%	
Numbers of observations	3,335	

4.2 Regression results

We proceed to the influence of Shanghaihua proficiency on migrants' Shanghainese identity recognition with estimates drawn from OLS and IV–2SLS regressions. The OLS model in Table 2, Panel A, confirms the positive relationship between Shanghaihua proficiency and migrants' identity. The following 2SLS model has shown consistent results, with the estimated coefficient of Shanghaihua proficiency changed from 0.530 ($p < 0.001$) to 1.165 ($p < 0.001$). The increase in this coefficient could be partly because IV methods produce estimations based on the local average treatment effect. The OLS estimates measure the average treatment effect of the whole population while IV estimates the effect of better Shanghaihua proficiency only for the population whose

choice of Shanghainese identity was affected by the instrument. The larger IV estimate could be due to the heterogeneity in the studied population. Also, the inclusion of an IV inevitably adds noise to a model, which is reflected by the increased standard deviation and results in a larger confidence interval. The Cragg-Donald Wald F statistic of 106.38 indicates a strong selected IV and valid IV estimation results, as it exceeds the thumb rule of 10. Panel B in Table 2 shows the detailed results of the first stage estimates.

From the results, Hypothesis 1 is supported. These findings suggest that higher Shanghaihua proficiency will lead to a higher level of Shanghainese identity recognition.

Considering the other variables, possessing an agricultural hukou at birth does not show signs of strong correlation with recognizing oneself as Shanghainese compared to those with a nonagricultural hukou at birth. We will continue to examine the role of local hukou acquisition in the following heterogeneous analysis. Moreover, achieving the educational level of a bachelor's degree and above is associated with a higher score of one's Shanghainese identity when compared to those who haven't been to high school, suggesting a coherence with Shanghai's elite image.

Table 2: **Shanghaihua proficiency and Shanghainese identity recognition (OLS and IV–2SLS)**

Panel A

VARIABLES	(1) Shanghainese identity-OLS	(2) Shanghainese ID-IV-2SLS
Shanghaihua proficiency	0.530	1.165
	(0.032)	(0.190)
	0.000	0.000
Age	0.030	0.018
	(0.002)	(0.004)
	0.000	0.000
Hukou status at birth (Reference: Nonagricultural)		
Agricultural hukou	−0.093	0.002
	(0.079)	(0.087)
	0.238	0.985
Been to SH before 14 (Reference: Yes)		
No	−0.815	−0.188
	(0.104)	(0.215)
	0.000	0.381
Male	−0.130	−0.059
	(0.062)	(0.069)
	0.037	0.394
Married	−0.057	−0.113
	(0.080)	(0.086)
	0.477	0.189
Total income (logged)	0.019	−0.006
	(0.014)	(0.017)
	0.186	0.738

Table 2:   (Continued)

Panel A

VARIABLES	(1) Shanghainese identity-OLS	(2) Shanghainese ID-IV-2SLS
Occupation type (Reference: Managers of national organizations)		
Professionals	–0.104	–0.034
	(0.165)	(0.175)
	0.528	0.848
Administrative, security, and emergency personnels	–0.233	–0.182
	(0.195)	(0.207)
	0.233	0.379
Business and service	–0.283	–0.250
	(0.166)	(0.175)
	0.088	0.154
Agriculture, forestry, husbandry, fishery, and water-related industry	–0.485	–0.188
	(0.229)	(0.257)
	0.034	0.466
Production, transportation, equipment operators, and related personnel	–0.077	0.041
	(0.174)	(0.186)
	0.658	0.824
Military officers	–0.398	–0.400
	(0.663)	(0.700)
	0.548	0.568
Others	–0.078	–0.012
	(0.177)	(0.188)
	0.661	0.951
Educational level (Reference: Below high school)		
High school	0.155	0.053
	(0.088)	(0.098)
	0.080	0.590
Associate degree	0.529	0.324
	(0.107)	(0.128)
	0.000	0.011
Bachelor's degree and above	0.640	0.595
	(0.096)	(0.103)
	0.000	0.000
Constant	1.517	0.300
	(0.256)	(0.449)
	0.000	0.504
Instrumental variable		COMP_Diadis
Cragg-Donald Wald F statistic		106.38
R^2	0.236	
Adjusted R^2	0.232	
Observations	3,335	3,335

Source: Shanghai Urban Neighborhood Survey (2017).
Note: Standard errors are in parentheses and clustered at the individual level. Standard errors in the second row, p-values in the third row.

Table 2:   (Continued)

Panel B first stage

VARIABLES	
COMP_Diadis	−0.044
	(0.004)
	0.000
Age	0.020
	(0.001)
	0.000
Hukou status at birth (Reference: Nonagricultural)	
Agricultural hukou	−0.128
	(0.042)
	0.002
Been to SH before 14 (Reference: Yes)	
No	−0.943
	(0.054)
	0.000
Male	−0.127
	(0.034)
	0.000
Married	0.093
	(0.043)
	0.031
Total income (logged)	0.034
	(0.008)
	0.000
Occupation type (Reference: Managers of national organizations)	
Professionals	−0.086
	(0.089)
	0.331
Administrative, security, and emergency personnels	−0.068
	(0.105)
	0.518
Business and service	−0.008
	(0.089)
	0.930
Agriculture, forestry, husbandry, fishery, and water-related industry	−0.335
	(0.124)
	0.007
Production, transportation, equipment operators, and related personnel	−0.127
	(0.094)
	0.173
Military officers	0.074
	(0.357)
	0.836
Others	−0.045
	(0.096)
	0.635
Educational level (Reference: Below high school)	
High school	0.094
	(0.048)
	0.050

Table 2:   (Continued)

Panel B first stage

VARIABLES	
Associate degree	0.220
	(0.058)
	0.000
Bachelor's degree and above	−0.041
	(0.053)
	0.438
Constant	2.143
	(0.136)
	0.000
R^2	0.221
Adjusted R^2	0.217
Observations	3,335

Source: Shanghai Urban Neighborhood Survey (2017).
Note: Standard errors are in parentheses and clustered at the individual level. Standard errors in the second row, p-values in the third row.

4.3 The moderating roles of Shanghai hukou and age

We then examine the moderating role of acquiring a local hukou and age in shaping the relationship between Shanghaihua proficiency and integration, applying the coherent OLS method. We did not conduct 2SLS because the constructed instrument for the interaction term with the previous IV did not pass the test for a strong instrument variable, falling short on its relevance. Thus, we cannot justify its good use. Also, the results from the 2SLS in Table 2 are larger than those of the OLS. We therefore prefer a more general coefficient estimation, which may also be a more conservative one. Column 1 in Table 3 presents the baseline model, which includes the estimates from the previous table. Column 2 represents the results of this interaction model.

After introducing the interaction of Shanghaihua proficiency and local hukou status, the results show that compared to those without a local hukou, speaking Shanghaihua might not improve one's identity integration as much if the migrant has a local hukou, although this statement falls short on its statistical significance. There is no evidence supporting that the link between Shanghaihua proficiency and migrants' integration differ depending on one's local hukou status. Hypothesis 2 is not supported.

Table 3: **The interactions with hukou and age and Shanghainese identity recognition (OLS regressions)**

VARIABLES	(1) Hukou baseline	(2) Hukou interaction	(3) Age baseline	(4) Age interaction
Shanghaihua proficiency	0.383	0.407	0.530	0.213
	(0.031)	(0.038)	(0.032)	(0.084)
	0.000	0.000	0.000	0.011
Local hukou status (Reference: Non-Shanghai)				
Shanghai hukou	1.670	1.854		
	(0.0926)	(0.181)		
	0.000	0.000		
Shanghaihua proficiency * Local hukou		-0.077		
		(0.066)		
		0.238		
Age	0.012	0.012	0.030	0.014
	(0.002)	(0.002)	(0.002)	(0.005)
	0.000	0.000	0.000	0.003
Shanghaihua proficiency *Age			0.007	
			(0.002)	
			0.000	
R^2	0.304	0.304	0.236	0.240
Adjusted R^2	0.300	0.300	0.232	0.235
Observations	3,335	3,335	3,335	3,335

Source: Shanghai Urban Neighborhood Survey (2017).
Notes: Standard errors are in parentheses and clustered at the individual level. Standard errors in the second row, p-values in the third row. Control variables are the same as in Table 2.

The moderating role of age in shaping the relationship between Shanghaihua proficiency and integration is presented in Column 3 of Table 3, with the expectation that Shanghaihua proficiency is more important for older people. The results have supported our Hypothesis 3. However, this situation might change in the future since both the people and the local government have come to realize the importance of dialect preservation: 83.99% of respondents believe that it is "very necessary" to protect local Wu dialects, and 14.8% of respondents choose "to a certain level of necessity" (Zhang and Fan 2017). From the policy perspective, the Chinese government has launched the China Language Resources Protection Project since 2015 to record and preserve local dialects. Meanwhile, the Shanghai government has encouraged Shanghaihua courses, which led to more than 200 preschools having pilot programs about Shanghaihua teaching by 2016 (Xia and Shen 2019). With a future young generation having better Shanghaihua proficiency, the importance of learning Shanghaihua for migrants' integration into the host society might be more significant.

4.4 Robustness checks

Considering that lack of details by using whether one has been to Shanghai for more than half a year as a proxy for one's previous encounter with the Shanghaihua environment, we substitute it with the variable of whether one has any educational experience in Shanghai to show his or her previous living experience in Shanghai. The results stay robust (see Table B-1, Appendix B), with the affirmation of Shanghaihua proficiency's facilitation effect on one's identity integration and the consistent confirmation that previous experience with Shanghai could be correlated with both a higher possibility of deeper identity integration and better command of Shanghaihua.

Regarding the limitation of self-reported Shanghaihua proficiency and the availability of the data, we substitute whether using Shanghaihua at work or for social purposes as an indicator of one's Shanghaihua proficiency. The results show consistent proof of how Shanghaihua proficiency could lead to a higher level of identity integration (see Table B-2, Appendix B).

Some may consider migrants from Zhejiang and Jiangsu provinces as different from those of other origins since these two areas have witnessed prevailing neighboring influence from Shanghai. Some might argue that those people feel more toward their hometown even though they can speak perfect Shanghaihua. Others might think that these people would experience easier identity integration into Shanghai, considering their similar cultural and dialectal backgrounds. We apply the OLS interaction of Shanghaihua proficiency with whether they come from these two provinces and find that the difference is not significant given the current dataset (see Table B-3, Appendix B).

Besides, local hukou acquisition and homeownership in Shanghai have long been considered important events during one's integration. Although these two events are supposed to occur during the process of one's identity integration, we still add them into the model to see if it would alter the results (see Table B-4, Appendix B). We could see that the relationship between Shanghaihua proficiency and migrant identity integration remains.

5. Discussion and conclusion

Internal migration in China has undergone significant changes due to the reform of the hukou system, the influx of skilled migrant workers into large cities, and the country's unique dual-language system. In this system, Putonghua serves as the unified language across China, while local dialects reflect individuals' sense of belonging and identity. The role of these dialects in migrant integration and identity recognition within the context of China's domestic migration remains underexplored. This study examines how

the host society's language influences the identity of migrants in Shanghai, a megacity with a rich dialectal history. Utilizing data from the Shanghai Urban Neighborhood Survey 2017, we employ an instrumental variable approach to assess the relationship between proficiency in Shanghaihua and migrant assimilation regarding identity recognition.

Our findings indicate that higher proficiency in Shanghaihua positively correlates with self-identification as Shanghainese, a result that is robust across various indicators. However, when considering interaction effects with age, we observe that the impact of Shanghaihua proficiency on identity integration diminishes among younger respondents. This finding suggests that the significance of the local dialect in shaping the sense of Shanghainese has weakened over time. Interestingly, we find no strong evidence that the effect of Shanghaihua on self-identification as Shanghainese differs between those who have acquired a local hukou and those who have not. The sign of the estimated coefficient is negative, indicating a weakened importance of local dialect for those who have Shanghai hukou. However, this estimation is not statistically proven. This finding suggests that proficiency in Shanghaihua may still serve as a significant barrier for migrants seeking to truly feel like a Shanghainese, regardless of their administrative recognition as local residents. A similar finding is noted in Xu's interviews with Shanghai migrants (2021), where the connection between the local identity and the vernacular is so entrenched that nonspeakers will always be considered Waidiren (outlanders), making the administrative designation of local hukou irrelevant. The term 'New Shanghairen' is even coined by the mainstream media to refer to migrants with Shanghai hukou, signaling a subtle but existing distinction between them and natives. For most migrants, they acquire the local hukou mainly for the social support and living environment, just as widely researched by the migration intention studies (e.g., Zhou et al. 2023). Therefore, the self-identification as Shanghainese reflects more of a sense of cultural adaptation and local belonging, which differs from the widely studied concept of economic integration.

Our work has some limitations. First, our paper focuses exclusively on the city of Shanghai. Shanghai is unique in China, being one of the four tier-1 cities with a long history of migration driven by trade and import-export activities. The identity of Shanghai has evolved over time, along with its local dialect. Findings from other major cities in China may differ. For instance, in Beijing, the local dialect is Putonghua, though it is spoken with a distinct Beijing accent. Self-identification as a Beijinger may rely more on administrative recognition than on language factors since people could not directly tell one's locality via the language. Hong Kong could be another illustration of dialect dominance, with even speaking Guangdong-accented Cantonese could be discriminated against by locals, hindering the identity integration of all migrants, regardless of their acquisition of a Hong Kong passport. More studies about different locations in China and around the world should be introduced to the discussion on this topic, and the situation

could differ based on the original unintelligibility and current popularity of local dialects. Second, although the use of the instrumental variable enhanced our confidence about the causal relationship between Shanghaihua proficiency and identity integration, there are still some critiques of using this method. Among these, the most important one, considering the generalizability of the results, should be that IV looks at a specific subpopulation. In our study, this subpopulation consists of those whose Shanghaihua proficiency is altered by the interaction between one's comprehension ability and the dialect distance between their original province and Shanghai. In response, we have consulted the previous literature in linguistics and have found that multiple studies have stressed the relationship between one's comprehension ability (Hendriks and Koster 2010; Ünal and Papafragou 2016), linguistic distance (Cenoz 2001; Isphording and Otten 2014; Lindgren and Muñoz 2013), and one's language acquisition. Besides, the evaluation of Shanghaihua proficiency is comparatively subjective, which can be refined in future research as researchers could adopt more professional linguistic measurements. Moreover, due to the limitation of the dataset, we could construct IV as the interaction between only a province-level dialect distance and the individual level, which could be improved if we could acquire county-level information of respondents since the original division of dialect regions is based on county. Also, despite our best efforts to meet the exclusion restriction by controlling possible channels through which IV could directly influence one's identity integration, there could still be possibilities that they are connected through unobserved ways. Finally, we would also like to note that integration is dynamic. It is important to highlight the cultural diversity brought by migrants into the host societies as well as the great benefits that the destination societies could gain from a younger population and expanded labor force market. Our analysis is limited by the cross-sectional nature of the data, which prevents us from distinguishing whether the declining significance of Shanghaihua proficiency in shaping migrants' identity is due to birth cohorts, age, or period effects. In an increasingly open and diverse society, both migrants and local Shanghainese may view Shanghaihua as less important in defining one's identity. Future studies could further investigate whether this decline is also influenced by age effects, as older individuals may become more conservative due to increased socialization across different groups.

Our finding regarding the significance of Shanghaihua in shaping migrants' self-identification as Shanghainese has important policy implications. As the government actively promotes the 'citizenization' of migrants in their host society, the substantial challenges of learning Shanghaihua pose particular difficulties for a city like Shanghai. Most migrants come from non-Wu regions, where their mother tongues differ significantly from Shanghaihua. In addition, the strong attachment to their original hometown or roots further complicates migrants' sense of belonging. Previous studies have shown that even children of nonskilled migrant workers, who were born and raised

in Shanghai, often do not feel they are Shanghainese. A poignant expression of this sentiment is reflected in one statement of a migrant's child: "I am not a rural child; I am a child of the city, but I am not a child of Shanghai" (Xiong 2009). This statement clearly highlights Shanghai's unique status in contemporary urban China. On the other hand, it is interesting to note that as more migrants who do not speak Shanghaihua or any Wu dialect come to Shanghai, and as younger generations of Shanghainese struggle with their proficiency in Shanghaihua, the significance of Shanghaihua in shaping one's sense of locality appears to be declining over time.

Considering the China Language Resources Protection Project since 2015 and local preservation programs, we might witness some by-products as they may create a language environment in which migrants experience more direct exclusiveness and a harder integration process. Meanwhile, the study reaffirms that migrants take dialect distances and the popularity of the local dialects into consideration while they choose their destinations, as this could indicate the cultural distances as well as one's experience of hardship when integrating. Consider Shenzhen as an example (Wang et al. 2011), where the migrant population takes up more than 90%. Due to the lack of strong local dialects and exclusive local communities, migrants have formed a city where Shenzhen identity is easy for every migrant to identify with, illustrated by the quote that could be seen both on the street and also in official government reports: 'As long as you are here, you are Shenzhenese' (Qin 2025). Contributing to the research on international migrant integration, this study suggests taking language prevalence into consideration since, as migrants increase, the dominant role of the original local language might subside, such as the growing Spanish-speaking community in the United States, Quebec in Canada, and Grisons in Switzerland. Future work can take advantage of these changing language dynamics to understand the evolving meaning of locality.

Drawing inspiration from new migration patterns and the national policy promoting the 'citizenization' of migrants in major cities, we anticipate that the integration of migrants – particularly regarding their identity and cultural adaptation – will become increasingly important. Using Shanghai as a case study, we examine the role of local dialects in shaping migrants' sense of identity. This study is one of the few that focuses on identity integration from a cultural and linguistic perspective to understand the potential and process of identity integration of domestic migration in China.

6. Funding

References

Ager, A. and Stranger, A. (2004). Indicators of integration: Final report. Home Office Development and Practice Report Home Office.

Alba, R. (1999). Immigration and the American realities of assimilation and multiculturalism. *Sociological Forum* 14(1): 3–25. doi:10.1023/A:1021632 626811.

Alba, R. and Nee, V. (1997). Rethinking assimilation theory for a new era of immigration. *International Migration Review* 31(4): 826–874. doi:10.1177/01979183970 3100403.

Alexander, C., Edwards, R., and Temple, B. (2007). Contesting cultural communities: Language, ethnicity and citizenship in Britain. *Journal of Ethnic and Migration Studies* 33(5): 783–800. doi:10.1080/13691830701359223.

Allport, G.W., Clark, K., and Pettigrew, T.F. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Appau, S., Churchill, S.A., and Farrell, L. (2019). Social integration and subjective wellbeing. *Applied Economics* 51(16): 1748–1761. doi:10.1080/00036846.2018. 1528340.

Aronsson, K. (1998). Identity-in-interaction and social choreography. *Research on Language and Social Interaction* 31(1): 75–89. doi:10.1207/s15327973rlsi 3101_5.

Berkman, L.F. (1995). The role of social relations in health promotion. *Psychosomatic Medicine* 57(3): 245–254. doi:10.1097/00006842-199505000-00006.

Bian, Y., Zhang, L., Yang, J., Guo, X., and Lei, M. (2015). Subjective wellbeing of Chinese people: A multifaceted view. *Social Indicators Research* 121(1): 75–92. doi:10.1007/s11205-014-0626-6.

Block, D. (2006). *Multilingual identities in a global city: London stories*. Houndmills, Basingstoke, Hampshire, New York: Palgrave Macmillan.

Bucholtz, M. and Hall, K. (2005). Language and identity. In: Duranti, A. (ed.). *A companion to linguistic anthropology*. 1st ed. Hoboken, NJ: Wiley: 369–394. doi:10.1002/9780470996522.ch16.

Cenoz, J. (2001). Chapter 1. The effect of linguistic distance, l2 status and age on cross-linguistic influence in third language acquisition. In: Cenoz, J., Hufeisen, B., and

Jessner, U. (eds.). *Cross-linguistic influence in third language acquisition*. Bristol, PA: Multilingual Matters: 8–20. doi:10.21832/9781853595509-002.

Census and Statistics Department (2016). *2016 population by-census*. Census and Statistics Department Hong Kong.

Chan, K. and Chan, W. (2008). *Hong Kong's professional immigrants from Mainland China and their strategies of adaptation*. Hong Kong: Department of Sociology Hong Kong Baptist University.

Chen, W., Wu, X., and Xu, D. (2023). Does the Hukou system reform promote migrants' social integration? Emperical evidence from Shanghai. *Journal of Cities and Society* 2(1): 86–107.

Chen, Y. and Wang, J. (2015). Social integration of new-generation migrants in Shanghai China. *Habitat International* 49: 419–425. doi:10.1016/j.habitatint.2015.06.014.

Chen, Z., Lu, M., and Xu, L. (2014). Returns to dialect: Identity exposure through language in the Chinese labor market. *China Economic Review* 30:27–43. doi:10.1016/j.chieco.2014.05.006.

Christofides, L.N. and Swidinsky, R. (2010). The economic returns to the knowledge and use of a second official language: English in Quebec and French in the rest-of-Canada. *Canadian Public Policy* 36(2): 137–158. doi:10.3138/cpp.36.2.137.

Clark, E.V. (2011). *Lexicon in acquisition*. Cambridge: Cambridge University Press.

Clark, E.V. and Hecht, B.F. (1983). Comprehension, production, and language acquisition. *Annual Review of Psychology* 34(1): 325–349. doi:10.1146/annurev.ps.34.020183.001545.

Cohen, S. (1988). Psychosocial models of the role of social support in the etiology of physical disease. *Health Psychology* 7(3): 269–297. doi:10.1037/0278-6133.7.3.269.

Connelly, R., Roberts, K., and Zheng, Z. (2011). The settlement of rural migrants in urban China – some of China's migrants are not 'floating' anymore. *Journal of Chinese Economic and Business Studies* 9(3): 283–300. doi:10.1080/14765284.2011.592356.

Darvin, R. and Norton, B. (2015). Identity and a model of investment in applied linguistics. *Annual Review of Applied Linguistics* 35: 36–56. doi:10.1017/S0267190514000191.

Delander, L., Hammarstedt, M., MÅnsson, J., and Nyberg, E. (2005). Integration of immigrants: The role of language proficiency and experience. *Evaluation Review* 29(1): 24–41. doi:10.1177/0193841X04270230.

Du, Z. (2015). *Chinese language demystified*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Dustmann, C. and Soest, A.V. (2001). Language fluency and earnings: Estimation with misclassified language indicators. *Review of Economics and Statistics* 83(4): 663–674. doi:10.1162/003465301753237740.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning* 54(2): 227–275. doi:10.1111/j.1467-9922.2004.00255.x.

European Commission: Directorate-General for Migration and Home Affairs (2013). Using EU indicators of immigrant integration – final report for Directorate-General for Home Affairs. Luxembourg: Publications Office. doi:10.2837/34091.

Fei, X., Hamilton, G.G., and Zheng, W. (1992). *From the soil: The foundations of Chinese society*. Oakland, CA: University of California Press. doi:10.1525/9780520912489.

Feuchtwang, S. (2003). *Popular religion in China*. Milton Park: Routledge. doi:10.4324/9780203392850.

Fong, E., Li, J.X., and Chan, C.C. (2018). Mainland migrants in Hong Kong. In: Lui, T.-l., Chiu, S.W.K., and Yep, R. (eds.). *Routledge handbook of contemporary Hong Kong*. Milton Park: Routledge: 275–289. doi:10.4324/9781315660530-17.

Francis, N. (2016). Language and dialect in China. *Chinese Language and Discourse. An International and Interdisciplinary Journal* 7(1): 136–149. doi:10.1075/cld.7.1.05fra.

Gardner, R.C. and Lambert, W.E. (1972). *Attitudes and motivation in second-language learning*. Rowley, MA: Newbury House Publishers.

Garfinkel, H. (2023). Studies in ethnomethodology. In: Longhofer, W. and Winchester, D. (eds.). *Social theory re-wired*. New York: Routledge: 58–66. doi:10.4324/9781003320609-8.

Goldman-Eisler, F. (1951). The measurement of time sequences in conversational behaviour. *British Journal of Psychology. General Section* 42(4): 355–362. doi:10.1111/j.2044-8295.1951.tb00314.x.

Gómez Cervantes, A. (2023). Language, race, and illegality: Indigenous migrants navigating the immigration regime in a new destination. *Journal of Ethnic and Migration Studies* 49(7): 1610–1629. doi:10.1080/1369183X.2021.1968294.

Goossaert, V. and Palmer, D.A. (2011). *The religious question in modern China*. Chicago, IL: University of Chicago Press. doi:10.7208/chicago/9780226304182.001.0001.

Gordon, M.M. (1964). *Assimilation in American life: The role of race, religion and national origins*. New York: Oxford University Press. doi:10.2307/3510186.

Gu, M.M. (2011). 'I am not qualified to be a Honkongese because of my accented Cantonese': Mainland Chinese immigrant students in Hong Kong. *Journal of Multilingual and Multicultural Development* 32(6): 515–529. doi:10.1080/01434632.2011.614350.

Gu, M.M. and Tong, H.K. (2012). Space, scale and languages: Identity construction of cross-boundary students in a multilingual university in Hong Kong. *Language and Education* 26(6): 501–515. doi:10.1080/09500782.2012.663553.

Hendriks, P. and Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua* 120(8): 1887–1897. doi:10.1016/j.lingua.2010.02.002.

Holmes, S.M. (2012). The clinical gaze in the practice of migrant health: Mexican migrants in the United States. *Social Science and Medicine* 74(6): 873–881. doi:10.1016/j.socscimed.2011.06.067.

House, J.S., Landis, K.R., and Umberson, D. (1988). Social relationships and health. *Science* 241(4865): 540–545. doi:10.1126/science.3399889.

Huang, G., Zhang, H., and Xue, D. (2018). Beyond unemployment: Informal employment and heterogeneous motivations for participating in street vending in present-day China. *Urban Studies* 55(12): 2743–2761. doi:10.1177/0042098017722738.

Huang, X., Liu, Y., Xue, D., Li, Z., and Shi, Z. (2018). The effects of social ties on rural-urban migrants' intention to settle in cities in China. *Cities* 83: 203–212. doi:10.1016/j.cities.2018.06.023.

Isphording, I.E. and Otten, S. (2014). Linguistic barriers in the destination language acquisition of immigrants. *Journal of Economic Behavior and Organization* 105: 30–50. doi:10.1016/j.jebo.2014.03.027.

Johnstone, B. (1996). *The linguistic individual: Self-expression in language and linguistics*. Oxford University Press. doi:10.1093/oso/9780195101843.001.0001.

Kalbach, M.A. and Pigott, B.S. (2005). Language effects on ethnic identity in Canada [Canadian Ethnic Studies Association 7th biennial conference, 2003]. *Canadian Ethnic Studies* 37(2): 3.

Kang, L. (2017). 'Who is the foreigner': Study on people's local identity and its factor in big cities – Shanghai as an example. *Journal of Huazhong University of Science and Technology (Social Science Edition)* 31(01): 58–67.

Kang, X. (2009). Two temples, three religions, and a tourist attraction: Contesting sacred space on China's ethnic frontier. *Modern China* 35(3): 227–255. doi:10.1177/0097700408329600.

Kanno, Y. (2000). Bilingualism and identity: The stories of Japanese returnees. *International Journal of Bilingual Education and Bilingualism* 3(1): 1–18. doi:10.1080/13670050008667697.

Krashen, S.D. (1985). *Second language acquisition and second language learning*. Reprinted. Oxford: Pergamon Press Language teaching methodology series.

Lai, M.L. (2011). Cultural identity and language attitudes – into the second decade of postcolonial Hong Kong. *Journal of Multilingual and Multicultural Development* 32(3): 249–264. doi:10.1080/01434632.2010.539692.

Larin, S.J. (2020). Is it really about values? Civic nationalism and migrant integration. *Journal of Ethnic and Migration Studies* 46(1): 127–141. doi:10.1080/1369183X.2019.1591943.

Li, D.C.S. (2017). *Multilingual Hong Kong: Languages, literacies and identities*. Cham: Springer International Publishing. doi:10.1007/978-3-319-44195-5.

Li, J. (2000). Analysis of the mentality structure of Ningbo migrants in Shanghai. *Academic Monthly* (12): 55–62.

Li, L. (2024). Reform makes it easier for migrants to gain city hukou. *China Daily*. https://www.chinadaily.com.cn/a/202408/03/WS66ad72eca3104e74fddb8471.html.

Li, Q. and Meng, L. (2014). Dialect, Mandarin and labor migration. *China Journal of Economics* 1(04): 68–84.

Li, Y.-T. (2024). Integration based on socio-economic status? Everyday interactions and boundary (un)making between mainland skilled professionals and locals in Hong Kong. *International Migration Review* 58(3): 1410–1433. doi:10.1177/01979183231171553.

Lindgren, E. and Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism* 10(1): 105–129. doi:10.1080/14790718. 2012.679275.

Liu, H., Fang, C., and Fan, Y. (2020). Mapping the inequalities of medical resource provision in China. *Regional Studies, Regional Science* 7(1): 568–570. doi:10.1080/21681376.2020.1848615.

Liu, L., Huang, Y., and Zhang, W. (2018). Residential segregation and perceptions of social integration in Shanghai, China. *Urban Studies* 55(7): 1484–1503. doi:10.1177/0042098016689012.

Liu, Y., Song, H., Wang, Tianying, Wang, Tianhao, Yang, H., Gong, J., Shen, Y., Dai, W., Zhou, J., Zhu, S., and Pan, Z. (2015). Determinants of tobacco smoking among rural-to-urban migrant workers: a cross-sectional survey in Shanghai. *BMC Public Health* 15(1): 131. doi:10.1186/s12889-015-1361-x.

Lu, H. (2002). Urban superiority, modernity and local identity – a think piece on the case of Shanghai. In: Faure, D. and Liu, T.T. (eds.). *Town and country in China*. London: Palgrave Macmillan UK: 126–144. doi:10.1007/978-1-137-07001-2_7.

Lu, S. and Chen, Y. (2019). The power of language: Does speaking local enhance the social integration of migrants? *Economic Science* (04 vo): 118–128.

Lu, Y. and Zhang, K. (2019). Geographic distance, dialect difference, and spatial labor mobility. *Statistical Research* 36(03): 88–99.

Nakhaie, R. (2020). Language proficiency and sociocultural integration of Canadian newcomers. *Applied Psycholinguistics* 41(6): 1437–1464. doi:10.1017/S01427 16420000375.

Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.

Norton, B. (2013). *Identity and language learning: Extending the conversation*. Harlow: Pearson Education. doi:10.21832/9781783090563.

Norton, B. and Toohey, K. (2011). Identity, language learning, and social change. *Language Teaching* 44(4): 412–446. doi:10.1017/S0261444811000309.

Office of the Leading Group of the State Council for the Seventh National Population Census (2021). *Communiqué of the Seventh National Population Census (No. 7)*.

Oktavianus, J. (2025). Words apart: Unravelling the language barrier plight of migrant domestic workers in Hong Kong. *Language and Intercultural Communication* 25(3): 349–364. doi:10.1080/14708477.2025.2520971.

Peirce, B.N. (1995). Social identity, investment, and language learning. *TESOL Quarterly* 29(1): 9. doi:10.2307/3587803.

Pendakur, K. and Pendakur, R. (2002). Language as both human capital and ethnicity. *International Migration Review* 36(1): 147–177. doi:10.1111/j.1747-7379.2002.tb00075.x.

Pettigrew, T.F. and Tropp, L.R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology* 90(5): 751. doi:10.1037/0022-3514.90.5.751.

Pong, D. (ed.) (2009). *Encyclopedia of modern China*. Detroit: Gale.

Potter, S.H. and Potter, J.M. (2009). *China's peasants*. Cambridge: Cambridge University Press.

Qin, W. (2025). The report on the work of Shenzhen City 2025. https://fao.sz.gov.cn/xxgk/zyxw/content/post_1533472.html.

Remennick, L. (2004). Language acquisition, ethnicity and social integration among former Soviet immigrants of the 1990s in Israel. *Ethnic and Racial Studies* 27(3): 431–454. doi:10.1080/01491987042000189213.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge. doi:10.4324/9780203851357.

Shanghai Bureau of Statistics (2014). *2013 Shanghai Citizen's Language Ability Survey Report*. Shanghai: Shanghai Bureau of Statistics. http://tjj.sh.gov.cn/tjfx/20140207/0014-266714.html.

Shanghai Statistical Bureau (2025). *Shanghai statistical yearbook 2024*. Shanghai: Shanghai Statistical Bureau.

Sidnell, J. (2003). Constructing and managing male exclusivity in talk-in-interaction. In: Holmes, J. and Meyerhoff, M. (eds) *The handbook of language and gender*. Hoboken, NJ: Wiley: 327–352. doi:10.1002/9780470756942.ch14.

Siu, H.F. (1995). Community festivals in Post-Mao South China: Economic transformation and cultural improvisation. *China Review* 16.1–16.17.

Solinger, D.J. (1999). *Contesting citizenship in urban China: Peasant migrants, the state, and the logic of the market*. Berkeley, CA: University of California Press. doi:10.1525/9780520922617.

Spolsky, B. and Shohami, I.G. (1999). *The languages of Israel: Policy, ideology and practice*. Clevedon: Multilingual Matters.

Stevens, G. (1992). The social and demographic context of language use in the United States. *American Sociological Review* 57(2): 171. doi:10.2307/2096203.

Sung, C.C.M. (2020). Cantonese learning, investments, and identities: Mainland Chinese university students' experiences during cross-border studies in Hong Kong. *Learning, Culture and Social Interaction* 26: 100415. doi:10.1016/j.lcsi.2020.100415.

Sutton, D.S. (2007). Ritual, cultural standardization, and orthopraxy in China: Reconsidering James L. Watson's ideas. *Modern China* 33(1): 3–21. doi:10.1177/0097700406294914.

Sze, M., Butow, P., Bell, M., Vaccaro, L., Dong, S., Eisenbruch, M., Jefford, M., Girgis, A., King, M., McGrane, J., Ng, W., Asghari, R., Parente, P., Liauw, W., Goldstein, D., and on behalf of the Psycho-oncology Co-operative Research Group Culturally and Linguistically Diverse (CALD) TEAM (2015). Migrant health in cancer: Outcome disparities and the determinant role of migrant-specific variables. *The Oncologist* 20(5): 523–531. doi:10.1634/theoncologist.2014-0274.

Tannenbaum, M. (2009). What's in a language? Language as a core value of minorities in Israel. *Journal of Ethnic and Migration Studies* 35(6): 977–995. doi:10.1080/13691830902957742.

Tian, K. (1995). Investigation, analysis and reflections on the urban adaptability of migrant workers. *Social Science Research* (05): 90–95.

Ünal, E. and Papafragou, A. (2016). Production–comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language* 89: 179–199. doi:10.1016/j.jml.2015.12.001.

Ushioda, E. (2009). 11. A person-in-context relational view of emergent motivation, self and identity. In: Dörnyei, Z. and Ushioda, E. (eds.). *Motivation, language identity and the L2 self*. Clevedon: Multilingual Matters: 215–228. doi:10.21832/9781847691293-012.

Ushioda, E. (2011). Language learning motivation, self and identity: Current theoretical perspectives. *Computer Assisted Language Learning* 24(3): 199–210. doi:10.1080/09588221.2010.538701.

Valenta, M. and Bunar, N. (2010). State assisted integration: Refugee integration policies in Scandinavian welfare states: The Swedish and Norwegian experience. *Journal of Refugee Studies* 23(4): 463–483. doi:10.1093/jrs/feq028.

Wang, D., Zhou, T., and Wang, M. (2021). Information and communication technology (ICT), digital divide and urbanization: Evidence from Chinese cities. *Technology in Society* 64: 101516. doi:10.1016/j.techsoc.2020.101516.

Wang, J., Wu, M., Jiang, H., Wang, G., Luo, L., and Lu, J. (2011). *Annual report on population and health development of Shenzhen*. Beijing. Ren kou yu jian kang lan pi shu. doi:10.1142/7648.

Wang, T. (2016). Law of the People's Republic of China on the standard spoken and written Chinese language. *Chinese Law and Government* 48(4): 275–278. doi:10.1080/00094609.2016.1118307.

Wang, W.W. and Fan, C.C. (2012). Migrant workers' integration in urban China: Experiences in employment, social adaptation, and self-identity. *Eurasian Geography and Economics* 53(6): 731–749. doi:10.2747/1539-7216.53.6.731.

Wang, Z., Lin, L., Tang, S., and Xiao, Y. (2023). Migration and migrants in Chinese cities: New trends, challenges and opportunities to theorise with urban China. *Transactions in Planning and Urban Research* 2(4): 317–331. doi:10.1177/27541223231217431.

Wei, L. and Gao, F. (2017). Social media, social integration and subjective well-being among new urban migrants in China. *Telematics and Informatics* 34(3): 786–796. doi:10.1016/j.tele.2016.05.017.

West, C. and Fenstermaker, S. (2002). *Doing gender, doing difference: social inequality, power and resistance*. New York: Routledge.

West, C. and Zimmerman, D.H. (1987). Doing gender. *Gender and Society* 1(2): 125–151. doi:10.1177/0891243287001002002.

Wu, X. and Yu, J. (2023). Does dialect difference impede patent transaction? Evidence from China's inter-city patent license data. *Technological Forecasting and Social Change* 187: 122197. doi:10.1016/j.techfore.2022.122197.

Xia, T. and Shen, Q. (2019). Individual agency in language acquisition planning for multi-dialectism: A Shanghai story. *International Journal of Bilingual Education and Bilingualism* 22(3): 338–351. doi:10.1080/13670050.2018.1530725.

Xiong, Y. (2009). The children of urbanization: The rural-urban awareness and self-identity of migrant workers' children in contemporary China. *China Rural Survey* (02): 2–11+45+95.

Xu, F. (2021). *Silencing Shanghai: Language and identity in urban China*. Lanham: Lexington Books. doi:10.5040/9781978727267.

Yang, J. (2015). Research on the assimilation of the floating population in China. *Social Sciences in China* 2: 61–79.

Yuan, R. and Wen, S. (2017). Determinants and effects of return migration in China. *Population Research* 2: 71–83.

Zhang, J., Huang, J., Wang, J., and Guo, L. (2020). Return migration and Hukou registration constraints in Chinese cities. *China Economic Review* 63: 101498. doi:10.1016/j.chieco.2020.101498.

Zhang, L. and Fan, C. (2017). Investigation on the survival situation of dialects under the general trend of popularization of Putonghua – taking Wu-speaking area as an example. *Research on Transmission Competence* 1(08): 227–228+232.

Zhang, L. and Wang, G. (2010). Urban citizenship of rural migrants in reform-era China. *Citizenship Studies* 14(2): 145–166. doi:10.1080/13621021003594809.

Zhao, Y. (2002). Causes and consequences of return migration: Recent evidence from China. *Journal of Comparative Economics* 30(2): 376–394. doi:10.1006/jcec.2002.1781.

Zhou, M., Wang, W., Pok Loa, S., and Kan, M.-Y. (2023). Moving in the time of COVID-19: How did the pandemic situations affect the migration decisions of Hong Kong people? *Asian Population Studies* 19(2): 204–227. doi:10.1080/17441730.2023.2193481.

Zhu, L. (2002). On the urban adaptability of the peasant-worker strata. *Jianghai Academic Journal* (06): 82–88+206.

Zhu, Y. (2018). Lifestyle mobility: Shifting conception of home in modern China. *International Journal of Tourism Anthropology* 6(4): 357–374. doi:10.1504/IJTA.2018.10017671.

Appendix A: Provincial dialect distance

In this study, we adopted the formula proposed by Liu et al. (2020), which was also used by Wu and Yu (2023) to account of dialect distance:

$$Diadis_{i,SH} = \sum_{m=1}^{M} spop_{im} \cdot dialect_{im,SH},$$

where $dialect_{im,SH}$ is the dialect distance between county *m* in province *i* to Shanghai. $spop_{im}$ represents the population share of county *m* in province *i*. According to Wurm et al. (1987), the whole area of Shanghai belongs to the subgroup of Taihupian (太湖片) under the group Wuyu (吴语), which belongs to the supergroup of Wuyu (吴语). If the county belongs to the same subgroup, $dialect_{im,SH}$=0. If it is from a different subgroup but the same group, $dialect_{im,SH}$=1. If from a different group but the same supergroup, then $dialect_{im,SH}$=2. If it is from another supergroup, then $dialect_{im,SH}$=3. The administrative division has been adjusted to today's version, and the population share is calculated using China's 2010 census.

The provincial dialect distance has a mean of 2.614373 and a standard deviation of 0.6711531. The comprehension ability has a mean of 2.409595 and standard deviation of 1.393112. The current IV has a mean of 6.299788 and a standard deviation of 4.075342.

PROVINCE	DIALECT DISTANCE
BEIJING	3
TIANJIN	3
HEBEI	3
SHANXI	3
INNER MONGOLIA	3
LIAONING	3
JILIN	3
HEILONGJIANG	3
JIANGSU	1.994
ZHEJIANG	0.543
ANHUI	2.914
FUJIAN	2.983
JIANGXI	2.890
SHANDONG	3
HENAN	3
HUBEI	3
HUNAN	3
GUANGDONG	3
GUANGXI	3
HAINAN	3
CHONGQING	3
SICHUAN	3
GUIZHOU	3
YUNNAN	3
TIBET	3
SHAANXI	3
GANSU	3
QINGHAI	3
NINGXIA	3
XINJIANG	3

Appendix B. Robustness check

Table B-1: **Shanghaihua proficiency and Shanghainese identity recognition (OLS and IV–2SLS with previous Shanghai education experience as control)**

Panel A

VARIABLES	(1) Shanghainese Identity-OLS	(2) Shanghainese ID-IV-2SLS
Shanghaihua proficiency	0.540	1.154
	(0.032)	(0.190)
	0.000	0.000
Age	0.029	0.018
	(0.002)	(0.004)
	0.000	0.000
Hukou status at birth (Reference: Nonagricultural)		
Agricultural hukou	−0.121	−0.005
	(0.078)	(0.090)
	0.121	0.960
Education in Shanghai (Reference: Yes)		
No	−0.668	−0.205
	(0.087)	(0.168)
	0.000	0.223
Male	−0.125	−0.060
	(0.062)	(0.069)
	0.045	0.379
Married	−0.087	−0.115
	(0.080)	(0.084)
	0.274	0.170
Total income (logged)	0.015	−0.006
	(0.014)	(0.016)
	0.282	0.731
Occupation type (Reference: Managers of national organizations)		
Professionals	−0.136	−0.041
	(0.165)	(0.176)
	0.408	0.814
Administrative, security, and emergency personnels	−0.256	−0.188
	(0.195)	(0.207)
	0.191	0.364
Business and service	−0.321	−0.257
	(0.166)	(0.175)
	0.053	0.143
Agriculture, forestry, husbandry, fishery, and water-related industry	−0.474	−0.187
	(0.229)	(0.257)
	0.038	0.467
Production, transportation, equipment operators, and related personnel	−0.106	0.035
	(0.173)	(0.188)
	0.542	0.853
Military officers	−0.468	−0.414
	(0.664)	(0.699)
	0.481	0.553

Table B-1: (Continued)

Panel A

VARIABLES	(1) Shanghainese Identity-OLS	(2) Shanghainese ID-IV-2SLS
Others	−0.094	−0.017
	(0.177)	(0.188)
	0.597	0.930
Educational level (Reference: Below high school)		0.0504
High school	0.142	0.050
	(0.089)	(0.097)
	0.108	0.604
Associate degree	0.416	0.292
	(0.108)	(0.120)
	0.000	0.015
Bachelor's degree and above	0.455	0.542
	(0.098)	(0.106)
	0.000	0.000
Constant	1.510	0.353
	(0.256)	(0.444)
	0.000	0.426
Instrumental variable		COMP_Diadis
Cragg-Donald Wald F statistic		104.24
R^2	0.235	
Adjusted R^2	0.231	
Observations	3,335	3,335

Source: Shanghai Urban Neighborhood Survey (2017)
Notes: Standard errors are in parentheses and clustered at the individual level. Standard errors in the second row, p-values in the third row.

Panel B first stage

VARIABLES	
COMP_Diadis	−0.044
	(0.004)
	0.000
Age	0.019
	(0.001)
	0.000
Hukou status at birth (Reference: Nonagricultural)	
Agricultural hukou	−0.169
	(0.043)
	0.000
Education in Shanghai (Reference: Yes)	
No	−0.711
	(0.046)
	0.000
Male	−0.120
	(0.034)
	0.000
Married	0.052
	(0.043)
	0.229
Total income (logged)	0.029
	(0.008)
	0.000

Table B-1:   (Continued)

Panel B first stage

VARIABLES	
Occupation type (Reference: Managers of national organizations)	
Professionals	−0.128
	(0.090)
	0.152
Administrative, security, and emergency personnels	−0.097
	(0.106)
	0.363
Business and service	−0.057
	(0.090)
	0.526
Agriculture, forestry, husbandry, fishery, and water-related industry	−0.335
	(0.125)
	0.007
Production, transportation, equipment operators, and related personnel	−0.169
	(0.094)
	0.073
Military officers	−0.012
	(0.361)
	0.973
Others	−0.066
	(0.097)
	0.492
Educational level (Reference: Below high school)	
High school	0.083
	(0.049)
	0.086
Associate degree	0.106
	(0.060)
	0.074
Bachelor's degree and above	−0.243
	(0.054)
	0.000
Constant	2.106
	(0.137)
	0.000
R^2	0.206
Adjusted R^2	0.202
Observations	3,335

Source: Shanghai Urban Neighborhood Survey (2017)
Notes: Standard errors are in parentheses and clustered at the individual level. Standard errors in the second row, p-values in the third row.

Table B-2: Shanghaihua proficiency and Shanghainese identity recognition (OLS and IV–2SLS with whether using Shanghaihua at work or for social purposes as proxy for Shanghaihua proficiency)

Panel A

VARIABLES	(1) Shanghainese identity–OLS	(2) Shanghainese ID–IV-2SLS
Shanghaihua proficiency	1.377	5.803
	(0.107)	(1.121)
	0.000	0.000
Age	0.029	−0.007
	(0.002)	(0.010)
	0.000	0.439
Hukou status at birth (Reference: Non–Agricultural)		
Agricultural hukou	−0.136	−0.020
	(0.080)	(0.102)
	0.089	0.842
Been to SH before 14 (Reference: Yes)		
No	−0.957	0.272
	(0.105)	(0.335)
	0.000	0.417
Male	−0.110	0.149
	(0.064)	(0.102)
	0.084	0.143
Married	0.030	0.160
	(0.081)	(0.105)
	0.711	0.127
Total income (logged)	0.028	−0.009
	(0.014)	(0.020)
	0.054	0.659
Occupation type (Reference: Managers of national organizations)		
Professionals	−0.144	−0.082
	(0.168)	(0.206)
	0.391	0.692
Administrative, security, and emergency personnels	−0.262	−0.214
	(0.199)	(0.244)
	0.188	0.380
Business and service	−0.289	−0.221
	(0.168)	(0.208)
	0.086	0.286
Agriculture, forestry, husbandry, fishery, and water-related industry	−0.663	−0.436
	(0.232)	(0.291)
	0.004	0.134
Production, transportation, equipment operators, and related personnel	−0.132	0.012
	(0.176)	(0.220)
	0.452	0.955
Military officers	−0.231	0.300
	(0.674)	(0.839)
	0.732	0.720
Others	−0.079	0.095
	(0.180)	(0.226)
	0.661	0.673

Table B-2:   (Continued)

Panel A

VARIABLES	(1) Shanghainese identity–OLS	(2) Shanghainese ID–IV-2SLS
Educational level (Reference: Below high school)		
High school	0.203	0.091
	(0.090)	(0.114)
	0.024	0.425
Associate degree	0.631	0.412
	(0.109)	(0.144)
	0.000	0.004
Bachelor's degree and above	0.708	0.811
	(0.098)	(0.123)
	0.000	0.000
Constant	2.447	2.161
	(0.253)	(0.318)
	0.000	0.000
Instrumental variable		COMP_Diadis
Cragg–Donald Wald F statistic		46.16
R^2	0.211	
Adjusted R^2	0.207	
Observations	3,334	

Source: Shanghai Urban Neighborhood Survey (2017)
Notes: Standard errors are in parentheses and clustered at the individual level. Standard errors in the second row, p-values in the third row.

Panel B First Stage

VARIABLES	
COMP_Diadis	−0.009
	(0.001)
	0.000
Age	0.008
	(0.000)
	0.000
Hukou status at birth (Reference: Nonagricultural)	
Agricultural hukou	−0.022
	(0.013)
	0.086
Been to SH before 14 (Reference: Yes)	
No	−0.269
	(0.016)
	0.000
Male	−0.061
	(0.010)
	0.000
Married	−0.028
	(0.013)
	0.030
Total income (logged)	0.007
	(0.002)
	0.002

Table B-2: (Continued)

Panel B First Stage

VARIABLES	
Occupation type (Reference: Managers of national organizations)	
Professionals	−0.009
	(0.027)
	0.739
Administrative, security, and emergency personnels	−0.008
	(0.032)
	0.800
Business and service	−0.007
	(0.027)
	0.811
Agriculture, forestry, husbandry, fishery, and water-related industry	−0.024
	(0.038)
	0.516
Production, transportation, equipment operators, and related personnel	−0.021
	(0.029)
	0.462
Military officers	−0.106
	(0.109)
	0.330
Others	−0.028
	(0.029)
	0.344
Educational level (Reference: Below high school)	
High school	0.012
	(0.015)
	0.415
Associate degree	0.029
	(0.018)
	0.102
Bachelor's degree and above	−0.046
	(0.016)
	0.005
Constant	0.110
	(0.041)
	0.008
R^2	0.245
Adjusted R^2	0.241
Observations	3,334

Source: Shanghai Urban Neighborhood Survey (2017)
Notes: Standard errors are in parentheses and clustered at the individual level. Standard errors in the second row, p-values in the third row.

Table B-3: **Shanghaihua proficiency and Shanghainese identity recognition (OLS for controlling whether one is from Zhejiang or Jiangsu [Wuyu provinces], separating the two subsamples, and doing the interaction with whether one is from a Wuyu province)**

VARIABLES	(1) OLS	(2) Non-Wuyu	(3) Wuyu	(4) WuyuINT
Shanghaihua proficiency	0.509	0.477	0.519	0.484
	(0.033)	(0.040)	(0.060)	(0.040)
	0.000	0.000	0.000	0.000
From Wuyu provinces (Reference: No)				
Yes	0.187			0.028
	(0.073)			(0.160)
	0.010			0.861
Shanghaihua proficiency * Wuyu (Reference: No)				0.072
Yes				(0.064)
				0.262
R^2	0.237	0.163	0.306	0.238
Adjusted R^2	0.233	0.157	0.292	0.233
Observations	3,335	2,461	874	3,335

Source: Shanghai Urban Neighborhood Survey (2017).
Notes: Standard errors are in parentheses and clustered at the individual level. Standard errors in the second row, p-values in the third row. Control variables are the same as in Table B2.

Table B-4: **Shanghaihua proficiency and Shanghainese identity recognition (IV–2SLS for adding whether having local hukou and homeownership control)**

VARIABLES	(1) Baseline	(2) SH hukou	(3) SH hukou + Homeownership
Shanghaihua proficiency	1.165	0.892	0.865
	(0.190)	(0.212)	(0.213)
	0.000	0.000	0.000
Local hukou status (Reference: Non-Shanghai)			
Shanghai		1.278	1.243
		(0.188)	(0.185)
		0.000	0.000
Own the house			0.243
			(0.087)
			0.005
Age	0.018	0.007	0.006
	(0.004)	(0.003)	(0.003)
	0.000	0.020	0.053
Hukou status at birth (Reference: Nonagricultural)			
Agricultural hukou	0.002	0.073	0.098
	(0.087)	(0.080)	(0.079)
	0.985	0.360	0.218
Been to SH before 14 (Reference: Yes)			
No	–0.188	–0.062	–0.059
	(0.215)	(0.189)	(0.188)
	0.381	0.744	0.754

Table B-4:   (Continued)

VARIABLES	(1) Baseline	(2) SH hukou	(3) SH hukou + Homeownership
Male	−0.059	−0.009	−0.001
	(0.069)	(0.063)	(0.063)
	0.394	0.889	0.990
Married	−0.113	−0.094	−0.123
	(0.086)	(0.081)	(0.081)
	0.189	0.244	0.127
Total income (logged)	−0.006	−0.003	−0.001
	(0.016)	(0.016)	(0.016)
	0.738	0.843	0.934
Occupation type (Reference: Managers of national organizations)			
Professionals	−0.034	−0.048	−0.030
	(0.175)	(0.164)	(0.164)
	0.848	0.769	0.855
Administrative, security, and emergency personnels	−0.182	−0.145	−0.121
	(0.207)	(0.193)	(0.193)
	0.379	0.454	0.529
Business and service	−0.250	−0.167	−0.138
	(0.175)	(0.164)	(0.164)
	0.154	0.308	0.399
Agriculture, forestry, husbandry, fishery, and water-related industry	−0.188	−0.153	−0.132
	(0.257)	(0.239)	(0.238)
	0.466	0.523	0.578
Production, transportation, equipment operators, and related personnel	0.041	0.064	0.092
	(0.186)	(0.174)	(0.173)
	0.824	0.715	0.595
Military officers	−0.400	−0.852	−0.849
	(0.700)	(0.659)	(0.656)
	0.568	0.196	0.195
Others	−0.012	0.060	0.089
	(0.188)	(0.176)	(0.175)
	0.951	0.734	0.611
Educational level (Reference: Below high school)			
High school	0.053	0.037	0.013
	(0.098)	(0.091)	(0.091)
	0.590	0.682	0.885
Associate degree	0.324	0.325	0.287
	(0.128)	(0.120)	(0.119)
	0.011	0.007	0.016
Bachelor's degree and above	0.595	0.331	0.244
	(0.103)	(0.100)	(0.106)
	0.000	0.001	0.021
Constant	0.300	0.809	0.844
	(0.449)	(0.473)	(0.473)
	0.504	0.087	0.075
Instrumental variable	COMP_Diadis	COMP_Diadis	COMP_Diadis
Cragg-Donald Wald F statistic	106.38	79.93	78.12
Observations	3,335	3,335	3,335

Source: Shanghai Urban Neighborhood Survey (2017).
Notes: Standard errors are in parentheses and clustered at the individual level. Standard errors in the second row, p-values in the third row.