IDB WORKING PAPER SERIES Nº 01460

Access to Language Training and the Local Integration of Refugees

Mette Foged
Cynthia van der Werf

Inter-American Development Bank
Migration Unit

May 2023

Access to Language Training and the Local Integration of Refugees

Mette Foged*
Cynthia van der Werf**

* University of Copenhagen
** Inter-American Development Bank

May 2023

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Foged, Mette.
Access to language training and the local integration of refugees / Mette Foged, Cynthia van der Werf.
p. cm. — (IDB Working Paper Series; 1460)
Includes bibliographic references.
1. Refugees-Language-Study and teaching-Denmark. 2. Refugees-Social aspects-Denmark. 3. Language and languages-Study and teaching-Denmark. I. Werf, Cynthia van der. II. Inter-American Development Bank. Migration Unit. III. Title. IV. Series.
IDB-WP-1460

http://www.iadb.org

Mette Foged: Mette.Foged@econ.ku.dk

Cynthia van der Werf: cvanderwerf@iadb.org

Access to Language Training and the Local Integration of Refugees*

Mette Foged† and Cynthia van der Werf‡

May 2023

Abstract

This paper examines whether language classes raises refugees' language proficiency and improves their socio-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 conditional on initial placement. Second, we show that a one-hour decrease in commuting time increases total hours of class attended by 46 to 71. Third, we use this novel identification strategy to show that 100 additional hours of language class increases fluency in the Danish language by 8-9 percent, post-language training human capital acquisition by 11-13 percent and improve the integration of the refugees in the communities where they were initially placed, as measured by the lower exit rates from those same communities and an almost 70 percent reduction in 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]

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] This paper contributes to this literature in two important ways. First, we use a novel identification strategy and show that a shorter commute to language training increases the total number of classes attended and language proficiency at the end of the three-year Integration rogram. Second, beyond analysing commonly used measures of economic integration (discussed below), we show that more hours of language training makes refugees better integrated into their local community and less likely to move to the main cities of Denmark with large pre-existing immigrant populations. Specifically, we see that mobility out of the immigrants' originally assigned municipality after the Integration Program and mobility to one of the main cities of Denmark decrease dramatically (by more than 50 percent of the sample mean). 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.[4]

[1] A 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.

[2] E.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.

[3] E.g., Sarvimäki and Hämäläinen (2016); Lochmann, Rapoport, and Speciale (2019); Foged et al. (2023); and Heller and Mumma (2022), which we describe below.

[4] While an even distribution of refugees is a popular policy in many countries, it may not be an economically

Consistent with Foged et al. (2023), we find that language training increases enrollment in education programs in Denmark. Our estimates imply that additional 100 hours of language training increase the enrollment probability by 3 percentage points or 11 to 13 percent of the sample mean.[5] The impact we estimate on the employment probability is statistically insignificant but the magnitude and timing is interesting to compare to the existing literature. We find that an additional 100 hours of language training increases the extensive margin of employment by 1.1 to 2.8 percentage points (2 to 5 percent of the sample mean, statistically insignificant) in year 5 after program start. We also consider alternative employment measures but do not detect a significant effect.

Foged et al. (2023) analyze a reform in Denmark in 1999 that expanded language training for adult immigrants. Their estimated impact on employment in year 5 after program start is in between our estimate (roughly 1.2) and also insignificant.[6] Foged et al. (2023) follow individuals 18 years and find that the impact on employment is marginally significant in the long run and accompanied by significantly higher earnings and a dramatic increase in the share working in occupations requiring good interpersonal communications skills. Lochmann, Rapoport, and Speciale (2019) use a discontinuity in an initial language screening test in France to estimate the impact of additional hours of language training assigned to participants who fail the initial test. Their estimates imply substantial employment effects of 100 hours of language classes 3 years after the onset of treatment.[7] Heller and Mumma (2022) use a lottery for places in the English for Speakers of Other Languages (ESOL) program in Massachusetts and find that individuals who win a seat in ESOL are 3.4 percentage point more likely to have some employer-

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.

[5]The point estimate is similar across specifications and statistical significant in some specifications. The IV estimate turns insignificant and the reduced form estimate is only significant at the 10 percent level in our most demanding specifications (zone fixed effects).

[6]Refugees treated by the reform attend 200 additional hours of language training. We divide their estimate by two such that it is comparable to the estimates reported in this paper.

[7]They estimate that 100 hours of language class increase labor force participation between 14.5 and 26.6 percentage points (Table 2). Employment is not affected conditional on labor force participation (Table 8), suggesting that employment as a share of the population increases roughly proportional to the labor force.

reported earnings in years 2 to 10 after enrollment.[8] Finally, Sarvimäki and Hämäläinen (2016) exploit a reform in Finland that restructured active labor market programs for immigrants to include more language training. The Finnish reform increased cumulative earnings over a 10-year period, but the authors did not find a robust effect on employment.

Our population consists of individuals who were granted refugee status in Denmark between 2003 and 2013 (1999 to 2008 when we analyze the impact up to 10 years after asylum). 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 that starts shortly after the individual is granted asylum (see Foged et al., 2023). 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, who are assigned to different neighborhoods in the same municipality (likely due to the availability of suitable housing at their date of arrival), and 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, access to educations and ethnic networks within the neighborhoods (zones). 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).[9]

[8] We scale the 4.2 percentage point increase in employment probability in Table 5 using the average number of hours (125) ESOL winners enroll (Table 3).

[9] Such an instrument is related to a large literature estimating the marginal returns to education using spatial variation in distance to college as an instrument (Card, 1995, 2001). We improve this identification by considering

We show that our instrument is as good as random for the individual refugees conditional on initial location. 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 and education.

Using our reduced form specification (impact of distance) and earlier arrival cohorts we also look into the effects up to 10 years after asylum. The estimated impact on mobility out of the initial municipality and into the largest cities of Denmark seems to shrink over time but remain large and significant. The effect on enrollment goes to zero by year eight after asylum. We do not detect significant changes in employment or unemployment within the 10 year 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.

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 following a quota system that aims to distribute refugees proportional to municipal inhabitants (see Azlor,

a group whose location is more likely exogenous and exploiting variation in distance generated by opening and closing of language training centers.

Damm, and Schultz-Nielsen, 2020; Foged et al., 2023, for details).[10] The assigned municipality 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, suitable housing conditional on demographic characteristics such as age and household size.

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.[11] Such sorting does not threaten 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.

Upon settlement in the municipality, the refugee starts a three-year Integration Program. The program is mandatory. Refugees are eligible for welfare benefits from arrival conditional on participation in the program, and access to permanent residency and citizenship require completion of the program.[12] 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

[10] We 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.

[11] This 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.

[12] See Arendt, Dustmann, and Ku (2021) for further requirements for permanent residency.

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 Integration 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 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).[13] 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.

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

[13]This 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).

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 determines fluency in the Danish language 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.[14] 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.[15] 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 we do not use the GPA from the final exams.[16] Similarly, hours is the cumulative hours during the Integration Program (in year 0 to 3).[17]

[14] Appendix 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.

[15] Employed refugees are not required to attend as stated in section 2. Few refugees find regular employment during the Integration Program. 28 percent hold fulltime jobs in year five from being granted asylum (two years after the three-year program, see Table 5). Language training is available in the evenings and at some large workplaces to make it easier for employed individuals to attend. Furthermore, refugees who become employed are more likely to qualify for permanent residency and citizenship, which increases their incentive to continue language training and take the final exam (consistent with Arendt, Dustmann, and Ku, 2021).

[16] We 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.

[17] Appendix Table A.3 shows results from similar regressions as Table 3 in the main text with the annual hours of class in year 0 to 3 after placement instead. We also look at the effect of cumulative hours on proficiency each year after program start in Appendix Table A.4. However, the right time to measure hours and proficiency with our data is in year 3 as we will argue in section 5.

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).[18] 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.[19] This distance measure covers urban and rural locations and is only missing if no one is observed traveling between the two zones.

Refugees are usually assigned to the nearest language training facility within the municipality but not all training facilities 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 (one way) 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.[20] Overall, as more state-approved training centers open during our period (Table 1), average distance decreased from 48 minutes to 42.5 minutes during our sample period.[21] We convert travel time in minutes into hours for readability in all our regressions.

[18] We have access to the pseudo-anonymized addresses within municipalities of individuals and language training centers that we can map to zone identification numbers and we have distances between zone pairs. The total number and the spatial distribution we observe (Appendix Figure A.4) match well available information about language training centers, see e.g. https://dedanskesprogcentre.dk/english/.

[19] We 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.

[20] Some 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 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.

[21] Language centers may open or close at specific locations, for instance, because private, state-approved suppliers win or lose the contract with the municipality. Some municipalities own language training centers, some contract with private suppliers or use a combination of public and private suppliers.

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

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.[22]

Individuals with severe health problems are exempt from the Integration Program. 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 be assigned to language training. Finally, we exclude a small number of individuals aged 50 or older because we have very low compliance for this group,[23] and we exclude individuals who are initially placed in one of the four largest cities of Denmark (Copenhagen including Frederiksberg, Århus, Odense and Aalborg).[24] These cities have received far less refugees than their population size would

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

[23]See Appendix Table A.2.

[24]This sample restriction eases the interpretation of the estimated coefficients when we analyse mobility in section 6.1 but does not affect any of our conclusions.

suggest due to large pre-existing immigrant populations. Hence, it is interesting to ask whether language skills can foster local integration and counteract post-placement mobility of refugees towards the largest cities. Our main sample consists of 10,170 refugees.[25]

It is worth noting that the refugees we exclude are those most likely to be devastated by war and those who for cultural or other reasons do not want to participate in the labor market. Hence, labor force participation is relatively high in our sample and three in four either have some employment or are enrolled in an education program (or both) in year five after arrival (Table 5).

3.4 Outcomes

The outcomes we consider are fluency in the Danish language, mobility, labor market status (employment and unemployment) and education. How we measure fluency in the Danish language is described in detail in section 3.1. We now define the remaining outcomes. Our labor market outcomes are a variable capturing employment as a fraction of a full working year, a dummy equal to one if the individual had positive labor market earnings, and a dummy equal to one if the individual is unemployed (in November). Education is an indicator equal to one if the individual is enrolled in an education in Denmark.

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 ($p = 0.001$).[26]

We create two measures of integration into the local community of initial placement. The first, is a dummy equal to one if the individual leaves the assigned municipality between com-

[25] We 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.

[26] 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 training on language proficiency and socio-economic integration.

pletion 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 sample selection no refugees start out in these cities). These measures of mobility are not significantly correlated with language proficiency in the raw data. Language skills are measured at the end of the three-year Integration Program. All other outcomes are measured in year 5 (or later in Figure 4).

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 socio-economic 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:

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

I_s is investment in language learning in period s, and A is individual characteristics that are predetermined upon arrival/placement in the municipality and influence language learning such as ears for language and education acquired abroad.[27] The function f_s is strictly increasing and

[27] We ignore subscripts for individuals. We introduce a subscript t to index the year of arrival/asylum in our empirical specification that uses a panel of individuals, who were granted asylum from 2003 to 2013.

strictly concave in I_t.[28,29]

It is reasonable to assume that the return to investing in language capital is increasing in the existing language capital ($\partial^2 f_s/\partial Q_s \partial I_s > 0$, 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)[30] 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. An extensive language training program like the Danish one (up to 1800 hours over a period of up to three years) could crowd out other activities that facilitate socio-economic integration. Hence, it is relevant to ask whether more hours of language training is an effective way to increase proficiency in the host country language (high returns to additional investments) or whether everyday life in the host country is equally effective (strong self-productivity, possibly after some basic introduction to the language).

4.2 Estimating Equations

We analyze whether more hours of language classes increase the language proficiency and socio-economic integration of adult immigrants. To answer this we estimate the following equation relating the outcome in year s, $y_{s,t}$, to the cumulative hours of class at the end of the three-year

[28]Cunha 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.

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

[30]The 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).

Integration Program $Hours_{3,t}$: [31]

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

ω_t and μ_z are year and municipality or zone of placement fixed effects. $X'_{0,t}$ and $Z'_{0,t}$ contain individual and zone characteristics measured at arrival. Individual characteristics are age (second order polynomial), gender, civil status, number of children by child age groups, country of origin and initial assignment to language training track which is a rough proxy for education from abroad, abilities and ambition. Individual characteristics are included in the regressions to account for the fact that caseworkers use the demographic information (e.g. household size) when assigning refugees to suitable housing.[32] We also constructed the following zone-level characteristics and run regressions with and without them (see section 4.3): foreign born share, refugee share, employment rate, unemployment rate and share enrolled in education.

Equation 3 describes the first stage equation.

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

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 outcomes provided that the assignment of refugees to public housing (distance) is random conditional on the included controls. Section 4.3 discusses the instrument in more detail and investigates the quasi-randomness of distance, relevance and monotonicity.

[31] The recursive form of the production function in equation 1, $Q_{s+1} = h_s(A, Q_0, I_0, I_1, ..., I_s)$, 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, $Hours_3 = I_0 + I_1 + I_2 + I_3$. Appendix Table A.3 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.

[32] The language training track fixed effects account for the fact that the tracks progress at different paces. Note, that this is not strictly needed in the 2SLS and reduced form estimations as placement/distance should be orthogonal to this.

4.3 Identification

The natural experiment we have in mind is one where refugees are placed in the same 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 distances to language training. This is our most demanding specification (zone fixed effects), which exploits a small subset of refugees and neighborhoods in Denmark. We also run specifications (with municipality fixed effects), where the zone is not identical but we can control for zone characteristics that could matter for refugee integration. Our results in section 5 to section 8 are stable across these alternative specifications which supports a casual interpretation of our estimates. 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. 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 to Danish II, and 17 percent to Danish III. The mean (median) refugee lives in a zone with 1 (1) percent refugees, 5 (6) percent foreign born including refugees and 11 (11) percent students.

Several of the individual characteristics predict hours of class (column 2 and 3). Importantly, the same characteristics are 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: refugees with school age children travel 0.016-0.017 hours (0.98-1.02 minutes) less to language training. This supports the hypothesis that availability of public housing when the refugee is assigned to the municipality (see section 2) is determining

the exact location of the refugee (conditional on household size). 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 do not correlate with distance.

Being assigned to a zone with a higher share of refugees and students is associated with shorter distance to language training within municipalities (column 4). A 11 percentage point increase - corresponding to a jump from zero to the mean/median - in the share of students decreases distance by $3.9 \times 0.11 = 0.43$ hour (26 minutes) in column 4. The coefficient drops to $0.802 \times 0.11 = 0.09$ hour (5 minutes) and is statistically insignificant in column 5 with zone fixed effects.[33] The coefficient on the refugee share follows a similar pattern.[34] The significant zone characteristics in column 4 suggest that zones closer to language training also have more refugees and students. This could be problematic if it correlates with proximity to education and other opportunities affecting our outcomes.[35] Zone fixed effects eliminate all time-invariant differences across zones. We will estimate models with and without (time-varying) zone characteristics and this allow us to evaluate whether imbalances in the zone characteristics influence our findings.

Figure 2 shows the density of our instrument (left y axis) and its correlation with hours of class (right y axis).[36] 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 specification with municipality fixed effects allow us to use much more variation in the instrument (Panel a) compared to the specification with zone fixed effects (Panel b). The distribution of the instrument, distance by public transport, is more compressed and the difference between the 10th and the 90th percentile shrinks to less than half of the corresponding difference in the municipality fixed effect specification. The locally fitted regressions of hours of

[33] We fix the zone characteristics the year the refugee obtained asylum.

[34] An increase in the refugee share equal to the mean/median (1 percent) decrease distance 0.24 hours (14 minutes) in the specification with municipality fixed and 0.11 (slightly less than 7 minutes) in the zone specification.

[35] The coefficient on the refugee share in column 5 could indicate that some municipalities have responded to a large influx of refugees by opening or moving language training centers but this correlation is economically small and far from significant.

[36] We first residualize the variables and then add back the mean. Appendix Figure **??** illustrates the reduced form relationship between our instrument and language proficiency.

class on distance show downward slopes in both specification. Furthermore, it seems reasonable to assume that hours of class is monotonically decreasing in distance and the relationship is close to linear.[37] Unsurprisingly, the slope is more precisely estimated for travel times where the density of our instrument is good, while we get 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).

5 The Impact on Language Proficiency

Table 3 reports the OLS, 2SLS, first stage and reduced form estimates of the impact of hours of language training on language proficiency using distance to language training as an instrument. We report four specifications: one with municipality fixed effects (columns 1 and 2), one with zone fixed effects (columns 3 and 4), and each of them without (columns 1 and 3) and with (columns 2 and 4) zone controls. All the regressions use the cumulative number of hours and the skills measured at the end of year 3.[38] 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. The coefficient suggests that refugees who live one hour further away from training accumulate between 46 and 71 hours less, depending on the specification.

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

[38] Appendix Table A.4 shows similar regressions for each year from one to six. The coefficient on hours and the F-test statistic of the first stage decrease from year one to five suggesting that the relationship weakens over time (we have fewer observations when we go beyond five years). Refugees who do not pass the exam required for permanent residency have an incentive to take the exam at a later date and we update their language skills while we do not observe progression for refugees who already completed the entire program. For that reason we think it is sensible to measure language skills by the end of the three-year Integration Program.

Second, the 2SLS 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 less motivated refugees may therefore take longer to complete. The estimated marginal return to an 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 CEFR level or 8 to 9 percent relative to the sample mean (2SLS estimates).

6 The Impact 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 these cities. Table 4 has a similar structure to Table 3 but more outcomes organized into panels. Panel A shows mobility out of the municipality of initial placement after the Integration Program, and Panel B shows mobility to the four largest cities in Denmark. The share residing in the four largest cities in year 5 equals mobility to these cities, because our population is placed outside the four largest cities in Denmark (see data section). The correlation between hours of language training and mobility is negative and quite strong (OLS and IV). This finding is consistent with the story that refugees who attend more hours of language training, learn more Danish and are less likely to leave their municipality of initial assignment and find residence in a minority-language enclave in one of the large cities.

The 2SLS estimates seem large but not unreasonable. An increase in in hours of class by 100 decreases the mobility rate out of the assigned municipality by 14 to 24 percentage points and reduces the probability of living in one of the main cities of Denmark by 5 to 7 percentage points or more than 50 percent of the sample mean.

The lower mobility towards the largest cities may reduce the economic returns to the acquired language skills. Foged et al. (2023) 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.[39] This evidence suggest that placement outside the largest urban municipalities and the lower mobility towards them reduce the economic gains from improved language skills in our population. However, the four largest cities are comparable to the country mean in terms of the employment rate of refugees and it is the employment rate of similar immigrants rather than the rural/urban division that seems to matter for the employment success Foged, Hasager, and Peri (2022). The next section studies the impact of language skills on key indicators of economic success in the host country.

6.2 Employment and Education

Table 5 shows the impact of hours of language training on employment, unemployment and education. Panels A to C show the effect on the employment probability (extensive margin), employment measured in fulltime equivalents and on unemployment (measured in November). Panel D shows the effect on the probability that the refugee is studying in Denmark. The OLS results of Table 5 exhibit similar positive relationships to those we saw between language skills and economic outcomes in Figure 1. 100 more hours of language training are associated with roughly 1 percentage point higher employment probability and 1 percentage point lower unemployment probability. The 2SLS estimates, generally, go in the same direction but we do not detect any significant impact of hours of language training during the Integration Program on employment and unemployment in year 5.

The impact of hours on enrollment in education programs in Denmark is strongly significant

[39]See 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).

using within municipality variation in the data. We get very similar point estimates with our specification relying on within zone variation but the 2SLS estimates are insignificant at conventional levels. The reduced form estimates support the hypothesis that more hours of language training increase the chances that refugees continue in education programs in Denmark after the completion of language training (significant at the 10 percent level with zone fixed effects).[40]The influx into education programs could delay any potential labor market effect, and hence, contribute to the insignificant effects on employment and unemployment as a fraction of the population.

7 Further Robustness Checks and Placebo Tests

Figure 3 shows the robustness of our findings regarding higher enrollment in education programs and lower mobility to the largest cities to alternative instruments in our reduced form specification and compare results to similar results for native born individuals, who should not be affected by travel distance to language training.

The hollow circles and the black bullets replicate our main results from the reduced form regressions with the municipality and the zone fixed effects, respectively. The hollow and black squares show the same for an alternative instrument that is an indicator for distances above or below the median. This instrument throws out useful information and generates more noisy estimates. The hollow and the black triangles use log distance instead of distance and produce estimates that are quite similar to our main results in Tables 4 and 5. Hence, our 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 by using a simple dummy for distances above the median.

The coefficients for natives (bottom of Figure 3) are very close to zero and generally insignificant. A few of them are marginally significant due to the very large sample size for the native

[40]We also generated a dummy for completion of an education program. The coefficient of completed formal education is generally positive but never significant.

born population.[41]

8 Persistence

Figure 4 look at the impact 5 to 10 years after asylum using refugees who were granted asylum from 1999 to 2008. The coefficients in year 5 for enrollment (in Panel a) are broadly similar in magnitude to those based on refugees who were granted asylum from 2003 to 2013 in Table 5. An increase in distance by one hour reduces enrollment by roughly 2 percentage points in both samples. The impact on enrollment in educations disappears from year eight. Some might complete the education and obtain employment. We looked into this and did not find clear evidence to document such transitions into employment and education.

The coefficients are larger for mobility to the main cities when we look at the earlier asylum cohorts (Panel b in Figure 4 compared to Panel b in Table 4). The estimated coefficient on living in the main city shrinks slightly over time but remains larger in all years between 5 and 10 compared to the more recent asylum cohorts in Table 4.

9 Conclusion

This paper contributes to a growing literature on the impact of language training. Compared to the existing literature on the topic, we employ a novel identification strategy and look at new outcomes related to mobility and the degree of integration into the local community where the refugee was initially placed.

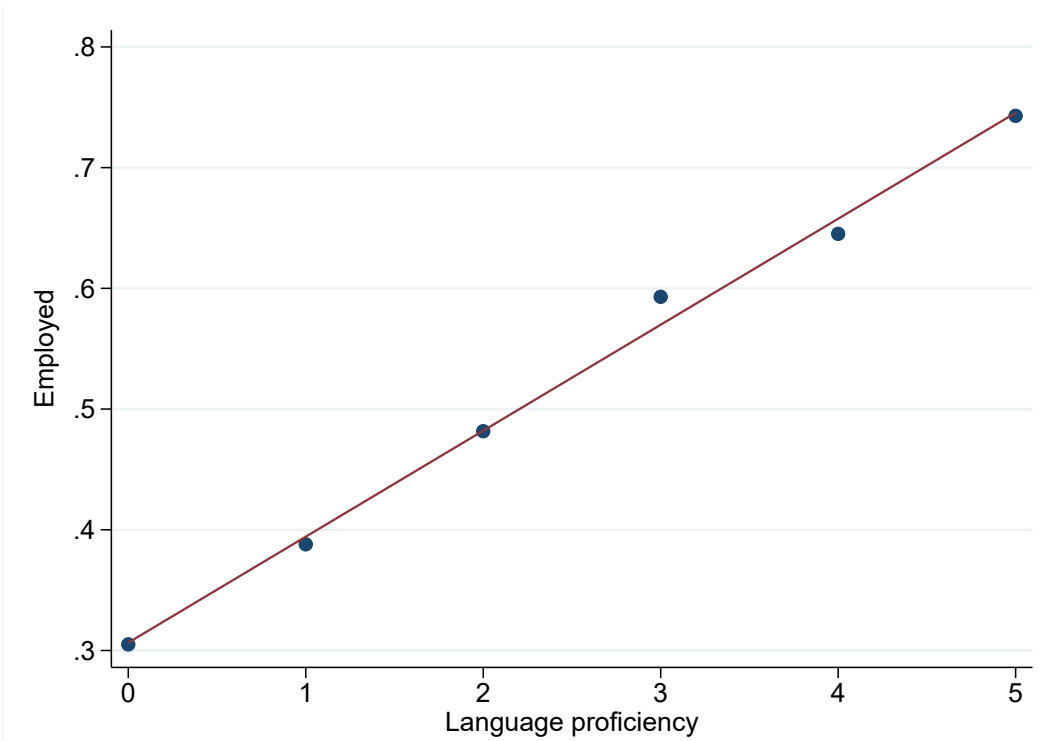
We exploit data on the location of refugees and language training centers in zones and a matrix of distances between all zone pairs to build an instrument that measures refugees' travel time by public transport to language training. RTC-type regressions show distance is unrelated to the characteristics of refugees (except for an economically small correlation with school age

[41] To 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.

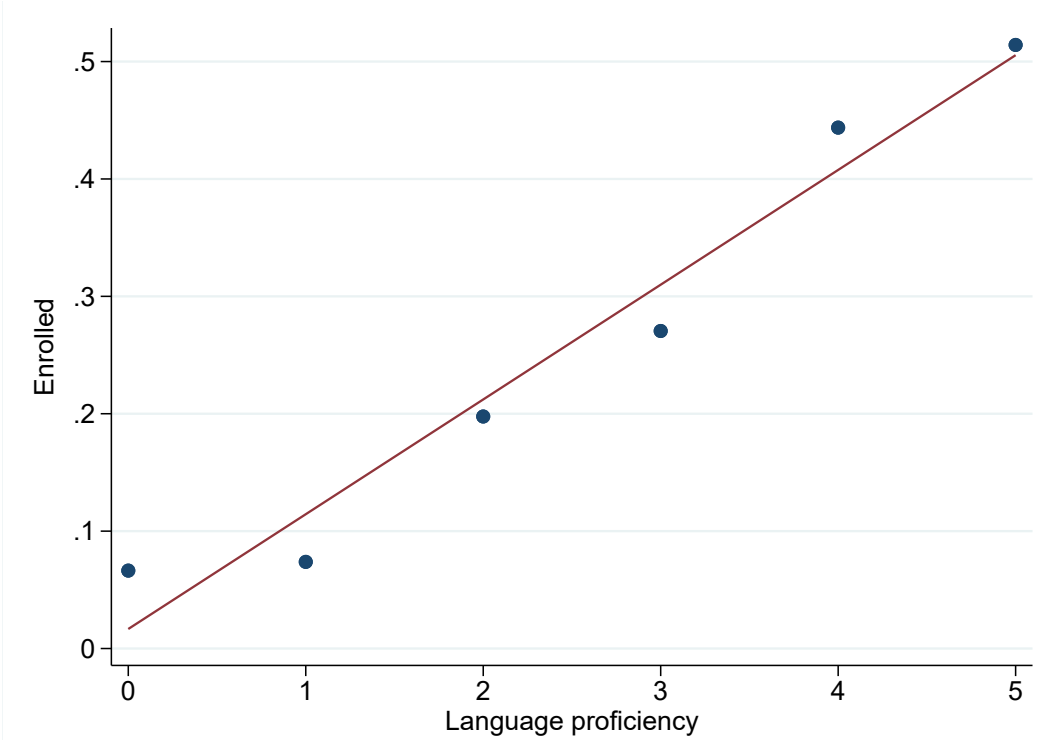
children). Within municipalities it seems to be the case that more refugees (and more students) are placed in the public housing areas closer to language training. To alleviate concerns that differences in zone characteristics that are correlated with distance to language training could impact our findings, we show estimates from four types of specifications: municipality fixed effects or zone fixed effects and each of them with and without time varying zone characteristics. Reassuringly, the point estimates are generally very stable across these specification, suggesting that our results are not confounded by differences across zones within municipalities or special features of the zones we are relying on when we exploit variation within zones over time.

Our findings suggest that longer commute to language training decreases the number of hours completed and decreases fluency in the Danish language by the end of the three-year Integration Program. This has implications for the attachment to the municipality of placement and mobility towards the main cities of Denmark with large immigrant populations. 100 additional hours of language training more than halve the mobility rate to the large cities. We find a positive effect on studying in Denmark but no (significant) effect on labor market integration.

10 Tables and Figures



(a) Share Employed by Language Proficiency Levels



(b) Share Enrolled by Language Proficiency Levels

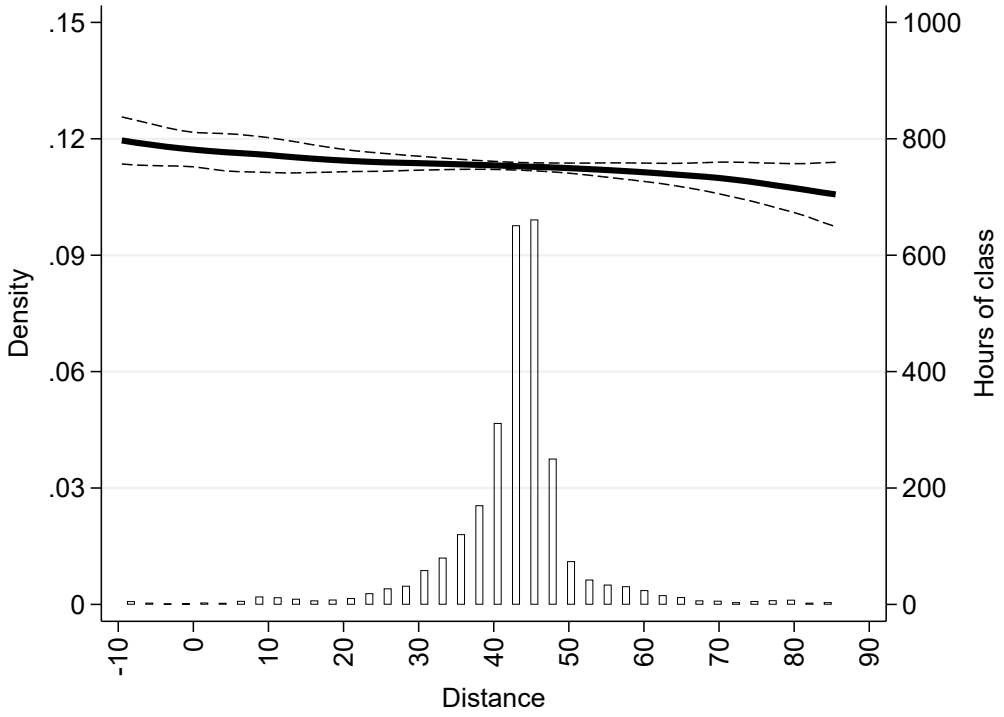
Figure 1: Economic Integration and Language Proficiency

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.

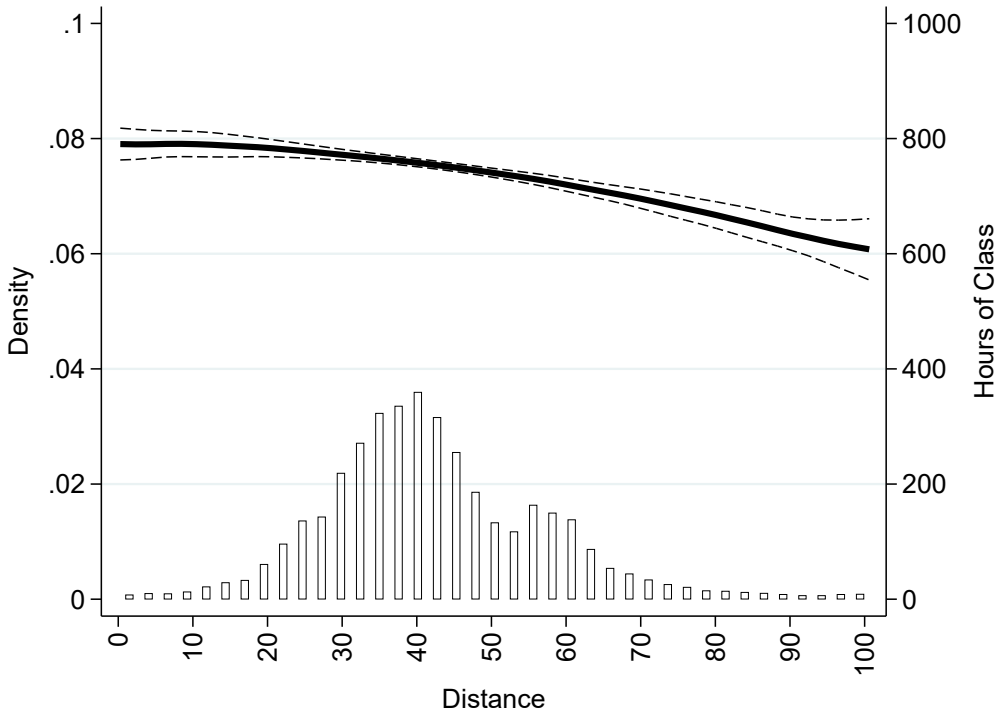
Table 2: Descriptive Statistics and Quasi-Random Assignment to Distance

	Summary statistics	Hours of class/100	Hours of class/100	Distance (hours)	Distance (hours)
	(1)	(2)	(3)	(4)	(5)
Age	30.45	0.123**	0.136***	0.009	0.005
	(7.76)	(0.050)	(0.050)	(0.006)	(0.006)
Age squared	987.49	-0.001	-0.001	-0.000	-0.000
	(505.85)	(0.001)	(0.001)	(0.000)	(0.000)
Female	0.41	-0.091	-0.128*	-0.014	-0.014
	(0.49)	(0.070)	(0.074)	(0.011)	(0.011)
Married	0.55	0.215**	0.225**	-0.004	-0.002
	(0.50)	(0.093)	(0.099)	(0.014)	(0.017)
Number of children (0-5)	0.41	-0.053	-0.086	-0.010	-0.004
	(0.71)	(0.065)	(0.068)	(0.008)	(0.005)
Number of children (6-12)	0.40	0.142**	0.150**	-0.017*	-0.016**
	(0.80)	(0.063)	(0.066)	(0.009)	(0.008)
Number of children (13-17)	0.18	0.203**	0.180*	-0.012	-0.011
	(0.55)	(0.092)	(0.091)	(0.011)	(0.011)
Danish II	0.52	-0.189*	-0.205**	-0.001	0.000
	(0.50)	(0.098)	(0.098)	(0.013)	(0.014)
Danish III	0.17	-0.663***	-0.723***	0.010	-0.006
	(0.38)	(0.138)	(0.145)	(0.023)	(0.022)
Myanmar	0.09	0.143	0.168	-0.004	-0.012
	(0.28)	(0.368)	(0.368)	(0.050)	(0.048)
Syria	0.22	-0.706***	-0.626***	-0.011	-0.013
	(0.41)	(0.156)	(0.148)	(0.037)	(0.032)
Iraq	0.08	-1.187***	-1.083***	0.039	0.025
	(0.27)	(0.225)	(0.233)	(0.030)	(0.037)
Iran	0.12	-0.707***	-0.663***	-0.026	-0.016
	(0.33)	(0.192)	(0.189)	(0.022)	(0.022)
Other	0.36	-0.368*	-0.305*	-0.040	-0.038
	(0.48)	(0.192)	(0.172)	(0.027)	(0.027)
Zone foreign born share	0.06	-12.465***	17.992	8.486***	2.596
	(0.04)	(4.532)	(13.113)	(1.451)	(2.657)
Zone refugee share	0.01	26.457*	-21.338	-23.568***	-11.315
	(0.01)	(13.386)	(39.219)	(3.491)	(8.466)
Zone employment rate	0.52	2.238	-4.190	0.171	1.499
	(0.04)	(2.961)	(7.556)	(0.817)	(1.011)
Zone unemployment rate	0.07	7.923	-15.619	-1.343	1.179
	(0.02)	(5.790)	(14.643)	(1.381)	(3.276)
Zone share in enrolled	0.11	-0.537	-12.056	-3.904***	-0.802
	(0.03)	(4.447)	(14.731)	(0.848)	(2.291)
Observations	10170	10170	10170	10170	10170
Municipality fixed effects		Yes	No	Yes	No
Zone fixed effect		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 hours of class (in hundred hours) or distance (in hours) on individual and zone characteristics as well as year (not shown) and municipality or zone fixed effects. The reference is a single man with no children, originating from Afghanistan, and assigned to Danish I.



(a) Within zones



(b) Within municipalities

Figure 2: First Stage Graph of Hours of Language Training 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 hours of class 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

	Municipality Fixed Effects		Zone Fixed Effects	
	(1)	(2)	(3)	(4)
OLS	0.168***	0.168***	0.167***	0.168***
	(0.005)	(0.005)	(0.005)	(0.005)
IV	0.201***	0.200***	0.218***	0.222***
	(0.031)	(0.032)	(0.048)	(0.046)
First Stage	-0.712***	-0.675***	-0.463***	-0.459***
	(0.143)	(0.149)	(0.104)	(0.096)
F-stat	24.93	20.43	19.97	22.84
P-value	0.00	0.00	0.00	0.00
Reduced Form	-0.143***	-0.135***	-0.101***	-0.102***
	(0.028)	(0.027)	(0.023)	(0.023)
F-stat	25.87	25.18	19.18	19.11
P-value	0	0	0.00	0.00
Observations	10170	10170	10170	10170
Outcome mean	2.54	2.54	2.54	2.54
Zone controls	No	Yes	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 Hours of Class on Integration Into the Local Community

	Municipality Fixed Effects		Zone Fixed Effects	
	(1)	(2)	(3)	(4)
Panel A: Left Municipality				
OLS	-0.013***	-0.012***	-0.012***	-0.012***
	(0.001)	(0.001)	(0.001)	(0.001)
IV	-0.152***	-0.144***	-0.238***	-0.240***
	(0.024)	(0.028)	(0.034)	(0.036)
Reduced Form	0.108***	0.096***	0.110***	0.110***
	(0.020)	(0.020)	(0.023)	(0.024)
Outcome mean	0.27	0.27	0.27	0.27
Panel B: Moved to main city				
OLS	-0.006***	-0.006***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
IV	-0.068***	-0.054***	-0.047***	-0.048***
	(0.014)	(0.014)	(0.016)	(0.018)
Reduced Form	0.049***	0.036***	0.022**	0.022**
	(0.013)	(0.011)	(0.011)	(0.011)
Observations	10170	10170	10170	10170
Outcome mean	0.10	0.10	0.10	0.10
Zone controls	No	Yes	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 2 and the first stage is based on equation 3. Hours of class is divided by 100. Distance is in hours.

Table 5: The Impact of Hours of Class on Economic Integration

	Municipality Fixed Effects		Zone Fixed Effects	
	(1)	(2)	(3)	(4)
Panel A: Employed (dummy)				
OLS	0.013***	0.013***	0.014***	0.014***
	(0.002)	(0.002)	(0.002)	(0.002)
IV	0.011	0.017	0.027	0.028
	(0.013)	(0.015)	(0.019)	(0.019)
Reduced Form	-0.008	-0.011	-0.012	-0.013
	(0.009)	(0.009)	(0.009)	(0.009)
Outcome mean	0.53	0.53	0.53	0.53
Panel B: Employed (fulltime equivalents)				
OLS	0.005***	0.005***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
IV	0.001	0.007	0.015	0.017
	(0.010)	(0.011)	(0.017)	(0.016)
Reduced Form	-0.001	-0.005	-0.007	-0.008
	(0.007)	(0.007)	(0.008)	(0.008)
Outcome mean	0.28	0.28	0.28	0.28
Panel C: Unemployed (dummy)				
OLS	-0.011***	-0.011***	-0.012***	-0.012***
	(0.002)	(0.002)	(0.002)	(0.002)
IV	-0.001	-0.004	0.012	0.010
	(0.010)	(0.011)	(0.018)	(0.018)
Reduced Form	0.001	0.003	-0.006	-0.005
	(0.008)	(0.007)	(0.009)	(0.009)
Outcome mean	0.49	0.49	0.49	0.49
Panel D: Enrolled				
OLS	0.015***	0.015***	0.015***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)
IV	0.033***	0.028**	0.034	0.033
	(0.010)	(0.011)	(0.021)	(0.021)
Reduced Form	-0.024***	-0.019***	-0.016*	-0.015*
	(0.005)	(0.005)	(0.008)	(0.008)
Observations	10170	10170	10170	10170
Outcome mean	0.26	0.26	0.26	0.26
Zone controls	No	Yes	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 2 and the first stage is based on equation 3. Hours of class is divided by 100. Distance is in hours.

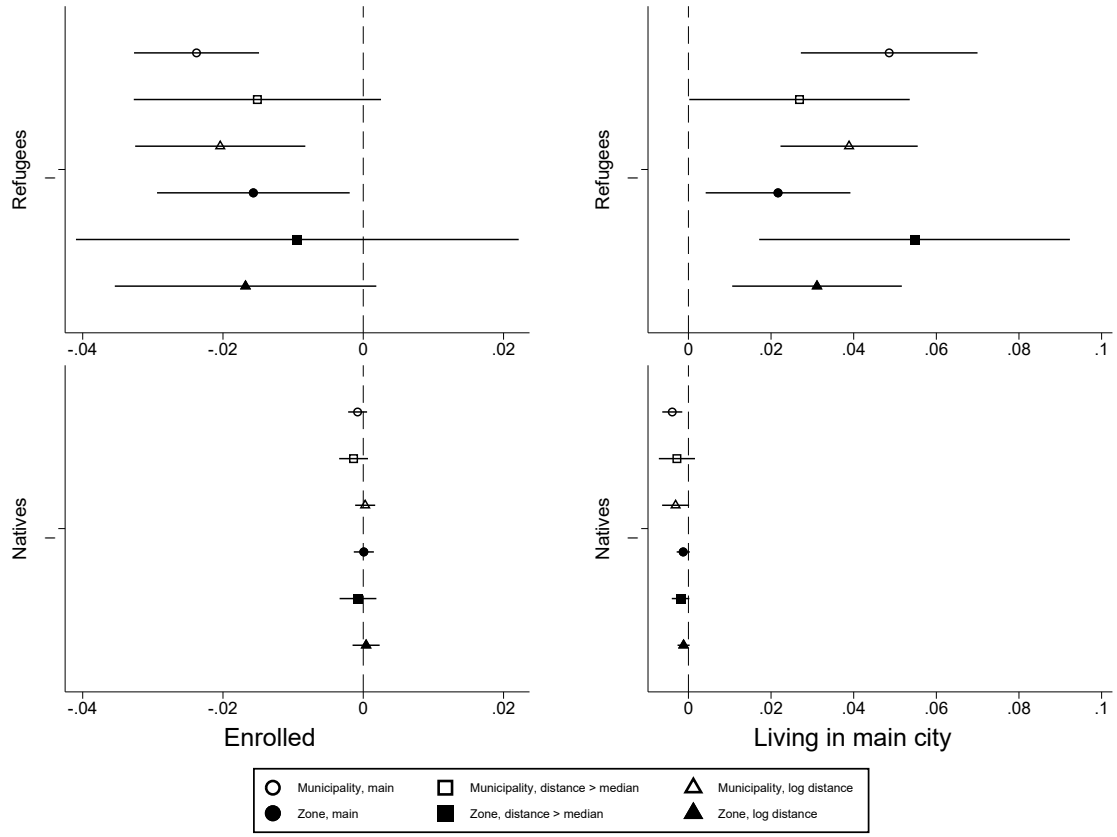
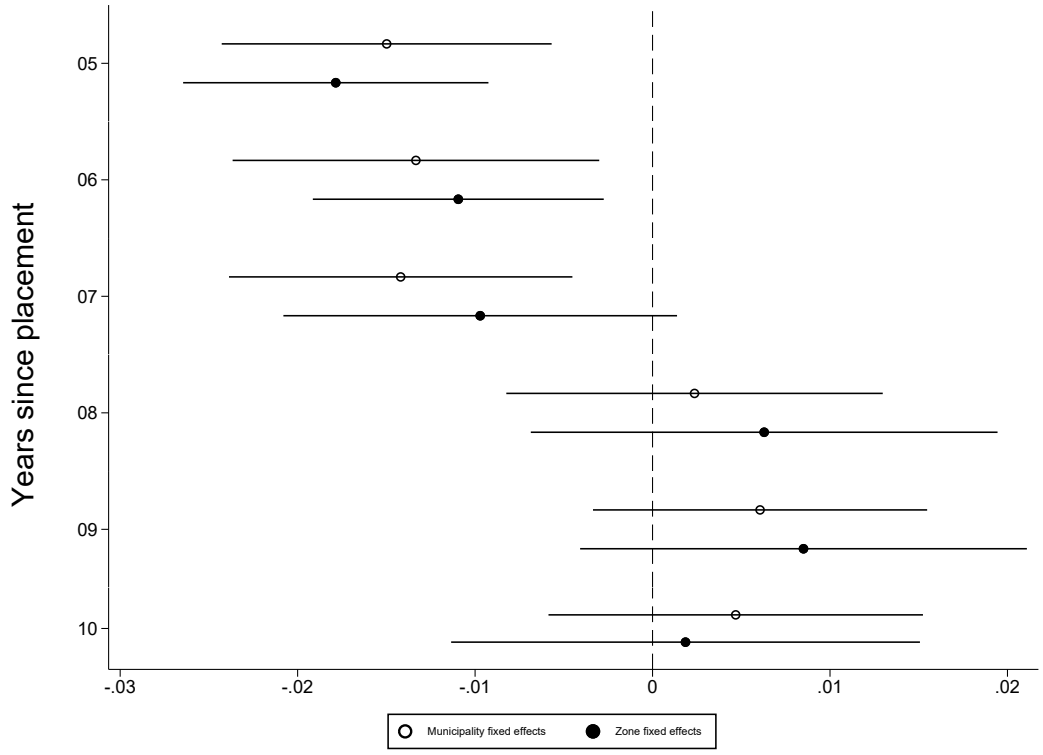
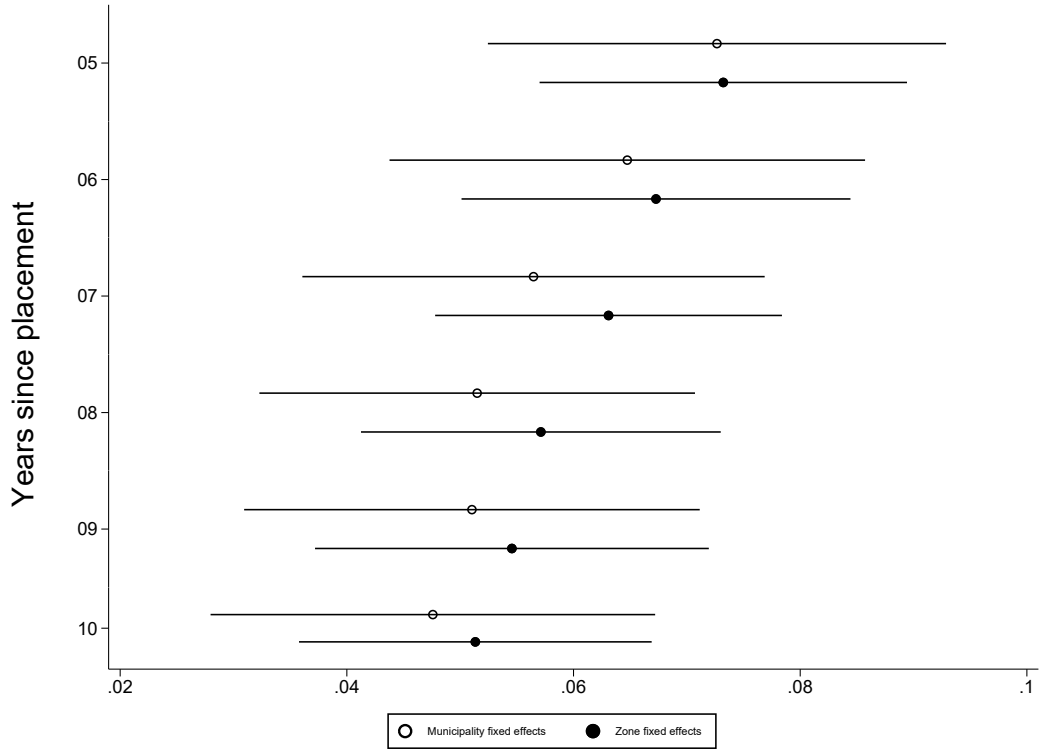


Figure 3: Placebo and Robustness Checks

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 (more arrival cohorts 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.

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, Christian Dustmann, and Hyejin Ku. 2021. "Permanent Residency and Refugee Immigrants' Skill Investment." *Rockwool Foundation Research Unit, Study Paper* 161.

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.

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 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.

Foged, Mette, Linea Hasager, and Giovanni Peri. 2022. "Comparing the Effects of Policies for the Labor Market Integration of Refugees." *NBER Working Paper* No. 30534.

Foged, Mette, Linea Hasager, Giovanni Peri, Jacob Nielsen Arendt, and Iben Bolvig. 2023. "Language Training and Refugees' Integration." *The Review of Economics and Statistics* URL `https://doi/10.1162/rest_a_01216`.

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

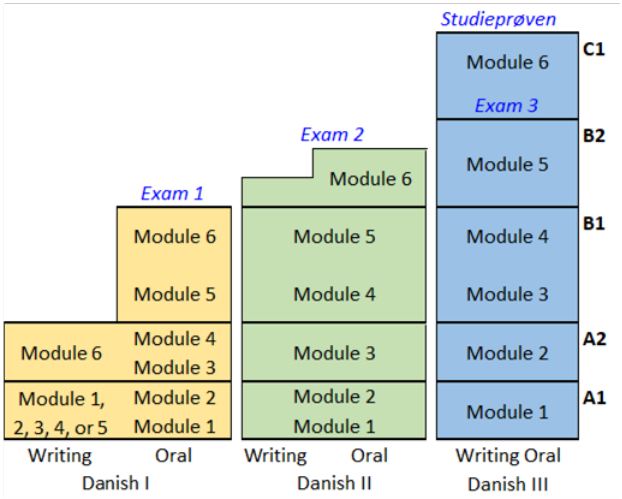


Figure A.1: Mapping Between Modules and CEFR Levels

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.

Table A.2: Compliance by Age

	Distance		
	Below Median (1)	Above Median (2)	Diff. (3)
18-29 of age	7.09	6.64	0.44***
30-49 of age	8.27	7.98	0.29**
50-64 of age (not included in analysis)	7.57	8.06	-0.49

Notes: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$ indicate results of simple test for differences in means. Table entries, in columns 1 and 2, are the mean cumulative hours of class (divided by 100) attended at the end of the three-year Integration Program (year 3 after being granted asylum). Column 3 shows the difference in means between the groups in columns 1 and 2.

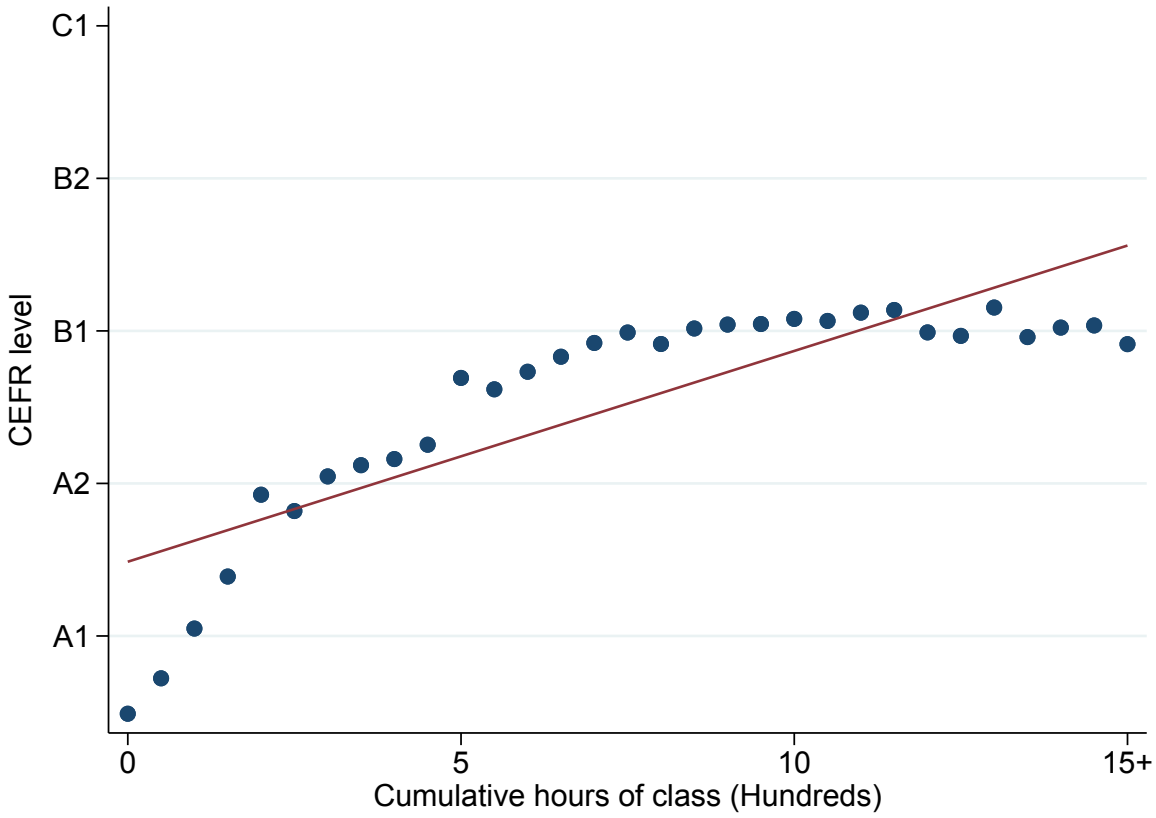
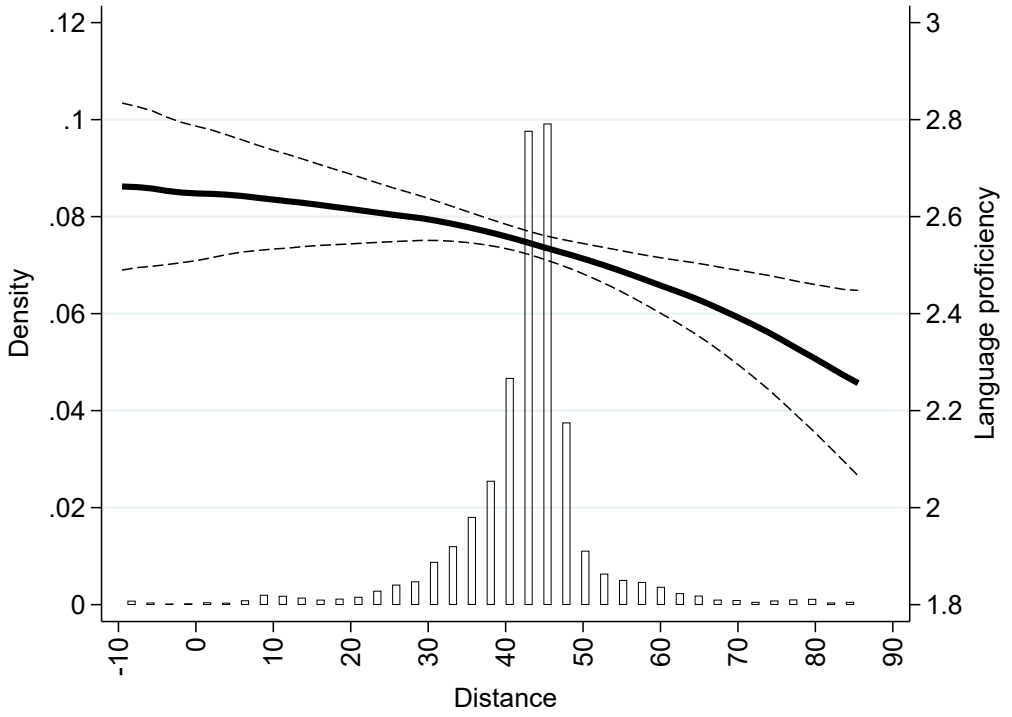
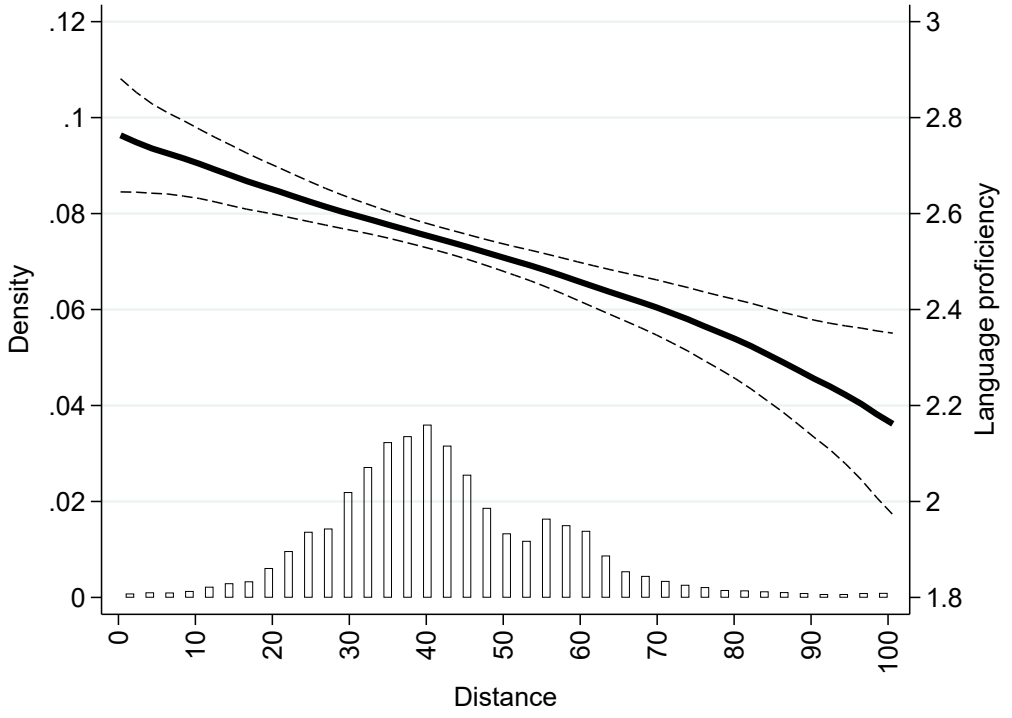


Figure A.2: Language Proficiency by hours of class

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



(a) Within zones



(b) Within municipalities

Reduced Form Graph of Language Proficiency on Distance

Notes: Reduced form 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 A.3: Hours of Class and Language Skills During the Integration Program

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS						
Hours in t=0	0.195***				0.165***	
	(0.013)				(0.011)	
Hours in t=1		0.329***			0.208***	
		(0.014)			(0.013)	
Hours in t=2			0.313***		0.212***	
			(0.013)		(0.012)	
Hours in t=3				0.094***	0.017	
				(0.014)	(0.011)	
Cumulative hours in t=3						0.167***
						(0.005)
Panel B: IV						
Hours in t=0	1.084***					
	(0.353)					
Hours in t=1		0.377***				
		(0.096)				
Hours in t=2			0.894***			
			(0.307)			
Hours in t=3				-8.840		
				(36.222)		
Cumulative hours in t=3						0.218***
						(0.048)
Panel C: First stage						
Public transportation	-0.093***	-0.268***	-0.113***	0.011		-0.463***
	(0.033)	(0.037)	(0.043)	(0.047)		(0.104)
F-stat	8.08	53.29	7.05	.06		19.97
P-value	.006	0	.009	.81		0
Observations	10170	10170	10170	10170	10170	10170
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Zone controls	No	No	No	No	No	No
Municipality fixed effects	No	No	No	No	No	No
Zone 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.4: The Impact of Hours of Class on Language Proficiency by Years Since Asylum

	s=1 (1)	s=2 (2)	s=3 (3)	s=4 (4)	s=5 (5)	s=6 (6)
OLS	0.210***	0.194***	0.168***	0.145***	0.130***	0.125***
	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
IV	0.220***	0.268***	0.222***	0.158***	0.115*	0.136*
	(0.054)	(0.054)	(0.046)	(0.049)	(0.059)	(0.072)
First Stage	-0.361***	-0.471***	-0.459***	-0.423***	-0.382***	-0.401***
	(0.049)	(0.077)	(0.096)	(0.094)	(0.095)	(0.093)
F-stat	55.00	37.11	22.84	20.26	16.19	18.54
P-value	0.00	0.00	0.00	0.00	0.00	0.00
Reduced Form	-0.079***	-0.126***	-0.102***	-0.067***	-0.044	-0.054*
	(0.023)	(0.032)	(0.023)	(0.025)	(0.027)	(0.030)
F-stat	11.89	15.19	19.11	7.40	2.72	3.23
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
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Zone controls	Yes	Yes	Yes	Yes	Yes	Yes
Municipality fixed effects	No	No	No	No	No	No
Zone fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

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

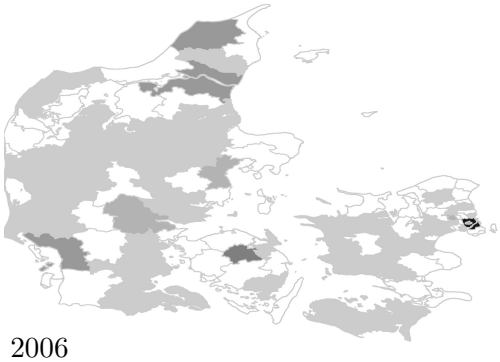
2000

2001

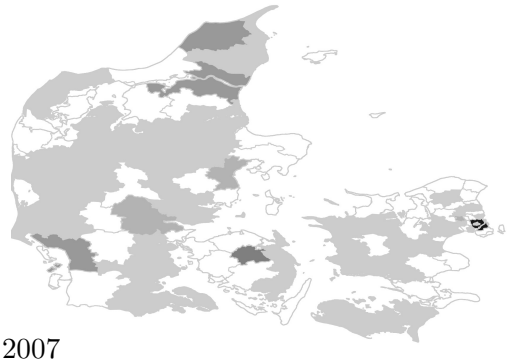
2002

2003

2004



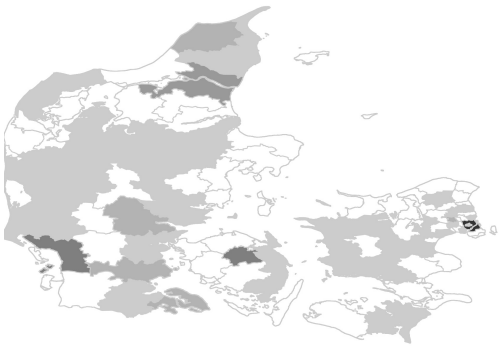
2005



2006

2007

2008



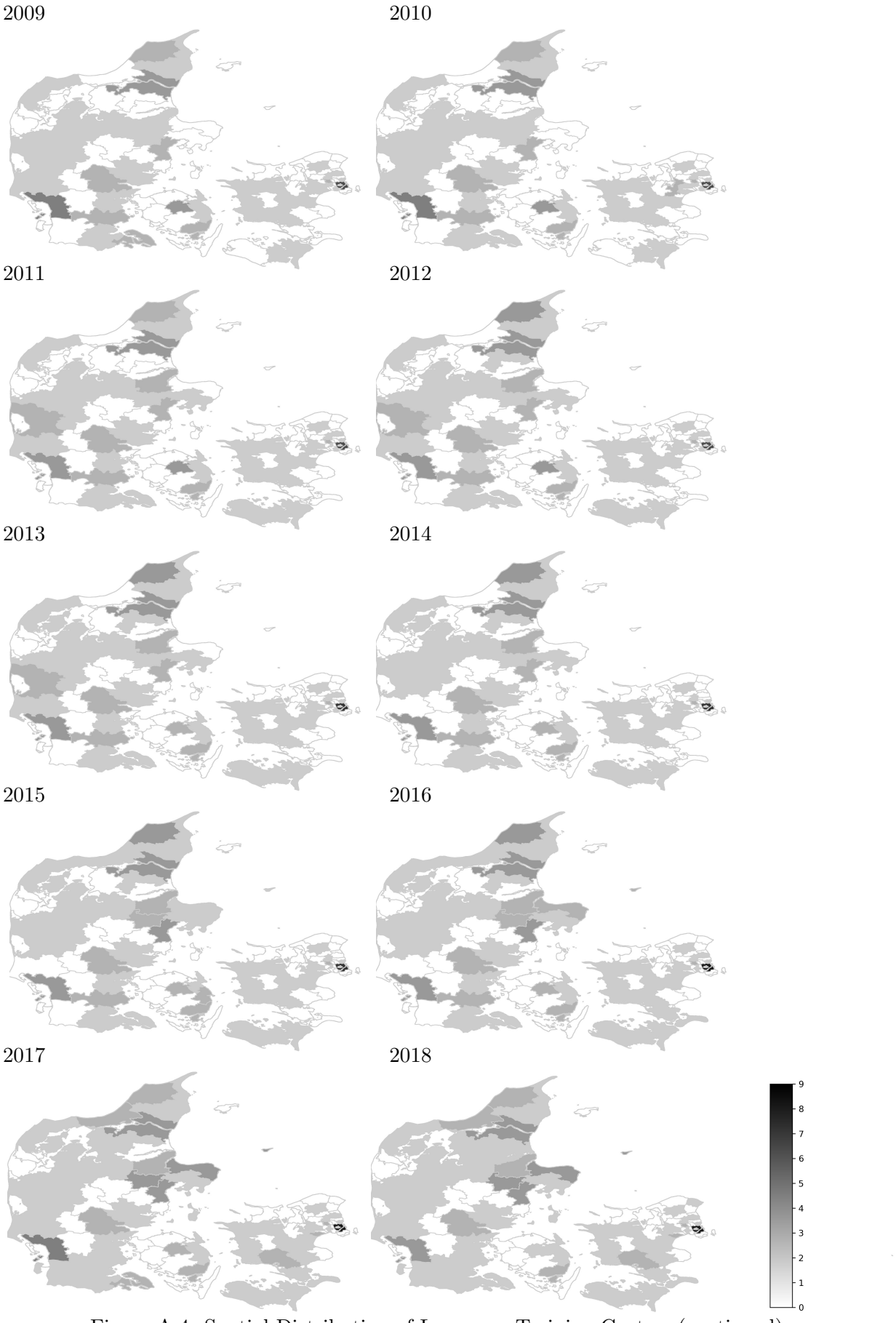


Figure A.4: Spatial Distribution of Language Training Centers (continued)

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