



Access to Language Training and Local Integration of Refugees

Mette Foged and Cynthia van der Werf

April 2022

Access to Language Training and Local Integration of Refugees*

Mette Foged[†] and Cynthia van der Werf[‡]

April 11, 2022

Abstract

This paper examines whether proximity to language classes raises refugees' language proficiency and improves their social and economic integration. Our identification strategy leverages the opening, closing, and gradual expansion of local language training centers in Denmark, as well as the quasi-random assignment of the refugees to locations with varying proximity to a language training center. First, we show that refugees' distance from the assigned language training center is as good as random. Second, we show that language skills decrease monotonically with commuting time such that a one-hour decrease in commuting time increases fluency in the Danish language by 4 to 6 percent relative to the sample mean. The exogenous variation in language proficiency generates substantial positive effects on post language training human capital acquisition and on the integration of the refugees in the local communities where they were initially placed, as measured by the lower exit rates from those same communities and lower mobility to the largest, most immigrant-dense, cities in Denmark.

JEL Classification: J60, J24

Keywords: Refugee Integration, Language Skills,

*This project received funding from the Economic Policy Research Network (EPRN grant no. 112784) and Mette also acknowledge generous financial support from the Innovation Fund Denmark (grant no. 6149-00024B). Giovanni Peri, Jacob Arendt, Iben Bolvig, Linea Hasager, Daphné Skandalis, Nikolaj Harmon, Courtney Brell, and participants at the 10th Annual Conference on “Immigration in OECD Countries” and the Annual Congress of EEA provided valuable discussions and feedback. We thank the Ministry of Immigration and Integration for assistance with data on language training and Janis Kreuder for assistance with the data preparation. We are indebted to Jeppe Rich and Christian Overgaard Hansen from the Technical University of Denmark for sharing their data on distances between the geographical zones from the Danish National Passenger Model with us.

[†]University of Copenhagen, Mette.Foged@econ.ku.dk.

[‡]Inter-American Development Bank, cvanderwerf@iadb.org.

1 Introduction

Most immigrants start their lives in the destination country at a substantial economic disadvantage, as measured by their employment rate or earnings relative to native born populations. The gap with natives gradually shrinks, but some immigrant groups never fully catch up.¹ A common interpretation for why immigrants' labor market outcomes improve over time since arrival is the acquisition of host country human capital, most importantly language capital. In fact, one of the most consistent findings in the economics of migration is the strong positive correlation between language proficiency and the economic success of immigrants.²

Still, we know relatively little about the causal effect of language skills. Recent papers have exploited exogenous variation in language training for adult immigrants to estimate the causal impact of language training on the labor market outcomes of the participants.³ While language training (hopefully) improves language skills, other mechanisms could be at play too. Attending language classes could improve immigrants' knowledge about the host country and allow them to form valuable networks, which could in turn reduce informational barriers and ease job searching and matching with potential employers.⁴

This paper contributes to this strand of literature in two important ways. First, we document a causal link between language training and language proficiency. Second, instead of estimating the impact of language training (allowing many potential mechanisms to play a role), we causally estimate the impact of language proficiency on the employment probability and the post-migration human capital acquisition of the immigrants. Furthermore, immigrants who become more proficient in the dominant language may also become better integrated into

¹A large literature, starting with Chiswick (1978) and Borjas (1985), documents these patterns. See Brell, Dustmann, and Preston (2020) for an overview and Schultz-Nielsen (2017) for a detailed analysis of the experience in Denmark.

²E.g., Chiswick (1991); Chiswick and Miller (1995); Dustmann (1994); Dustmann and van Soet (2001, 2002); Dustmann and Fabbri (2003); Bleakley and Chin (2004, 2010); Clausen et al. (2009); Akresh and Frank (2011). See also Figure 1 in this paper.

³See Sarvimäki and Hämäläinen (2016); Lochmann, Rapoport, and Speciale (2019); Arendt et al. (2021); and Heller and Mumma (2022), which we describe below.

⁴Battisti, Giesing, and Laurentsyeve (2019) show that reducing search and matching frictions for refugees and asylum seekers increases their employment probability. Belot, Kircher, and Muller (2018) and Altmann et al. (2022) demonstrate an important role of information frictions in the job search of native born workers.

their local community and be less likely to move to a minority-language enclave in a large city. We test this hypothesis using two measures of mobility: (1) mobility out of the immigrants' originally assigned neighborhood and (2) mobility to the four largest cities of Denmark. We think these are policy relevant outcomes because many countries have policies in place that distribute refugees evenly across the country, and these policies are often counteracted by substantial subsequent mobility.⁵

Our population consists of individuals who were granted refugee status in Denmark between 2003 and 2013 (1999 to 2013 when we analyze the impact over time). Refugees are quasi-randomly assigned to municipalities in Denmark in proportion to municipal inhabitants. The assigned municipality is responsible for finding affordable housing for the refugee, for the provision of language training, and for other support as part of the Danish Integration Program (see Arendt et al., 2021). The ideal experiment that we aim to mimic with our empirical strategy is one where individuals are randomly assigned to identical locations with different distances to language training. Our most demanding specification compares refugees placed in the same neighborhood (zone) at different points in time and with different distances to language training due to the gradual expansion of language training and the opening and closing of language training centers. Another specification, using much more variation in the data, compares refugees within the same municipality. These comparisons include individuals assigned to different neighborhoods in the same municipality (likely due to the availability of suitable housing at their date of arrival), who therefore face different length commutes to get to language training. To alleviate concerns that these neighborhoods differ in dimensions that matter for our outcomes, we include measures of job opportunities and residential networks in a third specification. Our instrument is travel time by public transport between very small geographical units in Denmark (zones used to study commuting behavior in Denmark, see Rich and Hansen, 2016).⁶

⁵Notice, that while an even distribution of refugees is a popular policy in many countries, it may not be an economically sound policy. Bansak et al. (2018) estimate large positive effects of matching refugees to locations that maximize their employment opportunities conditional on all their observed characteristics rather than pursuing an even distribution.

⁶Such an instrument is related to a large literature estimating the marginal returns to education using spatial

We show that our instrument is as good as random for the individual refugees and that it is monotonically decreasing in their language proficiency. There is negative selection into attending more hours of language training (likely due to the substitutability between ability and hours of class in the production of language skills). Once we eliminate this selection using our instrument in a 2SLS estimation, we find a larger, positive effect of hours of class on language proficiency. We use a novel, objective measure of language skills from standardized tests that map into the Common European Framework of Reference for Languages (CEFR).

We observe a strong correlation between language proficiency at the end of the three-year Integration Program and key indicators of economic integration two years later. Using our instrument, we find that compliers are significantly more likely to study in Denmark but not more likely to be employed (2SLS estimates are similar in magnitude to OLS but not significant). It is possible that the large effect we find on education postpones any impact of language skills on employment. Furthermore, refugees who become more fluent in Danish due to the quasi-random assignment to a location with a shorter commute to language training (the compliers) are less likely to leave their assigned neighborhood and less likely to move to one of the four main cities of Denmark with large pre-existing immigrant populations. This may hamper the economic returns to language skills (Arendt et al., 2021).

Using a longer panel, we analyze the impacts of distance to language training up to ten years after arrival in the municipality. The reduction in mobility to the largest cities of Denmark is persistent; the effect on enrollment decreases over time, but we do not detect significant changes in employment within the time horizon. We also test the robustness of our results to alternative functional forms of our instrument (log distance and an indicator for distance above median) and use natives in a placebo test of the distance to language training instrument. These placebo tests confirm that job opportunities and proximity of further education are unlikely to drive our results.

variation in distance to college as an instrument (Card, 1995, 2001). We improve this identification by considering a group whose location is more likely exogenous.

Our findings are consistent with Arendt et al. (2021), who show that a substantial expansion of the Danish language training in 1999, introduced with a sharp discontinuity in the date the individual is granted refugee status, increased refugees’ long-run earnings and employment. Refugees subject to the reform were also more likely to acquire further education in Denmark and to work in more communication-intensive occupations. The effects of language training emerge gradually after completion of the program and are absent or smaller in rural municipalities (the location of most of the refugees in this paper).

Sarvimäki and Hämäläinen (2016) exploit a similar reform in Finland that restructured active labor market programs for immigrants to include more language training. The Finnish reform increased compliers’ cumulative earnings over a 10-year period, but the authors did not find a significant effect on the time to first employment or the likelihood of being employed. Lochmann, Rapoport, and Speciale (2019) use a discontinuity in an initial language test to study the effect of attending more hours of language classes and find a positive effect on labor force participation but no effect on employment, which in their case could be due to a shorter time horizon. Interestingly, Lochmann, Rapoport, and Speciale measure language skills after program participation, but do not detect a clear effect. Finally, Heller and Mumma (2022) use a lottery for places in the English for Speakers of Other Languages (ESOL) program in Massachusetts and find substantial effects on immigrants’ earnings and voter registration. We document a significant positive effect of language classes on language proficiency and show that language proficiency improves key markers of immigrants’ economic and social integration.

The next section describes the institutional setting and how the location of refugees and training facilities create exogenous variation in language training proximity. Section 3 describes the data and the sample we use. Specifically, the high-quality data on language class attendance and language proficiency and our measure of distance to language classes. The empirical strategy and the identifying assumptions are discussed in section 4, before we present our results in sections 5 to 8 and conclude the paper in section 9.

2 Background

Individuals who are granted refugee status in Denmark are assigned to a municipality in Denmark following a quota system that aims to distribute refugees proportional to municipal inhabitants (see Azlor, Damm, and Schultz-Nielsen, 2020; Arendt et al., 2021, for details).⁷ The municipality of assignment has the month the individual is granted refugee status plus one month to find affordable housing for the refugee in the municipality. Public housing is in short supply, and placement officers in the municipality work continuously to find available housing and allocate new refugees to housing on a first-come-first-serve basis, where the next in line is offered the next available housing.

The Danish Immigration Service is responsible for the assignment of refugees to municipalities and it is not possible for the municipality to intervene in the process and select the refugees they would like to receive. Refugees may express their location preferences to a caseworker from the Danish Immigration Service and there is a chance that such location preferences are granted if the desired municipality has not yet met its annual quota. Azlor, Damm, and Schultz-Nielsen (2020) analyze the potential sorting of refugees across municipalities in the period we study in detail and find that highly educated refugees are more likely to be assigned to a municipality with better employment opportunities as measured by the local employment rate, early in the year.⁸ Such sorting is not problematic for our identification strategy as long as it is similar over time, because we do not rely on the spatial variation across municipalities in any of our specifications.

The refugee starts a three-year Integration Program upon settlement in the municipality. The program is mandatory for refugees. They are eligible for welfare benefits from arrival conditional on participation in this program, and access to permanent residency and citizenship

⁷We do not have asylum seekers in our data. Asylum seekers live in accommodation centers and have very limited access to the Danish society, while their case is being processed. They are not allowed to work while waiting on the decision (this changed May 1, 2013 but administrative barriers for employment remain high prior to asylum). This effectively means that refugees are unemployed upon placement in the municipality.

⁸This correlation disappears when the 10 most popular municipalities have met their annual quota and they show that after this point (roughly in June) the remaining placements are orthogonal to observable characteristics.

require completion of the program.⁹ The assigned municipality is responsible for the Integration Program, and the refugee can therefore only move during the Integration Program if a new municipality agrees to take over the integration responsibility. This rarely happens in practice and few refugees leave their assigned municipality within the first three years (Nielsen and Jensen, 2006; Azlor, Damm, and Schultz-Nielsen, 2020).

Extensive language training is a key element of the Danish Introduction Program. The refugees are offered up to 1800 hours of language training, a 20-hours course in civic understanding, and employment support. A refugee who obtains regular employment during the program can continue in language training free of charge but it is no longer required to attend.

Language training for adult immigrants in Denmark is organized into three tracks called Danish I, II and III and six modules within each track (see Appendix Figure A.1).¹⁰ Danish I to III progress at a different pace. Danish I is designed for individuals who have basic or no schooling, are illiterate or unfamiliar with the Latin alphabet, or individuals who for other reasons such as poor health need a relatively slow pace in the language instructions. Danish III is for immigrants who have college or university degrees or are in other ways regarded as fast-learners in the initial screening. For instance, young refugees who have not yet started a tertiary education can still be assigned to Danish III. Danish II is for those in between, typically immigrants with an short education from abroad. The next section describes the language training and our data on attendance and proficiency.

⁹See Arendt, Dustmann, and Ku (2021) for further requirements for permanent residency.

¹⁰This exact structure was only introduced January 1, 2004. Our sample contains refugees arriving in 2003. They start under the old rules but will follow the new structure like everybody else in our sample from January 1, 2004. We only measure their language skills starting in year 1 (2004) and our preferred specification considers the language skills after completion of the three-year Integration Program (at the end of year 3).

3 Data

3.1 Language Class Attendance and Language Proficiency

An important contribution from our paper is that we have very detailed and objective data on language skills and language investments from Danish administrative registers. We observe the assigned language training track (Danish I, II or III), when a refugee starts and finishes a module, when a refugee passes the test at the end of the module, and the hours of class attended (assigned hours - hours of absence). Our measure of language proficiency is based on the test results from the language training centers instead of the self-reported measures used in the previous literature (see, for example, Chiswick and Wang, 2019; Aoki and Santiago, 2018; Akresh, Massey, and Frank, 2014). This reduces misclassification error due to subjective perception of the knowledge scale.

Passing the test at the end of a module marks completion of the module and achievements according to the first five levels of Common European Framework of Reference for Languages (CEFR). The levels are A1, A2, B1, B2, and C1 and ranges from very basic to fluent.¹¹ We translate the ordinal proficiency scale into numbers from one to five and assign a zero to individuals who have not yet passed any of the tests. This measure has a couple of advantages over using individuals' GPA in the final exams (taken at the end of module 6 in each track). First, the CEFR level is comparable across the three tracks of language training, while the GPA is from exams that vary in difficulty across tracks. Second, progression in terms of CEFR levels is observable for everyone because language training is mandatory for refugees and we observe when individuals pass the tests while the final exam is not mandatory and only two-thirds of our population take it within five years from arrival. We, therefore, assign refugees a CEFR level at the end of the Integration Program (at the end of year 3 after placement in the municipality) based on the intermediate tests. This is our preferred measure of language skills and do not use

¹¹Appendix Figure A.1 and Appendix Table A.1 show how the modules of the Danish language training tracks map into CEFR levels and explain each proficiency level.

the GPA from the final exams.¹² Similarly, hours is the cumulative hours during the Integration Program (in year 0 to 3).¹³

3.2 Distance to Language Training

For each refugee and each language training center we observe an address that we map into zones from the Danish National Passenger Model (Rich and Hansen, 2016).¹⁴ Rich and Hansen kindly shared with us different measures of distance between these zone-pairs. Our preferred distance measure is travel time in minutes by public transport.¹⁵ This distance measure covers urban and rural locations and is only missing for a zone-pair if no one is observed traveling between the two zones.

Refugees are usually assigned to the nearest language training facility within the municipality (some municipalities own language training centers and some contract with private suppliers). Not all training facilities have the capacity or volume of participants to offer all tracks. Some refugees will, therefore, have to go to a different school to attend the track that matches their skills and aspirations upon arrival. Travel time varies from under 5 minutes to more than 5 hours in our sample. On average, refugees travel 45 minutes from their home to the language training center and half of them travel less than 40 minutes to attend the language class. Many rural municipalities do not have a language center in the municipality, and refugees in these areas travel more than two hours.¹⁶ Overall, as more state-approved training centers open

¹²We only use the information on whether they pass the final exam at the end of a language training track to update the CEFR level with the level that corresponds to passing the final module of that track.

¹³Appendix Table A.2 shows results from similar regressions as Table 3 in the main text where hours are instead the annual hours of class in year 0 to 3 after placement. We also look at the effect of cumulative hours on proficiency each year after program start in Appendix Table A.3. However, the right time to measure hours and proficiency with our data is in year 3 as we will argue in section 5.

¹⁴We have access to the pseudo-anonymized addresses within municipalities of individuals and language training centers that we can map to zones and we have distances between zone pairs. The total number and the spatial distribution we observe (Appendix Figure A.3) match well available information about language training centers, see e.g. <https://dedanskesprogcentre.dk/english/>.

¹⁵We also have distance by car, bike and walking as well as the ticket price by public transport. However, it is unlikely that our population has the resources to buy and to own a car during their first years in Denmark, and only very few who happen to live very close to the language training center will be able to walk or bike. Furthermore, it is possible that refugees get reimbursed for their travel expenses to language classes. For these reasons, the most reasonable measure of distance for our purpose seems to be the travel time by public transport.

¹⁶Some of these municipalities are islands and the refugee will need several modes of transport including a ferry. A nice source of exogenous variation arises when refugees on an island suddenly have a much shorter

during our period (Table 1), average distance decreased from 48 minutes to 42.5 minutes.¹⁷ We convert travel time in minutes into hours for readability in all our regressions.

Table 1: State-Approved Language Training Centers

	2003	2013	Changes
Public	47	37	
Private	20	35	
Exits			18
Entries			23
Total	67	72	

Notes: Local providers of Danish I, II and/or III and the official exams approved by the Danish state. The last column shows the total number of exits and entries from 2003 to 2013. A training center that changes location counts as an exit and an entry.

3.3 Sample and Outcomes

Our analysis considers refugees who settled in Denmark from 2003 to 2013. This allows us to follow all refugees for at least five years. We choose 2003 as the earliest immigration cohort, because the detailed data from the Danish language training centers (described in section 3.1) are available starting on January 1, 2004.¹⁸

Individuals with severe health problems are exempt from the Integration Program, as they will leave the labor market and be eligible for disability benefits. This process usually takes several years, but we observe whether the municipality assigned them to language training and exclude individuals who are never assigned to language training (individuals who are assigned may have zero hours and never pass a test). Similarly, some individuals may decide to leave the labor market (not claim welfare benefits), and hence, they would not have to participate either. Everyone else is unemployed when first placed in the municipality, because they were not allowed to work while waiting for the adjudication, and should therefore be assigned to language training. Finally, we exclude individuals aged 50 or older, because we have very low commute because a state-approved language training center opens on the islands. This happened at Samsøe in 2017. Unfortunately, we do not have enough refugees on islands and in similarly remote municipalities to focus on such events.

¹⁷Language centers may open or close at specific locations, for instance, because private, state-approved suppliers win or lose the contract with the municipality.

¹⁸We use a larger sample of refugees who entered between 1999 and 2013 in Section 8 to analyze the impact over time using our reduced form specification, which does not require language data.

compliance for this group, and we exclude individuals who are initially placed in one of the four largest urban municipalities in Denmark (Copenhagen including Frederiksberg, Århus, Odense and Aalborg). These municipalities have the largest pre-existing immigrant populations and we will analyze post-placement mobility into these cities. More importantly for our decision to exclude these cities is measurement error in our instrument. Distances to training is relatively short for immigrants in large urban areas and some refugees might be able to walk or bike in which case our distance measure (travel time by public transport) will be less accurate. These restrictions leave us with 10,170 refugees in our main sample.¹⁹

It is worth noting that the refugees we exclude are those most likely to be devastated by war and those whom for cultural or other reasons do not want to participate in the labor market. Hence, labor force participation in our sample is relatively high and 65 percent of the refugees are either employed or enrolled in an education program (or both) in year five after arrival. Employed is equal to one if the refugee had positive labor market earnings (extensive margin), and enrolled is equal to one if the individual is active in a study program (eligible for study grants). Figure 1 plots the share of immigrants employed (panel a) and enrolled (panel b) against our measure of Danish language skills. Consistent with the existing literature we find a strong positive correlation between host-country language capital and the share employed ($p < 0.001$) or enrolled ($p = 0.001$). These correlations reflect that more able and motivated individuals are more likely to work and study and more likely to acquire the language as well as any causal link from fluency in the language to employment and post-migration human capital investments. Our goal is to disentangle the causal effect of language skills on these key indicators of economic success.

As an additional contribution to the literature on the causal effects of language training and language skills, we create two measures of integration into the local community of initial

¹⁹We also dropped cases where distance between the refugee’s home and the language training center could not be calculated, individuals for whom we do not have information on all controls, and individuals who left the country or died within five years after arrival. Attrition is 17.6 percent.

placement. The first, is a dummy equal to one if the individual leaves the assigned neighborhood (zone) between completion of the Integration Program and year five. The second, is an indicator equal to one if the individual lives in one of the four largest cities in Denmark in year five (due to our sampling criteria no refugees start out in these cities). These variables are not significantly correlated with language proficiency in the raw data.

4 Empirical Strategy

This section describes in a formalized way how adult immigrants who arrive to the host country with no knowledge of the majority language acquire the language (section 4.1), how we estimate the impact of hours of class on language capital and the impact of language capital on measures of immigrant integration (section 4.2), and discuss the validity of our empirical strategy (section 4.3).

4.1 The Production of Language Skills

Adult immigrants acquire knowledge about the Danish language by attending language classes. Other inputs to the production of language skills may also be important, especially after some initial introduction to the Danish language. Suppose the following dynamic production function describes the production of language capital Q between year s and $s + 1$ after arrival in the host country:

$$Q_{s+1} = f_s(A, Q_s, I_s). \quad (1)$$

I_s is investment in language learning in period s after arrival, and A is individual characteristics that are predetermined upon placement in the municipality and influence language learning such as ears for language and education acquired abroad.²⁰ The function f_t is strictly increasing and

²⁰We ignore subscripts for individuals. We introduce a subscript t to index the year of arrival in our empirical specification that uses a panel of individuals arriving from 2003 to 2013.

strictly concave in I_t .^{21,22}

It is reasonable to assume that the return to investing in language learning is increasing in the existing language capital ($\partial^2 f_t / \partial Q_s \partial I_s > 0$, dynamic complementarity) and talent ($\partial^2 f_s / \partial A \partial I_s > 0$, more able individuals acquire the language faster). Refugees arrive with no prior knowledge of the Danish language ($Q_0 = 0$). It is therefore likely, that some initial investments (hours of class) are required before language skills can become self-productive ($\partial f_s / \partial Q_s > 0$, be acquired through everyday life without further language instruction).

Large investments in language learning early on (like in the Danish Integration Program prior to 2016)²³ are productive if dynamic complementarity and self-productivity are strong, i.e. when the stock of skills makes investments more productive and when skills produce more skills. These mechanisms will tend to put refugees with large initial language investments on a better trajectory, everything else held constant. However, attending language class takes time away from other activities that could facilitate language learning and enhance socio-economic integration. Hence, it is relevant to ask whether language training for immigrants is an effective way to increase their proficiency in the host country language (high returns to additional investments) or whether learning the language in the labor market and through everyday life is equally effective (strong self-productivity, possibly after some basic introduction to the language).

4.2 Estimating Equations

We first analyze whether language classes increase the language proficiency of adult immigrants or the counterfactual is equally effective. To answer this we estimate the following equation relating language skills measured by the obtained CEFR level ($CEFR_s$) to the cumulative hours

²¹Cunha and Heckman (2007) use a similar technology to describe skill formation in children where s is time since birth, A is parental characteristics and I is parental investments.

²²Figure A.2 shows that language capital as measured by the acquired CEFR level is increasing and concave in hours of class.

²³The policy changed in 2016 from a policy of language skills first and job search later to a strong emphasis and job search immediately upon placement in the municipality (see Arendt, 2022).

of class ($Hours_s$) at the end of the Integration Program ($s = 3$):²⁴

$$CEFR_{3,t} = \alpha + \beta Hours_{3,t} + X'_{0,t}\gamma + \omega_t + \mu_z + \xi_{3,t}. \quad (2)$$

ω_t and μ_z are year of placement and municipality or zone of placement fixed effects. $X'_{0,t}$ contains the individual characteristics measured at arrival. These are age, gender, civil status, number of children by child age groups, and initial assignment to language training tracks which is a rough proxy for education from abroad, abilities and ambition. The language training track fixed effects also account for the fact that the tracks progress at different paces. We also include country of origin dummies.

The municipality fixed effects capture time-constant differences across municipalities such as structural differences in employment opportunities and differences in how the Integration Program is administered. Similarly, zone effects capture time constant differences across zones which we sometimes refer to as neighborhoods. We also run a third type of specifications, which include municipality fixed effects and zone characteristics (calculated based on the entire population in each zone) to alleviate concerns that distance to language training within municipalities are somehow correlated with neighborhood (zone) characteristics that matter for our outcome. The included variables do not control for unobserved individual characteristics that influence hours of class and acquired language skills, we therefore employ an instrumental variable strategy. Equation 3 describes the first stage equation.

$$Hours_{3,t} = \alpha^1 + \beta^1 Distance_{0,t} + X'_{0,t}\gamma^1 + \omega_t^1 + \mu_z^1 + \varepsilon_{3,t}^1, \quad (3)$$

²⁴The recursive form of the production function in equation 1, $Q_{s+1} = h_s(A, Q_0, I_0, I_1, \dots, I_s)$, states that language skills Q is a function of all the investments in each period up to year $s + 1$. With only one instrument, we are not able to identify separately the effect of each period investment on the accumulated skills. We, therefore, replace the annual investments by total cumulative hours, $Hours_3 = I_0 + I_1 + I_2 + I_3$, when looking at the language skills at the end of period 3. Appendix Table A.2 show regressions of language proficiency at the end of year 3 on hours in each year from 0 to 3 estimated with OLS and 2SLS with a separate equation for each year. Attending classes increase proficiency each year during the program. This suggests that language training is productive beyond a basic introduction in the first year. The IV estimates increase with time since arrival, while the OLS estimates do not, suggesting a negative selection into still participating in language training in year three after arrival (language training ends when all modules have been completed).

where $Distance_{0,t}$ is travel time between the initial placement and the assigned language training center. Used in a 2SLS estimation, this instrument identifies the causal effect of hours on proficiency if individuals are quasi-randomly assigned to public housing within municipalities (due to capacity constraints as described in section 2) and the assigned neighborhood does not differ in ways that matter for our outcome. Alternatively, we can use the instrument within zones to rule out that structural differences, such as distance to city center, between zones in a municipality matter. Section 4.3 discusses the instrument in more detail and shows tests of random assignment to distance, of relevance and of monotonicity. Before turning to the identifying assumptions, we briefly describe the equation used to estimate the effect of language skills on socio-economic outcomes:

$$y_{5,t} = \alpha + \beta CEF R_{3,t} + X'_0 \gamma + \omega_c + \mu_z + \varepsilon_{5,t}. \quad (4)$$

This equation relates outcomes in year five after placement ($s = 5$) to language skills after the Integration Program has ended ($s = 3$).²⁵ We look at employment, education and mobility out of the assigned neighborhood. The first stage has the same right-hand side as equation 3 and $CEFR_{3,t}$ as the dependent variable:

$$CEFR_{3,t} = \alpha^1 + \beta^1 Distance_{0,t} + X'_{0,t} \gamma^1 + \omega_t^1 + \mu_z^1 + \varepsilon_{3,t}^1. \quad (5)$$

All standard errors are robust standard errors clustered at the municipality of initial placement.

4.3 Identification

The natural experiment we have in mind is one where refugees are quasi-randomly placed in a location at different points in time and due to the gradual expansion of language training centers and the opening and closing of centers (described in section 3.2), they have different

²⁵Section 8 considers the effect on outcomes over time since program completion.

distances to language training. This is our most demanding specification, which exploits a small subset of refugees and neighborhoods in Denmark. We therefore also run a specification (with municipality fixed effects instead of zone fixed effects), where the location may not be identical. But as we will show below neighborhood characteristics, that we think matter for refugee integration, are not correlated with distance. Together we think these two main types of estimations represent powerful evidence that the effects we estimate in section 5 to section 8 are causal and robust. The rest of this section discusses the validity, the relevance and the monotonicity of the instrument.

Table 2 shows summary statistics for our sample of refugees in column 1 and RCT-type tests of the randomness of distance in columns 2 to 5. All characteristics of the refugees and of their assigned zone are measured at arrival ($s = 0$). Refugees in our sample are young (30.5 years old on average), 59 percent are men and 55 percent are married at the time of placement in a municipality. Many arrive with children, especially young children. The top five countries of origin are Syria (22 percent), Afghanistan (14 percent), Iran (12 percent), Myanmar (9 percent) and Iraq (8 percent). 31 percent of refugees are assigned to Danish I, 52 percent are in the intermediate group, and 17 percent are assigned to Danish III.

Several of the individual characteristics and the share of foreign born in the neighborhood predict language proficiency (column 2 and 3). The same characteristics are essentially unrelated to distance to language training (column 4 and 5). Only one variable is significantly correlated with distance (which can be expected by chance) and the coefficient is economically small; women travel 0.022 hours (1.3 minutes) less to language training (column 4), and refugees with school age children travel 0.016 hours less (column 5). This supports the hypothesis that availability of public housing the date the refugee is assigned to the municipality (see section 2), rather than individual characteristics, is determining the exact location of a refugee within the municipality, and that this variation is as good as random because of the lack of public housing. It is especially reassuring that variables that are likely correlated with pre-migration education

and work-experience such as the assigned language training track and the country of origin of the refugees do not correlate with distance. Being assigned to an immigrant dense neighborhood decreases language proficiency (consistent with recent findings from Damm et al., 2022). An increase in the foreign born share from the 25th to the 75th percentile (4.2 percentage points) decreases proficiency significantly by $0.042 \times 1.466 = 0.06$ and is associated with an insignificant $0.042 \times 1.225 \times 60 = 3$ minutes longer commute. Hence, we conclude that distance is as good as random once municipality or more detailed zone fixed effects are accounted for.

Figure 2 shows the density of our instrument (left y axis) and its correlation with language proficiency (right y axis) in the specification with municipality fixed effects.²⁶ 80 percent of the refugees commute between 10 (p10) and 76 (p90) minutes to language training, and 50 percent commute between 17 (p25) and 55 (p75) minutes. The locally fitted regression of language proficiency on distance show a clear downward slope. Furthermore, it seems reasonable to assume that language proficiency is monotonically decreasing in distance and the relationship is close to linear.²⁷ Unsurprisingly, the slope is more precisely estimated for travel times between 10 and 80 minutes where the density of our instrument is good, while we get far more noise towards the tails of the distribution of our instrument due to few observations and possibly also due to measurement error at small and very large distances. Travel time by public transport may be a poor proxy for the actual time it takes for the refugee to commute to training when zones are close and especially in cases where the refugee attend language training in the same zone as their zone of residence. Our measure of travel time could also be less precise for very long distances, for instance, because small changes in the connections with public transport mean a lot (we do not have time variation in the distance measure).

²⁶We first residualize the variables and then add back the mean. The graph based on the specification with zone fixed effects looks similar but more compressed. For instance, the difference between the 10th and the 90th percentile shrinks to less than half of the corresponding difference in the municipality fixed effect specification.

²⁷The slope coefficient of the fitted line based on equation 3 is shown in Table 3 and discussed in the next section.

5 The Impact of Hours of Class on Language Proficiency

Language training is provided to refugees free of charge up to a maximum duration of three years from arrival (with possibilities for extensions due to sickness or parental leave). Most refugees complete the program within this time frame and we do not observe additional progression for them. However, refugees who do not pass the exam required for permanent residency have an incentive to take the exam at a later date.²⁸ The observable spread in language skills, therefore, diminish over time but it is unclear whether this reflects a real convergence in language skills or whether those who became proficient early continue to improve their language (due to the self-productivity of language skills).²⁹ Hence, language skills should be measured by the end of the Integration Program in (s=3) in order to capture the full effect of commuting time on language skills.

Table 3 reports the OLS (equation 2), 2SLS and first stage (equation 3) estimates as well as the reduced form effect of distance on language skills, which corresponds to our first stage when we turn to the impact of proficiency on socio-economic integration. We report three types of specifications: one with municipality fixed effects (column 1), one with municipality fixed effects and zone controls (column 2), and one with zone fixed effects (column 3). All the regressions use the cumulative number of hours and the skills measured at the end of year 3. Appendix Table A.3 shows similar regressions for each year from one to six. A couple of insights emerge from these tables.

First, our first stage (equation 3) is sufficiently strong with F-test statistics between 20 and 25 in year 3. The coefficient suggests that refugees who live one hour further away from training accumulate between 46 and 71 hours less, depending on the specification. The coefficient is

²⁸Refugees in our sample, who completed the Integration Program by November 29, 2006, are eligible for permanent residency after 7 years of residence in Denmark provided that they pass the exam the marks completion of Danish I (basic). After this date two additional requirements were added. First, the refugees now had to pass the exam that corresponds to passing Danish II (intermediate level). Second, they need 2.5 years of cumulative employment (in full-time equivalents) to be eligible. See Arendt, Dustmann, and Ku (2021) for details.

²⁹Reduced from estimates in Appendix Table A.3 shows the impact of distance on observed language proficiency measured in year 1 to 6 after program start. The coefficient and the F-test statistic decrease from year 2 to 5 suggesting that the relationship weakens over time. (We have fewer observations when we go beyond 5 years).

relatively stable after year 3 but less precisely estimated, while it is more precise in earlier years (distance has a more significant impact on cumulative hours earlier in the Integration Program) with F-statistics up to 77 and 52, depending on the specification. However, we think the difference in language skills at completion of the Integration Program is the more relevant for the outcomes later on.

Second, the IV estimates are larger than the OLS estimates. Refugees can progress to the next module once they pass the test that marks the end of the module, and less able and motivated refugees may therefore take longer to complete (even within language training tracks). The estimated marginal return to additional hours of language classes is therefore larger, once this negative selection into taking many hours is accounted for. We find that an increase in one hundred hours of language instruction raises the CEFR level by 0.20 to 0.22.

6 The Impact of Language Proficiency on Socio-Economic Integration

6.1 Mobility

The largest immigrant communities are found in the largest cities in Denmark and many refugees eventually leave their assigned municipality to move to one of the largest cities in Denmark. Our refugee population is placed outside the four largest cities in Denmark (see data section). Table 4 shows the effect of language proficiency on the mobility out of the local community where the refugee was initially placed in columns 1 to 3 and on the mobility to one of the four largest urban municipalities in Denmark in columns 4 to 6. Both coefficients are surprisingly stable across the specifications that exploits different variation in the data; within municipalities and within zones. The correlation between language proficiency and mobility is negative and quite strong. An increase in language proficiency by one tenth of a CEFR level is associated with a decrease in mobility out of the assigned neighborhood equal to 0.1 percentage point

and a decrease in mobility into the large cities of almost the same magnitude, using the OLS estimates.³⁰

The 2SLS estimates are even larger in magnitude and statistically very significant. This is consistent with a story where the refugees who are less able to learn the language are more likely to find residence in a minority-language enclave in a large city attractive and once this negative selection is taken care of (in the 2SLS estimates), we find an even larger role of language proficiency in integrating the refugees into the local community of assignment. An increase in language proficiency of one tenth of a CEFR level decreases the mobility rate out of the assigned neighborhood (after completion of language training) by 2 percentage points or 7 percent of the baseline and a reduction in the probability of living in one of the main cities of Denmark of 3 percentage points or 33 percent of the baseline. The 2SLS estimates seem large but not unreasonable. However, it is possible that the compliers are a special group whose mobility towards the most urban municipalities is very sensitive to the acquired language skills.

The lower mobility towards the largest cities may reduce the economic returns to the acquired language skills. Arendt et al. (2021) find that the long-run improvement in employment and earnings of refugees with access to more language training are small or absent in rural areas. The literature on the impact of the local conditions on refugee integration finds that refugees, who are placed in areas with worse employment opportunities, have worse labor market outcomes.³¹ Hence, the placement outside the largest urban municipalities with dense labor markets and the lower mobility towards them likely reduce the economic gains from improved language skills in our population. The next section studies the impact of language skills on key indicators of economic success in the host country.

³⁰An increase in language skills by one tenth of a CEFR level corresponds to the improvement generated by a decrease in distance by one hour (Table 3, “reduced form”).

³¹See Edin, Fredriksson, and Åslund (2004); Åslund and Rooth (2007); Damm and Rosholm (2010); Godøy (2017); Azlor, Damm, and Schultz-Nielsen (2020); Aksoy, Poutvaara, and Schikora (2021).

6.2 Employment and Education

Table 5 shows the impact of language proficiency on the employment and human capital investments of refugees. Columns 1 to 3 show the effect on the employment probability, and columns 4 to 6 show the effect on the probability that the refugee is studying (beyond the language training). The OLS results of Table 5 show that increasing Danish proficiency (CEFR level) by one unit is associated with an increase in the employment probability by 7.5-7.7 percentage points and an increase in the probability of studying in Denmark by 7.2.

Turning to the 2SLS estimates, the employment effect is no longer significant. Using within municipality variation in column 1 and 2, we find that a one unit increase in the CEFR level increases the likelihood of being employed by roughly six instead of by seven percentage points. This suggest that unobserved ability bias the OLS estimates upward in column 1 and 2, but the bias is small in our context. Furthermore, ability bias seem to play a minor role when comparing refugees placed in the same zone (column 3) or when analyzing enrollment (column 4 and 6), where the 2SLS estimates are roughly double the size of the OLS estimates. This suggest that classical measurement error in the CEFR levels generate a downward bias, which is large enough to offset potential ability bias in the OLS estimates. To give an example of the severity of attenuation bias that results from measurement error in the prior literature, Bleakley and Chin (2004) and Dustmann and van Soet (2002) conclude that the attenuation bias in their case is so severe that the IV coefficients are between two and three times their uncorrected OLS estimates.

The impact of language proficiency on enrollment in education programs in Denmark is strongly significant using within municipality variation in the data. In our most demanding specification that use a small subset of zones with variation in distance over time, we get a very similar point estimate that is significant at the 10 percent level.³²

³²We also generated a dummy for completion of an education program. We think our time horizon is too short to capture completion of programs that typically last two to four years after completion of language training. The coefficient of completed formal education is positive but not significant. Completion of an education program count as an alternative proof of language proficiency if applying for permanent residency. Educations are free in

Our employment variable captures the extensive margin. We also checked the intensive margin and the hourly wages and were unable to detect an impact on hours worked or hourly wages of employed refugees. The positive but not significant employment estimates are consistent with Arendt et al. (2021); Lochmann, Rapoport, and Speciale (2019), who also find insignificant employment effects in the first couple of years post treatment. Arendt et al. (2021) estimate a gradual increase in the employment probability in the years after completion of language training and find that it is marginally significant from around eight years post placement in the municipality. It is also worth noting, that the large influx into education programs contributes to delaying any potential labor market effect, and that decreased mobility towards the large cities (documented in Section 6.1) might decrease the labor market returns to language skills (Arendt et al., 2021).

7 Further Robustness Checks and Placebo Tests

Figure 3 shows the robustness of our main results to alternative instruments and check that outcomes of natives, who should not be affected by travel distance to language training, are not correlated with distance. We estimate precise zero effects for natives.³³ The coefficients are essentially zero, a few of them are on the margin of significance due to the very large sample size for the native born populations.

The hollow circle and the black bullet replicate our main result from the reduced form regressions with the municipality and the zone fixed effects, respectively. The hollow and black square show the same for an alternative instrument that simply distinguishes distances above

Denmark, and students receive study grants or are paid by the employer in the work-based learning programs. Hence, studying could be attractive route for talented refugees and it is possible that our OLS estimates of the impact of language proficiency on enrollment are negatively biased (reverse of the usual ability bias), if they are able to leave language training and unemployment to study without completing the language training. Since education programs in Denmark (with the exception of a few Masters) are taught in Danish it is not entirely clear how refugees would do this even if they might be accepted without proof of their Danish language skills. Notice, also that this only creates potential bias in the OLS estimates and not in the IV.

³³To estimate the effect of distance for natives, we take a random 10 percent sample of natives each year and assign them the distance to the most common language center for refugees who were placed in the zone in that year.

or below the median. This instrument throws out information that is useful and generates more noisy estimates. The hollow and the black triangle use log distance instead of distance and produce quite similar results. Overall, the results are robust to transformation of the instrument such as the log, but we lose precision if we throw out too much of the variation in distance. For instance, by using a simple dummy for distances above the median. This is especially true if we simultaneously restrict the estimations to variation within zones.

8 Persistence

Figure 4 look at the impact over time using a larger sample of refugees arriving from 1999 to 2013. The coefficients in year 5 for enrollment are comparable to the main sample (arrival cohorts between 2003 and 2013). An increase in distance by one hour (reduces language proficiency and) decreases enrollment by roughly 2 percentage points. The coefficients are larger for mobility to the main cities when we expand the sample period. The impact on enrollment in educations disappear from year eight, while the impact on residence in the large urban municipalities shrinks over time but remain large and significant. The impact on employment does not become significant in this larger sample nor over the extended time period.

9 Conclusion

Despite profound interest in the importance of language skills for the integration of immigrants, we know little about the effectiveness of language training on the acquisition of the host country language or the causal effect of the acquired language skills on important socio-economic outcomes. This paper makes progress on both these fronts.

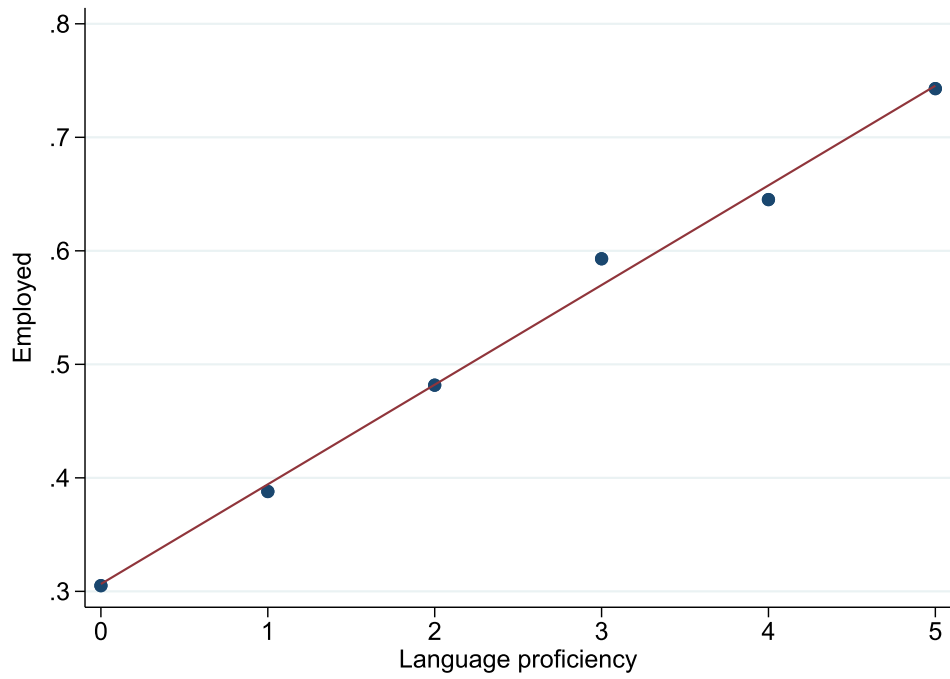
We first show that the marginal effect of additional hours of class on proficiency in the host country's language is positive and significant. The marginal effect is positive in all three year of the Danish integration program, suggesting that language classes are more productive than

the counterfactual (learning the language by living and working in the country), even in the second and third years after some basic knowledge about the language has been acquired. To overcome negative selection into attending more hours, we instrument attendance with travel time by public transport to the assigned language training center. The impact of distance on language skills is also of policy interest. It informs us about the (semi-)elasticity of language skills with respect to travel distance - a parameter that policy makers can influence directly. We show that a one hour decrease in commuting time increases proficiency by the end of the Danish Integration Program by 4 to 6 percent relative to the sample mean. Hence, the accessibility of language classes matters.

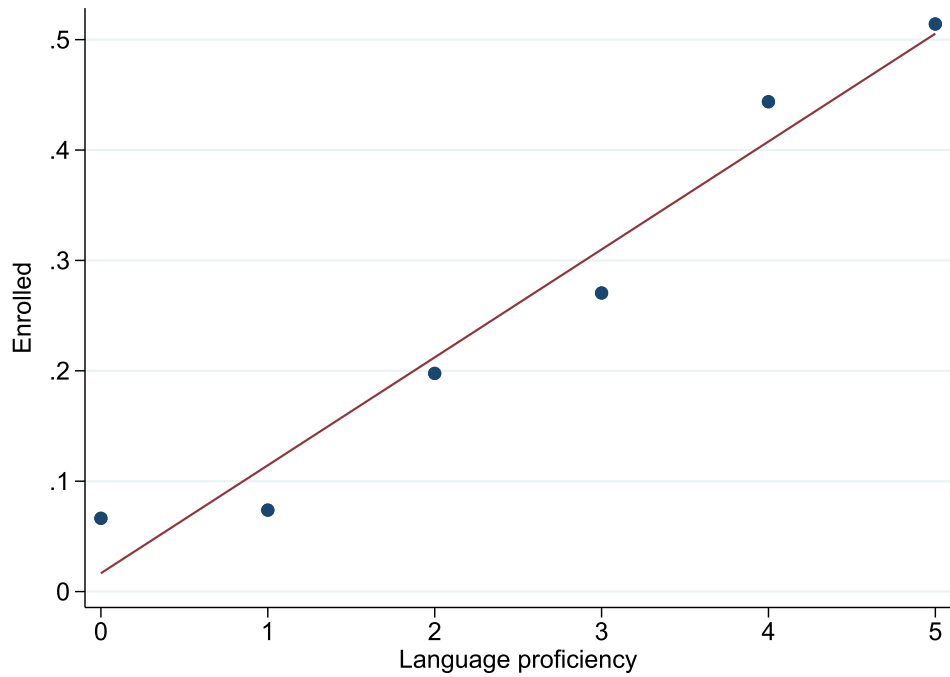
Next we use our instrument, exogenous variation in distance, to look at the impact of language capital on socio-economic outcomes. We find that immigrants who become more fluent in Danish further upgrade their skills in the free Danish education system. This likely postpones the employment effect and we do not estimate a significant impact on employment within the first ten years after placement.³⁴ Finally, we find that improving refugees' language skills facilitates integration into the local community as measured by lower mobility to the main cities in Denmark with large pre-existing immigrant populations. Hence, increasing language proficiency fosters social and economic integration.

³⁴This is consistent with the papers estimating a causal effect of language training. Sarvimäki and Hämäläinen (2016) finds large effects on cumulative earnings but do not detect any significant effect on employment in the first 10 years after program start. Lochmann, Rapoport, and Speciale (2019) finds an effect on labor force participation but no effect on employment looking only a few years after participation in language classes. Arendt et al. (2021) find that both earnings and employment increase in the long run as a consequence of more hours of language training but the effect emerge gradually and is only on the margin of significance 10 years after program start.

10 Tables and Figures



(a) Share Employed by Language Proficiency Levels



(b) Share Enrolled by Language Proficiency Levels

Figure 1: Economic Integration and Language Proficiency

Notes: Employment and enrolment are measured in year five and language proficiency is measured at the end of year three after arrival. See section 3 for detailed variable definitions.

Table 2: Descriptive Statistics and Quasi-Random Assignment to Distance

	Summary statistics (1)	Language proficiency (2)	Language proficiency (3)	Distance (4)	Distance (5)
Age	30.45 (7.76)	0.024 (0.015)	0.014 (0.016)	0.009 (0.006)	0.006 (0.006)
Age squared	987.49 (505.85)	-0.001** (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
Female	0.41 (0.49)	-0.133*** (0.029)	-0.149*** (0.028)	-0.022* (0.011)	-0.015 (0.011)
Married	0.55 (0.50)	0.039 (0.036)	0.037 (0.040)	-0.009 (0.014)	-0.003 (0.017)
Number of children (0-5)	0.41 (0.71)	-0.012 (0.021)	-0.005 (0.023)	-0.006 (0.009)	-0.004 (0.005)
Number of children (6-12)	0.40 (0.80)	0.045*** (0.017)	0.041** (0.018)	-0.014 (0.009)	-0.016** (0.008)
Number of children (13-17)	0.18 (0.55)	-0.007 (0.024)	-0.001 (0.025)	-0.013 (0.011)	-0.010 (0.011)
Danish II	0.52 (0.50)	0.893*** (0.033)	0.887*** (0.035)	-0.005 (0.013)	-0.001 (0.014)
Danish III	0.17 (0.38)	1.548*** (0.048)	1.536*** (0.050)	0.005 (0.024)	-0.007 (0.022)
Myanmar	0.09 (0.28)	-0.007 (0.106)	-0.010 (0.101)	0.005 (0.053)	-0.013 (0.050)
Syria	0.22 (0.41)	0.012 (0.053)	0.024 (0.056)	0.005 (0.037)	-0.011 (0.031)
Iraq	0.08 (0.27)	-0.341*** (0.085)	-0.331*** (0.088)	0.049 (0.033)	0.022 (0.036)
Iran	0.12 (0.33)	-0.015 (0.061)	-0.004 (0.066)	-0.017 (0.023)	-0.015 (0.021)
Other	0.36 (0.48)	-0.092* (0.054)	-0.087 (0.054)	-0.032 (0.027)	-0.039 (0.026)
2004	0.06 (0.24)	0.111 (0.095)	0.112 (0.097)	0.050 (0.061)	0.050 (0.062)
2005	0.05 (0.22)	0.183 (0.115)	0.188* (0.106)	0.031 (0.042)	-0.007 (0.037)
2006	0.04 (0.20)	0.203* (0.120)	0.189 (0.122)	0.041 (0.059)	0.067 (0.056)
2007	0.05 (0.22)	0.344*** (0.115)	0.321*** (0.112)	-0.089 (0.064)	0.003 (0.047)
2008	0.06 (0.23)	0.416*** (0.094)	0.376*** (0.104)	-0.090 (0.062)	0.038 (0.057)
2009	0.06 (0.24)	0.590*** (0.090)	0.552*** (0.102)	-0.034 (0.060)	0.049 (0.054)
2010	0.11 (0.31)	0.763*** (0.088)	0.712*** (0.089)	-0.053 (0.058)	-0.014 (0.052)
2011	0.13 (0.33)	0.636*** (0.091)	0.600*** (0.089)	-0.026 (0.060)	0.027 (0.055)
2012	0.15 (0.35)	0.624*** (0.090)	0.554*** (0.090)	-0.034 (0.076)	0.035 (0.069)
2013	0.20 (0.40)	0.670*** (0.085)	0.590*** (0.086)	-0.059 (0.085)	0.013 (0.072)
Zone foreign born share	0.06 (0.04)	-1.466* (0.782)	.	1.225 (0.766)	.
Zone employment rate	0.52 (0.04)	-0.052 (0.578)	.	0.907 (0.715)	.
Observations	10170	10170	10170	10170	10170
Municipality fixed effects		Yes	No	Yes	No
Zone fixed effects		No	Yes	No	Yes

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column 1 reports the mean and the standard deviation (in parentheses) for each of the variables shown in the rows. Columns 2 to 5 report the coefficients and the standard errors (in parentheses) from regressions of language proficiency or distance on the observable characteristics. The reference is a single man with no children arriving in 2003 from Afghanistan and assigned to Danish I.

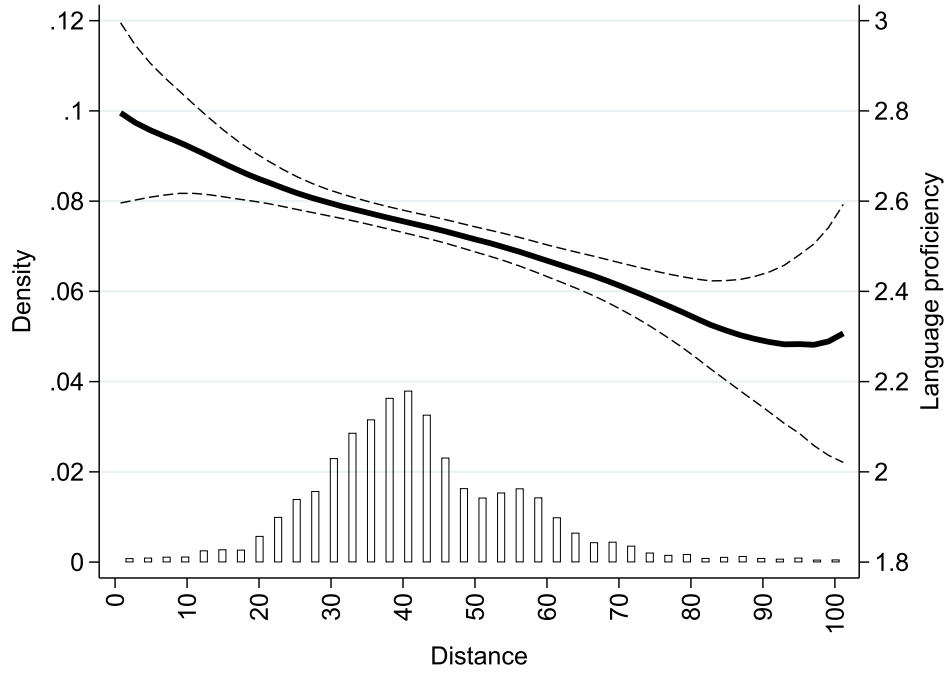


Figure 2: First Stage Graph of Language Proficiency on Distance

Notes: First-stage partial plot and density of instrument. Distance is travel time by public transport measured in minutes. The solid line is a local linear regression of language proficiency measured by the end of year three after arrival (right y axis) on distance (x axis), where both variables are the mean-standardized residuals from the specification with municipality fixed effects. Dashed lines show 90% confidence intervals. The histogram is the density of distance (left y axis). Top and bottom 2.5 percent of distance are excluded in the graph.

Table 3: The Impact of Hours of Class on Language Proficiency

	(1)	(2)	(3)
OLS	0.168*** (0.005)	0.168*** (0.005)	0.167*** (0.005)
IV	0.201*** (0.031)	0.200*** (0.032)	0.218*** (0.048)
First Stage	-0.712*** (0.143)	-0.702*** (0.141)	-0.463*** (0.104)
F-stat	24.93	24.71	19.97
P-value	0.00	0.00	0.00
Reduced Form	-0.143*** (0.028)	-0.140*** (0.027)	-0.101*** (0.023)
F-stat	25.87	26.15	19.18
P-value	0	0	0.00
Observations	10170	10170	10170
Outcome mean	2.54	2.54	2.54
Municipality fixed effects	Yes	Yes	No
Zone controls	No	Yes	No
Zone fixed effects	No	No	Yes

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Each table entry is a point estimate and standard error from a separate regression. OLS is based on equation 2 and the first stage is based on equation 3. Hours of class is divided by 100. Distance is in hours.

Table 4: The Impact of Language Proficiency on Integration Into the Local Community

	Left Zone (After Year 3)			Living in Main City		
	(1)	(2)	(3)	(4)	(5)	(6)
OLS	-0.008** (0.003)	-0.009** (0.003)	-0.011*** (0.003)	-0.008*** (0.003)	-0.007** (0.003)	-0.006* (0.003)
IV	-0.211*** (0.076)	-0.217*** (0.073)	-0.238** (0.115)	-0.340*** (0.077)	-0.324*** (0.075)	-0.215** (0.095)
F-stat	25.87	26.15	19.18	25.87	26.15	19.18
P-value	0.00	0.00	0.00	0.00	0.00	0.00
Reduced Form	0.030*** (0.011)	0.030*** (0.010)	0.024* (0.013)	0.049*** (0.013)	0.045*** (0.012)	0.022** (0.011)
Observations	10170	10170	10170	10170	10170	10170
Outcome mean	0.29	0.29	0.29	0.10	0.10	0.10
Zone controls	No	Yes	No	No	Yes	No
Municipality fixed effects	Yes	Yes	No	Yes	Yes	No
Zone fixed effects	No	No	Yes	No	No	Yes

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Each table entry is a point estimate and standard error (in parentheses) from a separate regression. OLS is based on equation 4 and the first stage is based on equation 5. Distance is in hours.

Table 5: The Impact of Language Proficiency on Economic Integration

	Employed			Enrolled		
	(1)	(2)	(3)	(4)	(5)	(6)
OLS	0.075*** (0.004)	0.075*** (0.004)	0.077*** (0.004)	0.072*** (0.003)	0.072*** (0.003)	0.072*** (0.003)
IV	0.056 (0.063)	0.061 (0.064)	0.123 (0.083)	0.166*** (0.046)	0.166*** (0.047)	0.155* (0.092)
F-stat	25.87	26.15	19.18	25.87	26.15	19.18
P-value	0.00	0.00	0.00	0.00	0.00	0.00
Reduced Form	-0.008 (0.009)	-0.009 (0.009)	-0.012 (0.009)	-0.024*** (0.005)	-0.023*** (0.005)	-0.016* (0.008)
Observations	10170	10170	10170	10170	10170	10170
Outcome mean	0.53	0.53	0.53	0.26	0.26	0.26
Zone controls	No	Yes	No	No	Yes	No
Municipality fixed effects	Yes	Yes	No	Yes	Yes	No
Zone fixed effects	No	No	Yes	No	No	Yes

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Each table entry is a point estimate and standard error (in parentheses) from a separate regression. OLS is based on equation 4 and the first stage is based on equation 5. Distance is in hours.

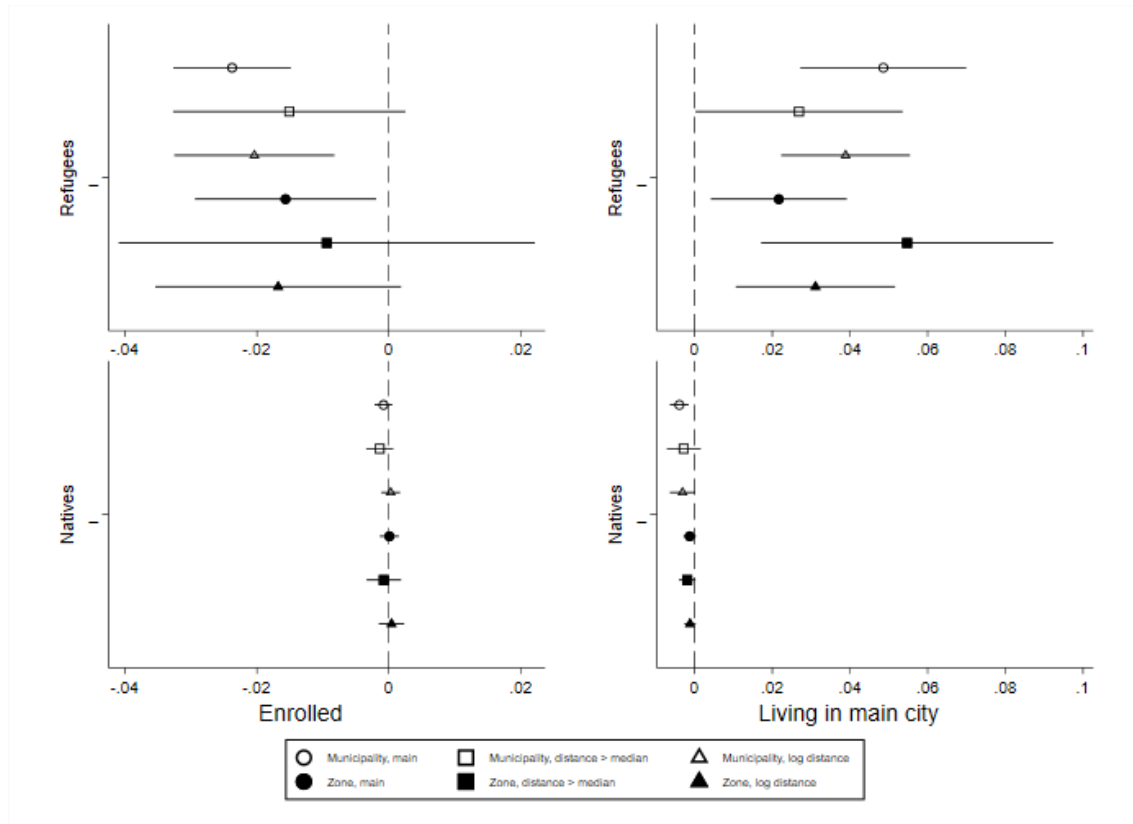
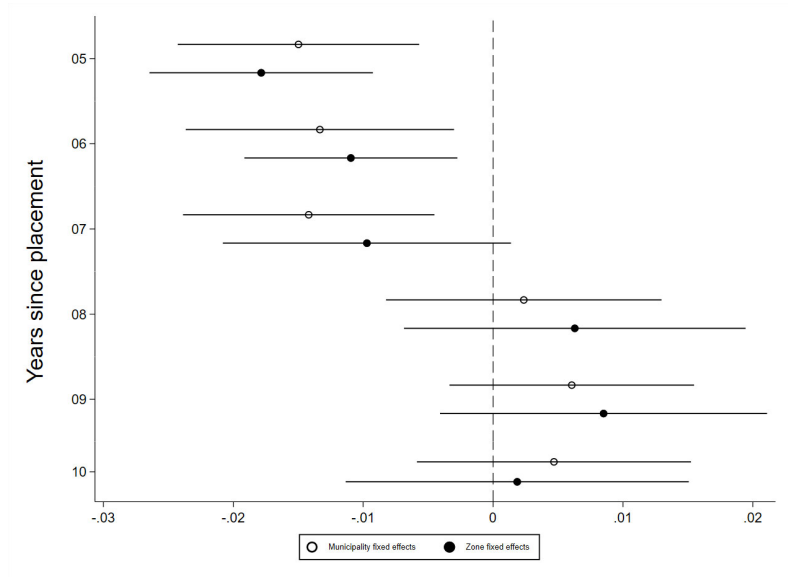
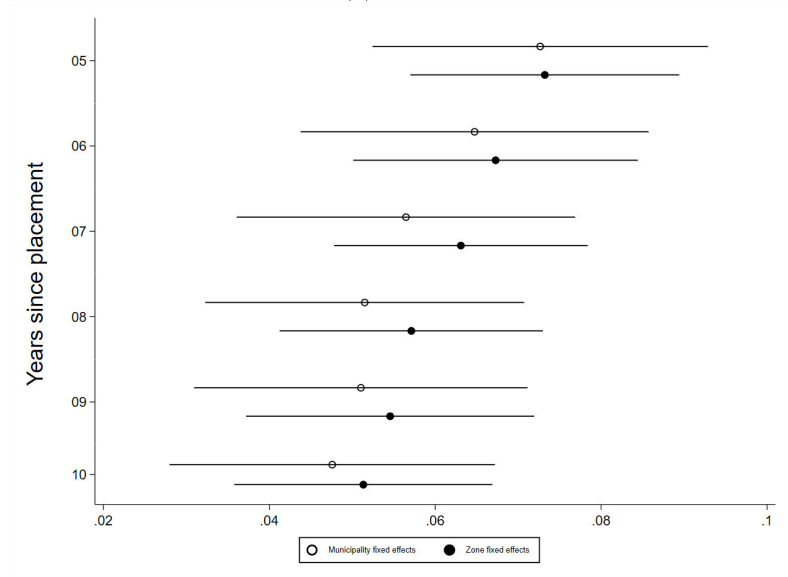


Figure 3: Placebo and Robustness Checks

Notes: The graph shows point estimates and 90-percent confidence intervals of the impact of distance to language training for refugees and natives with similar demographic characteristics.



(a) Enrolled



(b) Living in Main City

Figure 4: Persistence

Notes: The graph shows point estimates and 90-percent confidence intervals of the impact of distance on refugee admitted between 1999 and 2013 (larger sample than the remaining analysis).

References

- Akresh, Ilana Redstone and Reanne Frank. 2011. “At the intersection of self and other: English language ability and immigrant labor market outcomes.” *Social Science Research* 40 (5):1362–1370.
- Akresh, Ilana Redstone, Douglas S Massey, and Reanne Frank. 2014. “Beyond English proficiency: Rethinking immigrant integration.” *Social science research* 45:200–210.
- Aksoy, Cevat Giray, Panu Poutvaara, and Felicitas Schikora. 2021. “First Time around: Local Conditions and Multi-Dimensional Integration of Refugees.” ifo Working Papers 361, ifo Institute, Munich.
- Altmann, Steffen, Anita Glenney, Robert Mahlstedt, and Alexander Sebald. 2022. “Digital Tools to Facilitate Job Search.” Tech. rep., University of Copenhagen.
- Aoki, Yu and Lualhati Santiago. 2018. “Deprivation, Segregation, and Socioeconomic Class of UK Immigrants: Does English Proficiency Matter?” .
- Arendt, Jacob Nielsen. 2022. “Labor Market Effects of a Work-First Policy for Refugees.” *Journal of Population Economics* 35:169–196.
- Arendt, Jacob Nielsen, Iben Bolvig, Mette Foged, Linea Hasager, and Giovanni Peri. 2021. “Language Training and Refugees’ Integration.” Tech. rep., National Bureau of Economic Research.
- Arendt, Jacob Nielsen, Christian Dustmann, and Hyejin Ku. 2021. “Permanent Residency and Refugee Immigrants’ Skill Investment.” *Rockwool Foundation Research Unit, Study Paper* 160.
- Azlor, Luz, Anna Piil Damm, and Marie Louise Schultz-Nielsen. 2020. “Local Labor Demand and Immigrant Employment.” *Labour Economics* 63.
- Bansak, Kirk, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, and Duncan Lawrence. 2018. “Improving Refugee Integration Through Data-Driven Algorithmic Assignment.” *Science* 359 (6373):325–329.
- Battisti, Michele, Yvonne Giesing, and Nadzeya Laurentsyevea. 2019. “Can Job Search Assistance Improve the Labour Market Integration of Refugees? Evidence from a Field Experiment.” *Labour Economics* 61:101745.
- Belot, Michèle, Philipp Kircher, and Paul Muller. 2018. “Providing Advice to Jobseekers at Low Cost: An Experimental Study on Online Advice.” *The Review of Economic Studies* 86 (4):1411–1447.
- Bleakley, Hoyt and Aimee Chin. 2004. “Language Skills and Earnings: Evidence from Childhood Immigrants.” *Review of Economics and Statistics* 86 (2):481–494.
- . 2010. “Age at arrival, English proficiency, and social assimilation among US immigrants.” *American Economic Journal: Applied Economics* 2 (1):165–92.
- Borjas, George. 1985. “Assimilation, Changes in Cohort Quality and the Earnings of Immigrants.” *Journal of Labor Economics* 3 (4):463–489.
- Brell, Courtney, Christian Dustmann, and Ian Preston. 2020. “The Labor Market Integration of Refugee Migrants in High-Income Countries.” *Journal of Economic Perspectives* 34 (1):94–121.

- Card, David. 1995. "Using geographic variation in college proximity to estimate the return to schooling." In *Aspects of labour market behaviour: essays in honour of John Vanderkam*, edited by Robert Swidinsky Louis N. Christofides, E. Kenneth Grant. Toronto, Canada: University of Toronto Press, 201–222.
- . 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69:1127–1160.
- Chiswick, Barry R. 1978. "The Effect of Americanization on the Earnings of Foreign-Born Men." *Journal of Political Economy* 86 (5):897–921.
- . 1991. "Speaking, Reading, and Earnings Among Low-Skilled Immigrants." *Journal of Labor Economics* 9:149–170.
- Chiswick, Barry R. and P. W. Miller. 1995. "The Endogeneity between Language and Earnings: International Analyses." *Journal of Labor Economics* 13 (3):246–288.
- Chiswick, Barry R and Zhiling Wang. 2019. "Social Contacts, Dutch Language Proficiency and Immigrant Economic Performance in the Netherlands." Tech. rep., GLO Discussion Paper.
- Clausen, Jens, Eskil Heinesen, Hans Hummelgaard, Leif Husted, and Michael Rosholm. 2009. "The Effect of Integration Policies on the Time Until Regular Employment of Newly Arrived Immigrants: Evidence from Denmark." *Labour Economics* 16 (4):409–417.
- Cunha, Flavio and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review: Papers & Proceedings* 97 (2):31–47.
- Damm, Anna Piil, Ahmad Hassani, Trine Skriver Høholt Jensen, and Marie Louise Schultz-Nielsen. 2022. "Co-Ethnic Neighbors and Investment in Host-Country Language Skills." Study Paper no. 169. The Rockwool Foundation Research Unit.
- Damm, Anna Piil and Michael Rosholm. 2010. "Employment effects of spatial dispersal of refugees." *Review of Economics of the Household* 8 (1):105–146.
- Dustmann, Christian. 1994. "Speaking Fluency, Writing Fluency and Earnings of Migrants." *Journal of Population Economics* 7 (2):133–156.
- Dustmann, Christian and Francesca Fabbri. 2003. "Language Proficiency and Labour Market Performance of Immigrants in the UK." *The Economic Journal* 113 (July):695–717.
- Dustmann, Christian and Arthur van Soet. 2001. "Language Fluency and Earnings: Estimation With Misclassified Language Indicators." *The Review of Economics and Statistics* 83 (4):663–674.
- . 2002. "Language and the Earnings of Immigrants." *ILR Review* 55 (3):473–492.
- Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. 2004. "Settlement Policies and the Economic Success of Immigrants." *Journal of Population Economics* 17 (1):133–155.
- Godøy, Anna. 2017. "Local labor markets and earnings of refugee immigrants." *Empirical Economics* 52 (1):31–58.
- Heller, Blake and Kirsten Slungaard Mumma. 2022. "Immigrant Integration in the United States: The Role of Adult English Language Training." *American Economic Journal: Economic Policy* Revise and resubmit.

- Lochmann, Alexia, Hillel Rapoport, and Biagio Speciale. 2019. “The Effect of Language Training on Immigrants’ Economic Integration: Empirical Evidence from France.” 113:165–296. *European Economic Review*.
- Nielsen, C. P. and K. B. Jensen. 2006. “The Danish Integration Act’s Significance For the Settlement Patterns of Refugees.” *Institute for Local Government Studies, Copenhagen* .
- Rich, Jeppe and Christian Overgaard Hansen. 2016. “The Danish national passenger model – Model specification and results.” *European Journal of Transport and Infrastructure Research* 16 (4).
- Sarvimäki, Matti and Kari Hämäläinen. 2016. “Integrating Immigrants: The Impact of Restructuring Active Labor Market Programs.” *Journal of Labor Economics* 34 (2):479–508.
- Schultz-Nielsen, Marie Louise. 2017. “Labor Market Integration of Refugees in Denmark.” *Nordic Economic Policy Review* :55 – 90.
- Åslund, Olof and Dan-Olof Rooth. 2007. “Do When and Where Matter? Initial Labour Market Conditions and Immigrant Earnings.” *The Economic Journal* 117 (518):422–448.

A Appendix

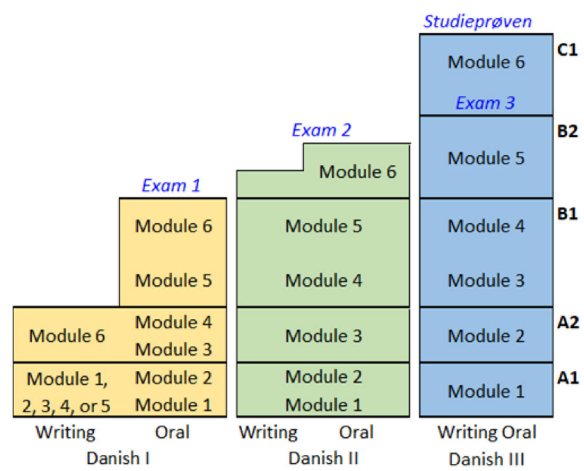


Figure A.1: Mapping Between Modules and CEFR Levels
Source: Ministry of Immigration and Integration.

Table A.1: Description of CEFR Levels

Proficient user	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
Independent user	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
Basic user	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Source: Council of Europe, the Common European Framework of Reference for Languages (CEFR), Global Scale - Table 1.

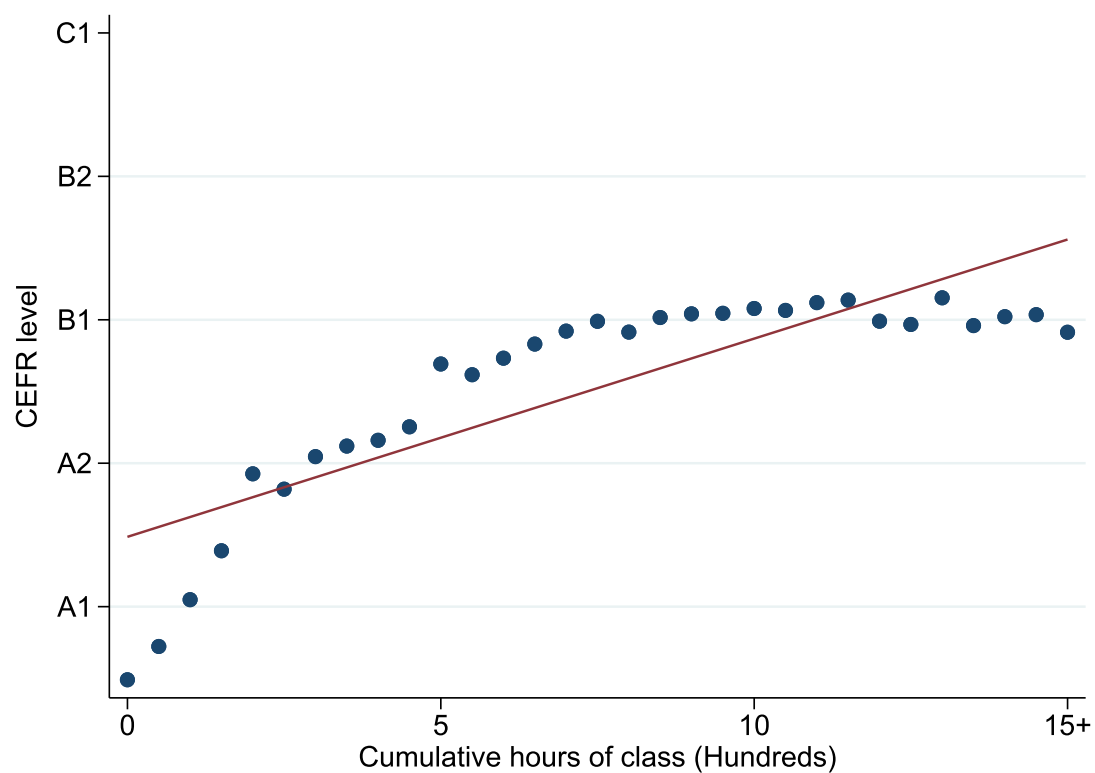


Figure A.2: Language Proficiency by hours of class

Notes: The hours of class attended in year zero to three (50 hours bins).

Table A.2: Hours of Class and Language Skills During the Integration Program

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: OLS</i>						
Hours in t=0	0.199*** (0.012)				0.164*** (0.012)	
Hours in t=1		0.327*** (0.014)			0.201*** (0.014)	
Hours in t=2			0.322*** (0.013)		0.217*** (0.012)	
Hours in t=3				0.104*** (0.014)	0.020* (0.011)	
Cumulative hours in t=3						0.168*** (0.005)
<i>Panel B: IV</i>						
Hours in t=0	0.851*** (0.154)					
Hours in t=1		0.379*** (0.072)				
Hours in t=2			0.887*** (0.236)			
Hours in t=3				27.392 (270.055)		
Cumulative hours in t=3						0.201*** (0.031)
<i>Panel C: First stage</i>						
Public transportation	-0.168*** (0.037)	-0.378*** (0.042)	-0.162*** (0.054)	-0.006 (0.051)		-0.712*** (0.143)
F-stat	20.81	81.95	9.02	.01		24.93
P-value	0	0	.003	.913		0
Observations	10170	10170	10170	10170	10170	10170
Municipality fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

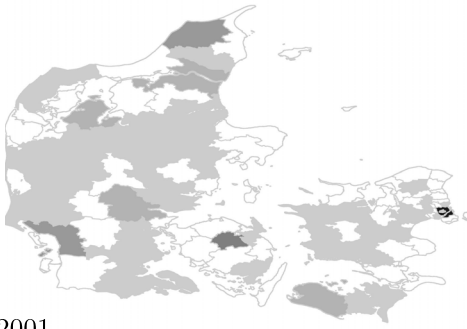
Notes: The outcome is always the language skills at the end of the Integration Program (year 3). Column 1 to 4 show the correlation between the hours attended each year from year 0 to 3 and the final language skills (year 3) estimated with OLS in Panel a and with 2SLS in Panel b. Notice, the instrument does not satisfy the exclusion restriction in column 1 to 4, because the end level is affected by each of the period investments. Column 5 shows all period investments together in one OLS regression. Column 6 replicates column 1 of Table 3 in the main text.

Table A.3: The Impact of Hours of Class on Language Proficiency by Years Since Admission

	Year 1 (1)	Year 2 (2)	Year 3 (3)	Year 4 (4)	Year 5 (5)	Year 6 (6)
Panel a: Municipality fixed effects						
OLS	0.208*** (0.007)	0.193*** (0.006)	0.168*** (0.005)	0.145*** (0.005)	0.130*** (0.005)	0.124*** (0.005)
IV	0.208*** (0.031)	0.244*** (0.039)	0.201*** (0.031)	0.147*** (0.031)	0.112*** (0.034)	0.131*** (0.043)
First Stage	-0.545*** (0.062)	-0.706*** (0.101)	-0.712*** (0.143)	-0.663*** (0.139)	-0.619*** (0.139)	-0.609*** (0.139)
F-stat	76.56	48.71	24.93	22.85	19.90	19.09
P-value	0.00	0.00	0.00	0.00	0.00	0.00
Reduced Form	-0.113*** (0.020)	-0.172*** (0.026)	-0.143*** (0.028)	-0.098*** (0.025)	-0.069*** (0.026)	-0.086*** (0.029)
F-stat	32.61	44.36	25.87	15.73	7.15	9.02
P-value	0.00	0.00	0.00	0.00	.01	.00
Panel a: Municipality fixed effects						
OLS	0.210*** (0.007)	0.194*** (0.006)	0.167*** (0.005)	0.145*** (0.005)	0.130*** (0.005)	0.124*** (0.005)
IV	0.219*** (0.052)	0.268*** (0.052)	0.218*** (0.048)	0.156*** (0.049)	0.113* (0.059)	0.131* (0.073)
First Stage	-0.362*** (0.050)	-0.475*** (0.081)	-0.463*** (0.104)	-0.426*** (0.101)	-0.386*** (0.100)	-0.384*** (0.105)
F-stat	51.62	33.95	19.97	17.82	14.82	13.42
P-value	0.00	0.00	0.00	0.00	0.00	0.00
Reduced Form	-0.079*** (0.023)	-0.127*** (0.032)	-0.101*** (0.023)	-0.067*** (0.024)	-0.044* (0.026)	-0.053* (0.030)
F-stat	12.36	16.22	19.18	7.73	2.79	3.10
P-value	0.00	0.00	0.00	0.01	0.10	0.08
Observations	10170	10170	10170	10170	10170	8180
Outcome mean	1.17	1.99	2.54	2.75	2.84	2.85

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Language proficiency is measured by the CEFR level converted to a scale from 0 to 5. Hours of class is divided by 100. Distance is in hours.

1999



2000



2001



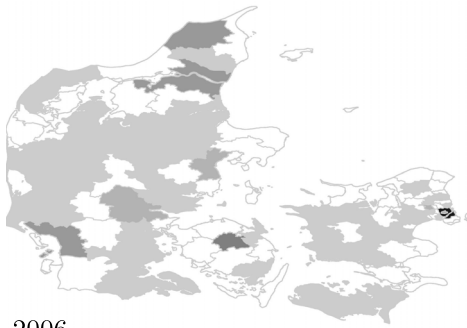
2002



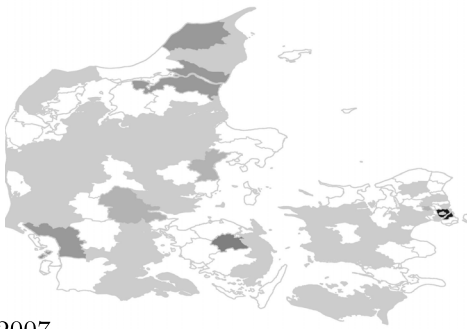
2003



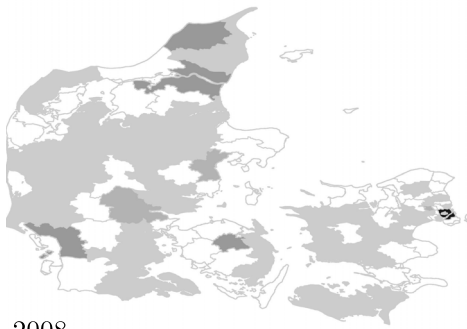
2004



2005



2006



2007



2008



2009



2010



2011



2012



2013



2014



2015



2016



2017



2018

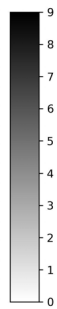


Figure A.3: Spatial Distribution of Language Training Centers (continued)

Notes: The maps show the number of state-approved language training centers in Danish municipalities from 1999 (top left) to 2018 (bottom right).