I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11331

The Effect of Language Training on Immigrants' Economic Integration: Empirical Evidence from France

Alexia Lochmann
Hillel Rapoport
Biagio Speciale

FEBRUARY 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11331

The Effect of Language Training on Immigrants' Economic Integration: Empirical Evidence from France

Alexia Lochmann
Paris School of Economics - Université Paris 1 Panthéon-Sorbonne

Hillel Rapoport
Paris School of Economics - Université Paris 1 Panthéon-Sorbonne, CEPII and IZA

Biagio Speciale
Paris School of Economics - Université Paris 1 Panthéon-Sorbonne

FEBRUARY 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Effect of Language Training on Immigrants' Economic Integration: Empirical Evidence from France*

We examine the impact of language training on the economic integration of immigrants in France. The assignment to this training, offered by the French Ministry of the Interior, depends mainly on a precise rule: the training is provided if the test score of an initial language exam is below a certain threshold. This eligibility rule creates a discontinuity in the relation between the test result and the variables of interest, which is used to estimate the causal effect of language training, through the method of Regression Discontinuity Design. We find that the number of assigned hours of training significantly increases labor force participation of the treated individuals. The language classes appear to have a larger effect for labor migrants and refugees relative to family migrants, for men and individuals below the median age, and for individuals with higher levels of education. Our estimated coefficients are remarkably similar when we rely on local linear regressions using the optimal bandwidth with few observations around the threshold and when we control parametrically for a polynomial of the forcing variable and use the whole estimation sample. We discuss extensively why manipulation of the entry test score is theoretically unlikely and show robustness checks that consider the possibility of misclassification. We conclude with a discussion of the candidate mechanisms for the improved labor market participation of immigrants.

JEL Classification: J15, J61, J68

Keywords: immigrants' integration, language training, Regression Discontinuity Design

Corresponding author:
Hillel Rapoport
Paris School of Economics
University Paris 1 Panthéon-Sorbonne
106-110 Boulevard de l'Hôpital
F-75013 Paris
France

E-mail: hillel.rapoport@psemail.eu

* This research project benefitted from the support of the *Département des statistiques, des études et de la documentation* (DSED) at the *Direction générale des étrangers en France* (DGEF) of the French Ministry of the Interior. We are particularly indebted to Marie-Hélène Amiel, Gérard Bouvier and Vir-ginie Jourdan. This research was also supported by the "Flash Asile" program of the Agence Nationale de la Recherche (ANR-16-FASI-0001) and the Labex OSE. We thank Toman Barsbai, Béatrice Boulu, Se-bastian Buhai, Ian Burn, Bastien Chabé-Ferret, Gabrielle Fack, Benjamin Feigenberg, Jesús Fernández-Huertas Moraga, François Fontaine, Delia Furtado, Christina Gathmann, Jeffrey Grogger, Ricardo Hausmann, Matthew Lindquist, Ole Monscheuer, Ljubica Nedelkoska, Frank Neffke, Francesc Ortega, Xavier Pauly, Paolo Pinotti, Marco Ranaldi, Dan-Olof Rooth, Melissa Siegel, and Andreas Steinmayr for their helpful comments and suggestions. We are also grateful to conference and seminar audiences at the 2017 Royal Economic Society Annual Conference in Bristol, 2017 European Society for Population Economics in Glasgow, 2017 Asian and Australasian Society of Labour Economics in Canberra, CEPII-OECD 6th Annual International Conference, University of Mannheim, RWI Essen, UNU-MERIT, SOFI at Stockholm University, and Center for International Development (CID) at Harvard University.

1 Introduction

Immigrants often represent the most vulnerable fringe of the labor force and France is no exception (Algan et al., 2010). In the mid-2000s, the French government introduced a series of policies aimed at improving the economic and social integration of immigrants and at providing immigrants the dignity guaranteed by fundamental human rights. Since the 1st of January 2007, every new legal immigrant to France who is older than 16 and is coming from a country outside the EU, has to sign a *Contrat d'accueil et d'intégration* (CAI). This contract imposes a civil training (e.g., on French institutions and the values of the French Republic), a language training, an information session on life in France and a statement of professional competence.[1]

This study evaluates the component of the CAI related to the language training. In fact, after the signature of the CAI contract, the immigrant has to pass a test on the knowledge of the French language, written and spoken. If the result is insufficient, the person is likely to receive a training. The assignment to this training depends, therefore, mainly on a precise rule: the treatment (i.e., the language training) is available when the test results are inferior or equal to a certain threshold. This eligibility rule creates a discontinuity in the relation between the test result and the variables of interest. We use this discontinuity to estimate the causal effect of the CAI language training, through the method of Regression Discontinuity (RD) design. Straightforward regression analysis could lead to bias in the estimation of the relation, since individuals who take the language classes are different in unobservable characteristics from the individuals who are not assigned the training. The proposed method is based on the hypothesis that immigrants who are just above and just below the eligibility threshold are comparable and similar in several dimensions, except for their participation in the linguistic training organized by the Ministry of the Interior.

This unique framework allows, thus, to assess one major factor that impacts the immigrants' efficient integration in the host country's labor market: the language barrier. Chiswick (1991) and Borjas (1994) were among the first scholars to recognize the language barrier as one of the major hurdles in the immigrant's integration. Since then, improving the language skills has been acknowledged to play an important role in the assimilation and integration process. Indeed, the data we use in our analysis – the ELIPA dataset – show that 45% of all individuals in our sample think that the lack of French language

[1] On the 1st of July 2016, the French Government modified some aspects of these integration policies. The new integration plan is denoted *Contrat d'intégration républicaine*, CIR.

skills is the main obstacle to immigrants' integration.

The econometric analysis of the relationship between language training and economic integration is challenging, because if the least skilled immigrants attend the language classes and some skill dimensions are unobservable, then traditional econometric tools would underestimate the true effect of the training. Our study is unique in its novel approach towards the assessment of language classes on several outcomes related to immigrants' integration. We use a local randomized experiment in the form of a Regression Discontinuity Design to correct for the selection-into-language-training due to unobservables, such as ability.

Our empirical analysis shows a higher probability of participating in the labor force due to the language classes. Looking at the heterogeneity of effects by education level, we find that the higher the level of education, the bigger the impact of the language training on labor force participation. Furthermore, there appears to be a higher effect of the training on labor force participation for labor migrants and refugees, as compared to family migrants. Women tend to benefit less from the language classes in terms of labor force participation. The same is valid for individuals above the median age. Instead our estimates do not show significant improvements in objective measures of language skills – which might be due to the basic level of the CAI French classes – nor changes in the probability of employment for the immigrants who were assigned the language training.[2] The mechanisms hint towards an information effect: before, during and after the classes, immigrants can obtain useful information on job search strategies from interactions with classmates and teachers. We also find negative behavioral effects of the language classes on "feeling at home" in France, interest in French politics and interest in the origin country's politics, which might be associated to the disappointment of the absence of impact of the integration policies on employment.

Since we have few observations around the threshold of the assignment variable (i.e., the entry test score), we compare estimates from local linear regressions using the optimal bandwidth with parametric estimates over the whole sample, controlling for a polynomial of the forcing variable. The two approaches provide results that are remarkably similar. We also discuss extensively why manipulation of the entry test score is theoretically un-

[2] A major difference between the CAI French classes and the language training of the new integration plan - CIR - is that an objective of the latter training is to achieve a higher level of French language, which can improve the immigrant's employability level. In the concluding remarks we discuss whether our findings on the effects on employment and knowledge of language skills can be generalized to other contexts.

likely. The immigrants do not necessarily know the threshold needed to succeed the entry test. Moreover there is no incentive to try to fail the entry test marginally – for instance targeting a test score close to the eligibility threshold -, as this strategy induces the risk of not failing. With regard to possible manipulation by the government, the concern might be that the test scores of individuals who would have marginally failed, were pushed upwards to pass the test, or vice versa. This hypothesis can be conceptually ruled out for several reasons. First, if the government wants to assign - or not to assign - language classes to a specific immigrant it can simply do so, without manipulating the test result. Second, the test consists of five parts, administered by different people, meaning that one individual examiner can hardly decide by herself whether a person fails or passes. Nevertheless, our data show excess bunching at the first passing grade and a missing mass at the first failing grade. This effect is, however, entirely mechanical coming from the construction of the exam itself. The overall exam grade comes from an accumulation of points in intervals of 5 or 10 points, which leads to the last passing grade as being less likely to be obtained.[3] For the sake of completeness, we nonetheless estimate several alternative specifications that take into account the possibility of misclassification by the government or the immigrants. These checks confirm that our results are not driven by non-random sorting around the threshold.

Our paper is related to several branches of literature. There are works on the relationship between language skills and labor market performance, see among others Chiswick (1991), Chiswick and Miller (1995), Dustmann and van Soest (2001), Dustmann and Fabbri (2003), and Bleakley and Chin (2004). Chiswick (1991) studies the determinants of English language fluency among immigrants in the US and the effects of self-reported language skills on earnings. He finds that reading fluency is more important than speaking fluency as a determinant of income. Chiswick and Miller (1995) provide an international comparison of the relationship between language and income, using data from Australia, the US, Canada and Israel. Dustmann and van Soest (2001) show that neglecting measurement error in self-reported assessments of language proficiency leads to a substantial downward bias of the impact of speaking fluency on earnings. For nonparametric identification of the earnings equation, they use the education level of the immigrant's father as instrument in the speaking fluency equation. Dustmann and Fabbri (2003) analyze data from the UK.

[3]See section "Validity of RD design" for detailed explanation. In addition to the last passing grade, there are other grades that are less likely to be obtained by construction of the exam. There is a low mass for these other grades as well.

They combine a matching estimator that addresses the problem of endogenous choice of language acquisition with an IV estimator that eliminates the bias due to measurement error in the self-assessed measures of language skills. The instrument they use is a dummy variable equal to one if the interview was done in English only. Their estimates show that fluency in English increases employment probabilities by about 22 percentage points. Instead, the estimates on earnings are not statistically significant, when both endogenous selection and measurement error are taken into account. Finally, Bleakley and Chin (2004) solve the endogeneity of language skills in immigrants' earnings equations, using age at arrival interacted with a dummy for non-English-speaking country as instrumental variable. We complement this branch of the literature by exploring how integration plans proposed by the government affect labor market integration, not only through their potential effect on objective measures of language skills, but through other channels as well. We also innovate using a RD design, which has the properties of a local randomized experiment.

The second branch of the literature this paper is related to, uses test scores or other eligibility rules as assignment variables of local randomized experiments in contexts of human capital investment. Matsurdaira (2008) exploits the fact that students are assigned to a summer school based on their score on year-end exams and shows that summer schools may be a less costly alternative to class-size reductions when trying to increase student achievement. Urquiola and Verhoogen (2009) study the effect of class-size on student outcomes, focusing on the relationship between schools' choices of class size and households' choices of schools.[4] Manacorda (2012) employs data on junior high school students in Uruguay and finds that grade failure increases subsequent dropout and reduces subsequent educational attainment. Chin et al. (2013) investigate the effect of bilingual education on the academic achievement of students with low English proficiency and their peers in Texas. In case of bilingual education, the students attend some classes in English and other classes in the native language (with Spanish-English programs by far the most common). The authors exploit a policy rule requiring a school district to offer bilingual education depending on a cutoff value of enrollment of limited English proficient

[4] Another work analyzing class size effects is Angrist et al. (2016).

students.[5] Our work contributes to this literature by focusing on the effects of language training for immigrants, which is a relevant research question given that poor language skills is a major obstacle for the immigrants' economic and social integration in host countries.

Finally, the article that comes closer to our study is by Sarvimäki and Hämäläinen (2016). They analyze the effect of a change in Active Labor Market Policies for unemployed immigrants in Finland, using a fuzzy RD design that relies on a cutoff date of arrival determining whether immigrants are affected or not by the reform. Sarvimäki and Hämäläinen (2016) estimate the effect of implementing tutoring of caseworkers in integration plans, which aims to improve the communication between caseworkers and immigrants and to make sure the immigrants understand their integration plan. Over a follow-up period of 10 years, they find remarkable benefits of the tutoring policy for the immigrants. Differently from that study, our work focuses on the impact of the language training itself (rather than tutoring of caseworkers in integration plans), on a broader set of outcomes related to the immigrant's integration (among others, the formation of networks and objective measures of language skills), and on the test scores of an initial language exam as assignment variable (rather than a cutoff date determining eligibility for the reform). Moreover, we empirically assess several mechanisms that might explain the relationship between languages classes and immigrants' economic integration.

The rest of the paper is structured as follows. Section 2 provides background information on immigrants' integration policies in France. Section 3 presents the data this study relies on and descriptive statistics that motivate our analysis. Section 4 explains the empirical strategy. Section 5 presents the results, as well as the mechanisms at work. Finally, section 6 concludes.

2 The French "*Contrat d'Accueil et d'Intégration*" (CAI)

As previously explained, every new legal immigrant to France who is older than 16 and is coming from a country outside the EU, has to sign a contract, which imposes a civil training (e.g., on French institutions and on the values of the French Republic), a lan-

[5] See Cappellari and Di Paolo (2015) for a study of bilingual schooling and earnings in Catalonia, which relies on a difference-in-differences framework and exploits a reform that introduced Catalan alongside Spanish as medium of instruction in Catalan schools. In addition to the articles summarized in this introduction, examples of other works that study human capital investment using a RD design are Van der Klaauw (2002) and Jacob and Lefgren (2004).

guage training, an information session on life in France and a statement of professional competence. Until recently this integration contract was the so-called *Contrat d'accueil et d'intégration* (CAI), introduced in 2007. In July 2016, the French Government launched an updated version of this program, under the name of *Contrat d'intégration republicaine* (CIR). In this paper, we evaluate the language training component of the CAI.

Following the signature of the CAI, all individuals had to take a language test which is designed by "FLE" educationalists (*Français langue étrangère* or French as a Foreign Language) . This language exam did not evaluate immigrants' actual language skills, but their skills relative to an A1.1 (a "survival") level. Individuals with language skills above this level were likely to succeed at the highest grade, even though their French language skills might have been very basic. For the individuals who were assigned this training - designed by "FLE" educationalists as well - the number of assigned hours ranged from 60 to 400, depending on how much training was needed to achieve an A1.1 level. The scope was, indeed, to bring all individuals' language skills to the same level: A1.1. The content of the training relied on three blocks: public life (institutions, values, secularism, etc.), practical life (health, education, etc.) and professional life (job search, employment contract types, etc.). Once the training was over, an evaluation was made and the individuals got a diploma of basic French skills, which was an important requirement to renew the residence permit. After the first renewal, this one-time language course allowed to extend the residence permit as often as the individuals wished. Only in case the immigrants wanted to naturalize, they had to present a B1 certificate level. In sum, the CAI system foresaw no system of progressive learning of the new language: once the level A1.1 was attained, individuals did not have to further improve their French - unless they wanted to apply for naturalization. Thus the offered training stopped at levels that were generally too low for economic integration. The individuals did not speak French well enough to be considered employable.

While our work focuses on the evaluation of the language training of the CAI, it is interesting to compare the old version of the integration policies with the new version of the contract. The CIR too imposes the language entry test to all individuals who sign the contract. In contrast to the CAI, however the language test of the CIR evaluates the actual language skills, to position individuals in courses according to their skills at levels A1, A2 or B1. The number of assigned hours varies between 50, 100 and 200 hours – according to the needs of the individual to achieve one of these three levels. Once the training is over, an evaluation of at least the basic level A1 is carried out. In order to renew the

residence permit, the immigrants must show progress in speaking and writing French. If the individuals progress, they can move from the 1-year-residence permit to the "multi-annual" permit (2 to 4 years, a new type of residence permit introduced with the CIR). As a next step, the individuals must follow a free A2 level course offered by the government. This time, to renew the "multiannual" residence permit, the individuals must successfully complete an exam that shows their A2 French skills. Finally, only if individuals want to naturalize, they need to have a B1 level. The French government offers classes for this level as well. The major difference between the CAI and the CIR's language classes concerns the system of progressive learning of French in the latter training, which follows the individuals up to the B1 level and allows them to learn French at employability levels.

3 Data

The study is based on the database *Enquête Longitudinale sur l'Intégration des Primo-Arrivants* (ELIPA), created by the *Département des statistiques, des études et de la documentation* (DSED) of the French Ministry of the Interior. These data contain detailed socio-demographic information on the interviewed immigrants. For each individual in the sample, the dataset also includes administrative information by the *Office français de l'immigration et de l'intégration* (OFII) on the test results that determine admission to the language training program.

3.1 ELIPA

The *Enquête Longitudinale sur l'Intégration des Primo-Arrivants* (ELIPA) is a longitudinal survey on the integration of first-time arrived immigrants in France. The first wave was carried out in 2010 (baseline survey), the second wave in 2011, and the final wave took place in 2013. The survey aims at following the path of the individuals and collects a large set of information regarding socio-demographic characteristics, bureaucratic itinerary, employment, language skills, living conditions and social integration. The individuals are immigrants from countries outside the EU and Switzerland who are at least 18 years old and were granted their residence permit at the end of 2009.[6] The survey considers individ-

[6]The dataset only includes information on immigrants residing in the host country with a regular residence permit. Examples of works that analyze the effects of legal status on several aspects of immigrants' integration are Kossoudji and Cobb-Clark (2002), Dustmann, Fasani and Speciale (2017), Pinotti (2017) and Devillanova, Fasani and Frattini (forthcoming).

uals who asked for a residence permit of at least one year and want to settle permanently in the country.[7]

3.2 Data on Test Results and Outcome Variables

In order to evaluate the integration of immigrants as broadly as possible, we use several different outcome variables. We measure labor market outcomes in 2013, three years after the baseline survey and the entry test determining assignment to the language classes. Our main dependent variables are labor force participation (equal to 1 if being employed or registered as unemployed, 0 otherwise)[8], employment status (equal to 1 if employed, 0 otherwise), type of contract (equal to 1 if permanent contract) and type of employment (a dummy equal to 1 if full time job; another dummy equal to 1 if informal work). Furthermore we also measure income per household individual.[9]

The ELIPA dataset provides detailed information on the language training and results of the entry test. The latter information is not self-reported, but it is provided by the OFII. The training is not assigned to all individuals. It is exclusively offered to, and compulsory for, a subset of individuals. After signing the CAI, each immigrant has to take a language test, which is composed of five parts: one oral part that amounts to a total of 70 points maximum and four written exams that, together with the oral exam, amount to a total of 100 points. We create a running variable that sums the points gained from each of the tests. The total grade is in intervals of five points and its maximum value is 100 points. 45 is the last failing grade and 50 is the first passing grade.

[7]The following categories of residence permits are excluded from the survey: students and trainees; employees on assignment; holders of residence permits with endorsements for "skills and talents", "research" or "business"; high-level executives and family members accompanying them; holders of a permit endorsed as "visitor" or "arts and culture professional"; veterans; and persons from third countries entering France for medical treatments (Direction générale des étrangers en France, Département des statistiques, des études et de la documentation, 2010).

[8]In the section where we present the results, we mainly focus on labor force participation because this is the dependent variable for which we find statistically significant effects of the language classes.

[9]Income per household individual is defined as the total income of the household divided by its "consumption units". These are calculated following the equivalence scale of the OECD, which attributes one consumption unit to the head of the household, 0.5 to other members of age 14 and older, and 0.3 to children below the age of 14.

3.3 Descriptive Statistics

Most of the new immigrants (69%) were granted a residence permit on the grounds of family migration. The other two main administrative reasons are the granting of refugee status (14%) or economic migration, which is migration for employment reasons (12%). The slight majority of the new immigrants (51%) are women. The average age at arrival is 29. All the individuals in the sample received their residence permit in 2009, however, they have not all arrived in the same year: on average, immigrants have been living in France 3.5 years before the survey. The main region of origin is Africa: 33% are from the Maghreb and 28% from sub-Saharan Africa. The reason why a residence permit was issued widely differs by origin. Family migrants more often come from the Maghreb countries (43%), economic migrants from sub-Saharan Africa (53%) and refugees from Asia (40%).

The immigrants' employment rate rose steadily: from 46% in 2010 (67% for men and 26% for women), to 54% in 2011 (74% and 35%) and 59% in 2013 (79% and 42%). In 2010 more than 50% of all immigrants who worked found their current job through relations and recommendation (family, friends, third persons), whereas in 2011 this number decreased to 39% for those people who had a new job in 2011 relative to 2010, replaced by responses to job offers in newspapers, from the *Pôle Emploi* (the French national employment agency), and spontaneous applications. In 2013, 38% of the workers found their current job through relations and recommendations. The numbers of responses to job offers from newspapers and spontaneous job applications increased further. Tables A1 and A2 in the appendix show additional descriptive statistics on socio-demographic characteristics by migration reason, as well as labor market outcomes by gender and year.

25% of the immigrants (equally distributed between men and women) were proposed the language training program. Out of a minimum of 0 and a maximum of 400 hours, the average number of prescribed hours is 264, the median number is 200. 73% of all individuals offered a language training program completed their training by 2011. Of those who had completed the training, 95% passed the DILF French language exam (*Diplôme Initial de Langue Française*), which provides them with a recognized diploma. Nonetheless, 70% of the individuals who took the CAI language classes think that this training was not enough to read and speak French – they would have wished to have additional hours.

The compliance rate – i.e., the percentage of immigrants who were assigned language training and actually attended the classes – was high: 91%. The reason for this high com-

pliance rate is that attending the training sessions of the CAI is a requirement enhancing the probability of renewal of the residence permit. For the individuals who abandoned or did not follow their language training, the most important reasons involved were being pregnant, having to take care of children or health issues. By 2013 almost 90% completed their language training, 97% of whom passed the DILF.

Grouping immigrants by nationality shows that immigrants from Asia are the most likely to be assigned language training (47%), second most likely are immigrants from Europe (33%), whereas immigrants from Maghreb countries, sub-Saharan Africa and other African countries are much less likely to be assigned language training. Australia and Oceania account for the lower range. These numbers may reflect former colonial ties and French language skills before arrival.

Immigrants with lower language skills are less likely to have a job. In 2013, 65% of immigrants with very good French skills had a job, while only 56% of immigrants with little or no French had one. In 2011, 59% of immigrants with very good French had a job, while only 43% of immigrants with little or no French had one. In 2010, the percentages were 48% versus 32%.

4 Regression discontinuity design (RDD)

As explained, each new legal immigrant to France has to take a language exam assessing her knowledge of the French language, written and spoken. If the result is insufficient, the person is likely to be assigned to a language program. The rule is the following: the treatment (i.e., the linguistic training) is likely to be offered when the test results are inferior or equal to a certain threshold. More precisely, the exam is composed of five parts. The first part is an oral examination, where the individual can obtain a score between zero and 70 points. The four additional written tests give a maximum of 30 possible points. The total score ranges from zero to 100 points and is measured in intervals of five points. 45 is the last failing grade and 50 is the first passing grade. This eligibility rule – which is not necessarily known by the immigrants – creates a discontinuity in the relation between the test result and the variables of interest. The discontinuity will be used to estimate the causal effect of the linguistic training, through the method of Regression Discontinuity Design.[10]

[10] See Lee and Lemieux (2010) for a survey of RD design in economics and Pinotti (2017) for a recent application of this methodology to assess the effect of immigrants' legal status on crime.

4.1 Fuzzy RDD

The treatment is heterogeneous in the sense that different individuals can be assigned a different number of hours of language training. The range goes from 0 until 400 hours, in steps of 10. The assignment to the language training and the number of hours do however not solely depend on the fact of passing or not the language exam. The government may take into account other personal and socio-demographic characteristics of the immigrant as well. So, for the same test score, an immigrant coming from a non-francophone country may have a higher probability of being assigned language training (and more hours of classes) than an immigrant from a francophone country (Le Quentrec-Creven, 2014). Nonetheless, the language examination is the most important variable taken into consideration. As Figure 1 shows – considering the total test score of the entry exam as horizontal axis and the number of hours of language training as vertical axis – there is a clear cutoff at the passing threshold.

Given that our running variable is not the only determinant of the treatment status, we employ the method of fuzzy RD design, using the test result to build an instrument for the number of hours of language classes the immigrant was assigned.

Our first and second stage regressions are set up the following way. Equation 1 represents the second stage of the 2SLS specification:

$$Y_i = \beta_0 + \beta_1 LC_i + \beta_2 X_i + \varepsilon_i \tag{1}$$

Y_i is the outcome of interest for individual i (e.g. labor force participation, employment status, objective language skills, etc.) and X_i is a vector of control variables. Our main explanatory variable is LC_i, which is equal to 0 if no language classes were assigned or the number of prescribed hours if the language training was assigned. We are, thus, interested in the coefficient β_1. Equation 2 is the first stage of our specification:

$$LC_i = \alpha_0 + \alpha_1 T_i + \alpha_2 X_i + \upsilon_i \tag{2}$$

T_i is the assignment variable, that is a dummy equal to 1 if the immigrant did not succeed the language test or 0 if she did succeed. More precisely, $T = 1$ if $TestResult < cutoff(50)$ and $T = 0$ if $TestResult \geq cutoff(50)$.

The set of control variables of vector X includes education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the "Ile-de-France" region (i.e.

the geographical area around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the immigrant has already spent in France, dummy variables indicating the reason to migrate (the three main options being labor migration, refugee migration, and family migration), a dummy variable equal to 1 if the individual was employed in 2010 (i.e., at the time of the language entry test), and country of origin fixed effects.

4.2 Two approaches: local linear regressions and parametric estimates using the whole sample

A peculiarity of our setting is that we have few observations around the threshold (see Figure 3). For this reason, we compare estimates from local linear regressions where we focus on a small neighbourhood around the threshold with estimates from parametric regressions using the whole estimation sample. To choose the subsample of immigrants for the first approach, we rely on the optimal bandwidth selector suggested by Calonico, Cattaneo and Titiunik (2014), which is compatible with fuzzy RD designs. We also check the robustness of our results using alternative bandwidths. The second approach we rely on instead uses the whole estimation sample. This adds statistical power and robustness to the results, since we do not have a sufficiently large sample size around the threshold. We choose a vector of polynomials to approximate the functional form of our model. Equations (1) and (2) can be rewritten as follows:

$$Y_i = \beta_0 + \beta_1 LC_i + \beta_2 Dist_i + \beta_3 Dist_i * T_i + \beta_4 X_i + \varepsilon_i \tag{3}$$

$$LC_i = \alpha_0 + \alpha_1 T_i + \alpha_2 Dist_i + \alpha_3 Dist_i * T_i + \alpha_4 X_i + \upsilon_i \tag{4}$$

where the variables are defined as in equations (1) and (2). $Dist_i$ is a row vector of polynomial terms of the test result normalized around its cutoff value, 45 (i.e., the first failing grade). These terms control for the distance from the cutoff and define the functional form of our model. $Dist_i * T_i$ allows the functional form to vary on each side of the cutoff. For our estimates we will gradually increase the polynomial degree, from the first to the second. We stop at the second order polynomial, following the suggestions of Gelman and Imbens (2016) about the flaws of controlling for higher-order polynomials. We follow the

literature in considering T as the only excluded instrument in equation (4).[11]
In all specifications we cluster standard errors by country of origin times test score of the initial language exam. In our context, this approach has the advantage of exploiting the information on the assignment variable – as suggested by Card and Lee (2008) – and generating a sufficiently large number of clusters: 45 clusters for the local linear regressions using the 30-65 bandwidth and 325 clusters for the parametric estimates using the whole sample.[12]

4.3 Validity of the RD design

Figure 1 depicts the first stage of our model. The graph plots the relationship between the test score and language training, both with raw data and with residuals from a regression of language training on the control variables at baseline.[13] Panels A and B of Figure 1 plot the test score against the number of assigned hours, and Panels C and D the test score against the probability of being assigned training. The line shows the predicted values of a local quadratic smoother with rectangular kernel. We can clearly identify a jump around the cutoff: the number of hours of language classes are, indeed, remarkably higher for individuals who failed the test (i.e., who have a test result of 45 or less) with respect to individuals who passed the test.
In Figure 2 we plot labor force participation in 2013 against the running variable (the test result in 2010), again with raw data and after baseline characteristics have been controlled for. The line shows the predicted values of a local quadratic smoother with rectangular kernel. In this figure as well there is a clear jump around the cutoff: individuals who marginally failed the test, and were thus more likely to take up the language training pro-

[11] As Angrist and Pischke (2009) suggest, the resulting just-identified IV estimator has the advantage of having good finite sample properties.

[12] In Table A9, we show that our main findings are robust to the use of alternative ways of clustering the standard errors: two-way clustering by country of origin and test score of the initial language exam (Cameron et al., 2011), one-way clustering by country of origin, one-way clustering by test score of the initial language exam, wild bootstrap (Davidson and MacKinnon, 2010) by test score of initial language exam and robust standard errors. In our setting the first and the third alternative approach suffer from the few clusters issue, which may lead to downward-biased cluster-robust variance matrix estimates and over-rejection of the null hypothesis $H_0 : LC = 0$ (Cameron et al., 2015). The second alternative approach does not cluster the standard errors on the values of the forcing variable, as Lee and Card (2008) suggest to do when the assignment variable is discrete. Estimates are less precise when using wild bootstrapped standard errors clustering on test scores only.

[13] Although there are potentially 10 points to the left of the threshold for the 10 possible grades, Figure 1 depicts 8 points, since there are no observations for the grades 10 and 20.

gram, appear to have a higher probability of participating in the labor force in 2013 than those who marginally passed the test.

Next, in the spirit of the McCrary (2008) test, we plot the density of the forcing variable, to make sure the test scores have not been manipulated around the cutoff. Given that in our context the running variable is discrete, we have checked its density by plotting histograms. Figure 3 shows the density plot over the full sample and around the threshold. Focusing on the density around the threshold, we see that there is a higher mass around the test result of 50, which is the first passing result. However, this is a low density overall: 42 individuals have a test result of 50. The figure also shows that there is a lower mass for the first failing grade, i.e. 45. Again, these histograms emphasize the importance of the two approaches towards the estimation of the impact of the language classes. Since the density around the passing threshold is low, considering the whole sample of individuals – while controlling parametrically for a polynomial of the distance from the threshold – adds statistical power and robustness to the results.

In our context, manipulation of the entry test scores by the individuals on the one hand and the provider of the language classes on the other hand, is unlikely.[14] In order for our methodology to be valid, we need to be sure that the individuals close to the cutoff did not purposely fail the language test. This case is, intuitively, to be ruled out, since those who wanted to fail would have failed the test by more than just five or ten points. There is no incentive to try to fail the test marginally – for instance targeting a test score of 40 or 45 –, as this strategy induces the risk of not failing, as well as having a lower number of assigned hours. Also, the test consists of five parts, the individual does not know how many points she gets for each part of the test and the number of points needed to get exactly 45. In addition, the individual does not necessarily know the threshold needed to succeed the entry test. On the other hand, it is equally important that the results were not manipulated by the language class provider. The risk here is that the test scores of individuals who would have marginally failed, were pushed upwards to pass the test, or vice versa. Also this hypothesis can likely be ruled out. As discussed in the previous section, the test score of the entry exam is not the only variable that matters for the assignment to language training. There is no incentive, thus, from the side of the language class

[14] While in this section we explain why manipulation of the assignment variable around the threshold is conceptually unlikely, in Section 5.1 we consider robustness checks to address this potential threat to the validity of our empirical strategy. These checks are important because Figure 3 shows excess bunching at the first passing grade and at the zero test score.

provider, to manipulate the test result. If the government wants to assign - or not to assign - language classes to a specific individual it simply does so. In addition, the test consists of five parts, administered by different people, meaning that one individual examiner can hardly decide by herself whether to fail or pass a person.

Nonetheless, we still observe a missing mass at the test score of 45. If we look a bit more in detail on the test structure, we see that this missing mass derives from the structure of the exam itself, and not from possible manipulation by any involved party. In order to explain the missing mass around the passing threshold (45 points) it is important to point out the structure of the test. The exam grades go from zero to 100, with 50 being the first passing grade. The test is composed by an oral and a written part. From the oral part the participant can get a minimum of zero and a maximum of 70 points, where 35 points from the oral part are necessary to pass the overall exam. In our data, the variable "oral exam" has three values: 0, 35 and 70. The written part is composed of four sub-exams, in ascending levels of difficulty, which reward respectively the following amount of points: 5, 10, 10, 5 and hence a total of 30 points. In our sample (and from Figure 3) we can see the missing mass between zero and 35. These are individuals who got zero in their language exam but managed to get a few points from their written exams. The next threshold is 35 – and from there individuals can get 35 points if they pass no part of the written exam, 40 points if they pass the first (easiest) part of the written exam, 50 points if the pass the first and the second part of the written exam and so on. It is to notice here, that in order to receive the grade of 45, the individual needs to get 35 points the oral exam, and 10 points in the written exam. From our sample we know that the few people who actually have the 45 grade, failed the first (easiest) part of the written exam but passed the second or third (more difficult) part of the written exam. The following conditional probabilities show that it is unlikely to pass Part 2 or Part 3 of the written exam - which both provide either 10 or 0 points - conditional on having failed Part 1 of the written exam (which is graded either 5 or 0): Pr(P2=10 | P1=5) = 90.5%, Pr(P2=10 | P1=0) = 9%, Pr(P3=10 | P1=5) = 79.7% and Pr(P3=10 | P1=0) = 8.4%. Hence the structure of the test explains the low frequency of the 45 grade. Similarly and importantly, in Figure 3 we observe a decrease in mass at the value of 55 for the same mechanical reason, and at the value of 80 – which is, again, 10 points above the grade of 70 (next oral exam grade threshold).

Finally, we follow the literature on RD design and carry out a test of balance of pre-treatment characteristics, in order to check whether there is sorting around the threshold. Table 1 shows the reduced form estimates, with 19 different dependent variables

in their 2010 values (i.e., observed at the time of the entry test). We regress each one of these variables[15] on the dummy variable equal to 1 if the test result is strictly inferior to 50, including polynomials of the forcing variable (centered around the threshold) to the second degree. Ideally, we expect no significant coefficients, which would confirm random sorting around the threshold. And indeed, for all 19 variables we find no significant coefficient, showing no evidence of a discontinuity between test scores and baseline characteristics.

5 Results

Table 2 presents our first stage results (Panel A), as well as the second stage estimates (Panel B), using labor force participation as dependent variable.[16] Labor force participation is equal to 1 for all individuals who are employed or registered as unemployed in 2013. By then, almost 90 percent of the individuals have already completed their language training. It is, thus, a point in time that is advanced enough to show first effects on economic integration, but does not allow yet to study long run effects.

Our results are presented in eight columns, where each odd-numbered column shows the estimates without control variables and country of origin fixed effects, and each even-numbered column presents the results from the regressions that include a set of control variables and country of origin fixed effects. The set of control variables is the one presented in Section 4.1.

The first four columns in the table follow our first approach: local linear regressions around the cutoff. For this approach, in line with the literature on RD design, we determine the optimal bandwidth around the cutoff. For this purpose, we use the optimal bandwidth selector suggested by Calonico, Cattaneo and Titiunik (2014), which is compatible with fuzzy RD designs. The preferred bandwidth includes individuals with test

[15] Education level, age, time already spent in France, a dummy equal to 1 if the individual is resident in the "Ile-de-France" region, a dummy equal to 1 if the individual is male, a dummy equal to 1 if the person is married, number of children in the household, number of individuals in the household, a dummy equal to 1 if the individual is employed at the time of the entry test, a dummy equal to 1 if the individual is taking other French classes, three dummies for the different migration reasons, and regions of origin. All these variables are measured in 2010.

[16] In this section, the main focus is on labor force participation as dependent variable, because for this outcome of interest we find statistically significant effects of the language classes. In Table 6, we show results from regressions using other labor market outcomes as dependent variables.

results between 35 and 60 points. For the sake of robustness, we also report the estimation results of a specification using a larger bandwidth, considering individuals with test results between 30 and 65 points.

The second set of results, columns five to eight, follows our second approach, the parametric estimates over the whole sample. We increase the polynomial of the forcing variable centered around the threshold (i.e., the distance of the test score from the threshold) from the first to the second degree and include as well interactions with our instrumental variable equal to 1 if the test result is lower than 50.

Looking at the first stage in Table 2, in all specifications we find a highly significant impact of failing the entry test on the number of assigned hours of French language training. The second stage shows the effect of additional 100 hours of training on the probability of participating in the labor force. We observe a significant positive impact, which is robust throughout the specifications. An increase by 100 hours of training raises the probability of participating in the labor force between 14.5 and 26.6 percentage points. This is a sizeable effect, considering the average probability of labor force participation in 2013 being 77%.

Table 3 shows heterogeneity of effects by education level, which is defined as the number of years of education. The range goes from 0 to 17, with a median value of 12, which is equal to a high school degree. The most frequent levels of education are 0 years at school – i.e., no schooling – (26% of the individuals in the sample), 12 years at school (19% of the individuals) and 17 years at school (17% of the individuals). The table shows that the higher the level of education, the bigger the impact of the language training on labor force participation. Illiterate immigrants have little benefit from the language classes, which may suggest the importance of combining language classes and literacy policies for individuals with no schooling. For all our specifications, the coefficients a_0 and a_1 are jointly significant at the median value of education level (12 years).

Table 4 analyzes the impact of the language classes on labor force participation by migration category.[17] Panel A considers labor migration, relative to other types of migration (family, refugee, other). Looking at the coefficients a_0 and a_1 (hours of French classes and the interaction term), as well as the joint p-value in columns 3 and 4, there seems to be a higher impact of language classes for labor migrants relative to the other categories. The effect is less precisely estimated in the local linear regressions. A similar conclusion can be

[17] In the specifications of Table 4, we compare each type of migration to all other types combined together, because we prefer to estimate a more parsimonius specification with fewer endogenous regressors.

drawn looking at Panel B, where we analyze refugee migration. Although the coefficients are significant, positive and more precisely estimated, overall the effects for refugees appear to be smaller in magnitude compared to the effects that we find for labor migrants. The coefficients in columns 1, 3 and 4 of Panel C, on the other hand, show that family migrants are likely to benefit relatively less from the language training in terms of labor force participation, compared to labor migrants and refugees.

In Table 5 we consider labor force participation by gender and age category. Women appear to benefit less from the language training than men. This result is in line with the lower effectiveness of the language classes for family migrants, since women represent the majority of this group. It is also in line with the conclusions of Gathmann (2015), who looks at the impact of waiting time for naturalization on several economic and social integration indicators. She finds that women are more likely to participate in the labor force, but do not invest in education and language skills. In the context of the CAI language training, a large part of the few non-compliers are women. The main reasons for non-compliance among women include being pregnant and having to take care of children at home.

Furthermore, we find that the language training appears more efficient for people below the median age and below the age of 40. In fact, learning a new language might be easier for individuals who arrive in the country at a young age, in line with the findings of Chiswick and Miller (1995).

Table 6 presents regressions using other labor market outcomes as dependent variables. The results for employment, permanent contract, informal work and income per household member are ambiguous. The estimates hint towards a negative impact on the probability of having a full-time job. Although these findings show little robustness throughout the specifications in terms of significance, the negative effect on full-time job may be explained considering the language training as an investment into future employability. Due to the language classes, the individuals might not be able to work full-time, an effect that could last for few years after the classes.

To summarize the main results, we find a strong labor force participation effect without an increase in the probability of employment. However, it is important to stress that our data do not allow to assess the long run effects of the language classes on employment.

5.1 Selection into or out of training: Robustness checks for possible manipulation of the assignment variable

In Section 4.3, we discuss that in principle neither the individuals themselves, nor the language class provider (i.e., the government), have an incentive or the power to manipulate the test result around the cutoff. Nonetheless, Figures 2 and 3 show potential issues of selection into or out of training. In Figure 3 we observe a jump in the density of the assignment variable around the cutoff. Notably, there are more individuals who received the first passing grade (50) than individuals who received the first failing grade (45). This may hint towards possible manipulation of the test results by the government, by increasing the number of immigrants who obtained the first passing grade.[18] The same figure also shows a high density at the "zero" test result and Figure 2 a high value of residuals in the labor force participation equation for observations with a test score equal to "zero". This may indicate a situation of manipulation in which highly motivated individuals hand in a blank exam copy and do not respond to the oral questions, in order to make sure they are assigned language training. To test for the robustness of our results to possible manipulation of the assignment variable, we run two additional sets of regressions.

First, to build an instrumental variable that takes into account possible manipulation by the language class provider, we reclassify certain individuals' test score from 50 to 45. We choose the individuals who marginally passed the test, but are most prone to potentially have been pushed by the government from 45 to 50 points. For this purpose we choose four characteristics that may induce such a positive selection out of training: education level, permanence in France (i.e., how long the immigrant has already been in France), distance of French to the native language (see this variable construction in Section 5.2) and the fact of being employed at the time of the test. Each of these characteristics may affect the immigrants' integration potential. For each variable, we calculate the 50th percentile for the individuals from our estimation sample who received 50 points in their test. Then we reclassify from test score 50 to test score 45 the immigrants with values of education or permanence in France above the 50th percentile, or with values of linguistic distance below the median value. With regard to employment at the time of the test – which is a binary variable – we reclassify from test score 50 to 45 all individuals who were employed

[18] In Section 4.3, we explain that the missing mass for the last passing grade is mainly due to the structure of the exam itself. For the same reason, there is a low density for other grades too, for instance 55 and 80 (see Figure 3).

in 2010 among those who had a test score equal to 50.[19]
Denoting with "*" the "true" unobserved variables and taking education level as an example, we assume that the language class provider manipulated the test score in line with the following misclassification rule:

$$TestResult_i = 50 \quad \text{if} \quad TestResult_i^* = 45 \quad \text{and} \quad EducationLevel_i > 9 \tag{5}$$

$$TestResult_i = TestResult_i^* \quad \text{otherwise} \tag{6}$$

where 9 is the median level of education for individuals with an *observed* test score equal to 50. To build the new instrumental variable, we create a binary variable r equal to 1 if the individual is reclassified and 0 otherwise. We then generate an IV that takes into account the probability p of being reclassified: $T^* - pr$, where $T^* = 1$ if $TestResult^* < cutoff(50)$ and $T^* = 0$ if $TestResult^* \geq cutoff(50)$.[20]
Using this modified instrumental variable in a regression with the *observed* outcome variable as dependent variable, we replicate our main estimates, showing the effect of language classes on labor force participation (i.e., considering specifications similar to those in Table 2). Panel A in Table 7 shows the results following the reclassification according to the education level. The reasoning behind this reclassification comes from a higher language learning potential for higher levels of education. This might induce the language class provider to raise the test result from 45 to 50 for individuals with higher education levels, since they may be considered relatively less in need of language classes. In Panel B, we reclassify individuals according to their permanence in France. The more time immigrants have already spent in France at the time of the entry test, the easier it is for them to learn the new language. For Panel C, we choose to reclassify from 50 to 45 the individuals whose native language has lower distance relative to French. These immigrants might need relatively less language training compared to those whose native language is more distant from French. Finally, we look at employment at the time of the entry test, in Panel D. The individuals who were employed in 2010 might be considered less vulnerable in terms of integration and also less in need of language classes.
Comparing the results with each other and with the estimates in Table 2, the coefficients

[19] In this case, we reconstruct our instrument considering that these individuals form 59% of all individuals in our sample with a test result equal to 50.

[20] See Manacorda (2012) for a similar exercise in the context of grade repetition in school using data from Uruguay. In our setting, we exploit the information from government's documents to choose the criteria that may determine selection out of language classes.

remain similar. We can conclude that our estimates are robust to this test. Our findings do not depend on positive selection out of language classes due to manipulation of the test score by the government. In Tables A3 to A7 of the Appendix, we show that findings from this robustness test are similar when we consider estimates on other labor market outcomes.

As a second robustness test, we check whether the high density of "zero" test scores in Figure 3 and the high value of the residuals in the labor force participation equation for immigrants with a "zero" test score (see Figure 2) may signal the individuals' manipulation of the assignment variable. Highly motivated immigrants may want to participate in free language classes even though their level of French is high enough to achieve the passing grade in the test. We replicate the estimates of specifications similar to those in Table 2, excluding from the estimation sample all the individuals with a "zero" test score. This exercise implicitly considers a worst-case scenario where all immigrants with a "zero" test score cheated their entry test to be assigned to the language classes. Table 8 shows that our results are also robust to this check. In Table A8 of the Appendix, we repeat this test on other labor market outcomes. These additional checks confirm the robustness of our findings.

5.2 Mechanisms

An important question is to assess the mechanisms through which the language classes impact on economic integration. We point out five possible channels, which we present in Tables 9 and 10.

The first and most direct mechanism is an objective improvement in the language skills. To check for this channel, we use the results of a written and oral language exam that all individuals in the sample had to take in 2013, as well self-reported language skills. The estimates are presented in Panel A of Table 9. There seems to be no robust significant impact of the language classes on the actual language skills of immigrants. There is a hint towards a negative subjective perception of one's writing skills, which might be related to a negative behavioral effect associated to the disappointment of the absence of impact of the integration policies on employment. This result - whose interpretation seems to be confirmed by other findings below – needs however to be considered with caution. By 2011 over 70% of the language class participants had already completed their training.

Measuring language skills in 2013 means quantifying the skills two years after completion of the classes for most individuals. During this time, language skills might have decreased again if no further training was taken. Nonetheless, descriptive statistics show that over 70% of the individuals who were assigned language training do not think the training was enough to learn French, and about the same percentage would have liked to have more hours of classes. Column 3 in Table 9 reports estimates of the heterogeneity of effects, which allow to analyze the role of the distance of the native language to French. To compute measures of language distance, we follow previous literature (for instance, Fearon, 2003; Desmet, Ortuño and Wacziarg, 2012; Alesina, Harnoss and Rapoport, 2016; Laitin and Ramachandran, 2016) and use Ethnologue's linguistic tree diagrams to determine the number of common nodes between the native language of the immigrants – there are about 100 different native languages in our estimation sample – and French. In particular, we compute language distance d as follows:

$$d_{French\ Other\ Language} = 1 - \left[\frac{\#\ of\ common\ nodes\ between\ French\ and\ Other\ Language}{\frac{1}{2}(\#\ of\ nodes\ for\ French + \#\ of\ nodes\ for\ Other\ Language)}\right]^{\lambda} \quad (7)$$

where in line with previous literature we consider λ equal to 0.5. For instance, French is classified as: Indo-European, Italic, Romance, Italo-Western, Western, Gallo-Iberian, Gallo-Romance, Gallo-Rhaetian, Oïl, French. Spanish is instead classified as Indo-European, Italic, Romance, Italo-Western, Western, Gallo-Iberian, Ibero-Romance, West Iberian, Castilian. From equation (7), the two classifications imply that the language distance between French and Spanish is 0.205.

For the objective oral language skills we observe an improvement through language classes for individuals whose native language is not too distant from French (the coefficient is statistically significant up to the 42nd percentile of the language distance distribution), while the effect is not statistically significant when computed at the median value of the language distance variable (0.62). The estimates also show a negative effect of the language classes on subjective writing skills: this coefficient is significant at the median value of language distance.

A second possible channel works through a signaling effect. Potential employers could see a language diploma as a signal for language proficiency and higher ability of the immigrant. Indeed, after completion of the number of assigned hours, the participants take the DILF exam, which provides them with a diploma recognizing the acquisition of basic

French skills. By 2013, 97% of the participants passed this test, which hints at actual efficiency of the language training. If the signaling effect were an important driver of our main findings, we would expect to observe a significant increase in the employment rate. However, as we have already discussed, Table 6 shows that this is not the case. This finding presents evidence against a signaling effect at work.

The third channel we look at is based on network effects. There is literature showing the role of networks (especially high quality networks) in the integration process (e.g. McKenzie and Rapoport, 2010). The language training could give access to new networks of French friends, international friends or friends from the same country of origin. For instance, this may happen if the individual improves her language skills and becomes more sociable. Also, during the period of the language classes, the immigrant might interact more with her classmates, while having less time to socialize with other people. Panel B in Table 9 shows the estimates for the probability of having new friends in 2011 and 2013. We find some evidence of a negative impact on making new French friends in 2011 and no impact on other potential networks.

As a fourth channel, we consider behavioral effects. In principle, we would expect positive effects if immigrants are encouraged to participate in the labor force because the government implements an integration plan. Instead, we find suggestive evidence of negative behavioral effects, probably because immigrants are demotivated after they realize the integration policies are not effective in increasing the probability of employment. Indicators for this channel are the lack of an impact on self-reported language skills (and the negative effect on subjective perception of writing skills), as well as a negative effect of the language classes on "feeling at home" in France, interest in French politics and interest in the origin country's politics (see Panel B of Table 10).

Finally, we find evidence of a positive effect of the training through our fifth channel, the channel of information. The participants to the training are likely to benefit from positive interactions that the language classes produce. Before, during and after the classes the participants can use their time to exchange about their experiences in France and to give each other important advice on the French life and labor market. This exchange involves not only the participants, but also the teacher. For immigrants, it is simple and time-saving to talk to the teacher and other participants in class, as compared to trying to find the information on their own elsewhere. Furthermore, the language classes are applied to a specific content which consists of three blocks: public life, practical life and professional life. Hence, the training itself is an important source of information about life in France.

This is how attendance to the classes can provide a valuable source of information on life in France and job search strategies.[21] In Panel A of Table 10, outcome variables related to this channel – such as searching and / or having found a job through *Pôle Emploi* (the French national employment agency), having passed the driver's license in France and having asked for the recognition of the latest academic diploma – all show positive significant coefficients. It is possible to rule out the possibility that the information channel works through increased access to welfare benefits.[22] There are two potential welfare benefit allocations. The first is the so-called RSA (*"Revenu de solidarité active")* which is a specific type of minimum income allocated to individuals with no income - for immigrants it is limited to those residing in France for at least 5 years or for individuals with refugee status. The second type are the "*allocations chômage*" (unemployment benefits), for which there is an eligibility rule of at least 4 months of work in France. However, none of these allocations are linked to participation to the CAI language training or to registration to *Pôle Emploi*. This suggests that the mechanism at work is likely linked to information related to job search strategies rather than to access to welfare benefits.

6 Concluding remarks

We examine the impact of language training on the economic integration of immigrants in France. The language classes are offered by the French Ministry of the Interior after the immigrant signs the *Contrat d'accueil et d'intégration*. The training is more likely to be available when the test score of an initial language exam is below a certain threshold.
Our study is novel and important in at least two directions. First, we use a local randomized experiment in the form of a Regression Discontinuity Design to estimate the effects of language training on immigrants' integration. Second, we look at the mechanisms through which language classes impact on a set of outcome variables related to economic integration.
We find that the number of hours of training significantly increases labor force participation. The effect appears to be smaller for women, individuals above the median age, and family migrants relative to labor migrants and refugees. The language training shows a

[21] The *Contrat d'accueil et d'intégration* also includes an information session on life in France, which lasts 6 hours only. The advantage of the interactions during the language classes is that the average time length is higher, and therefore the immigrant has more time to think and ask about information that is useful for her integration.

[22] See "Service-Public.fr" for detailed information on the French social welfare system.

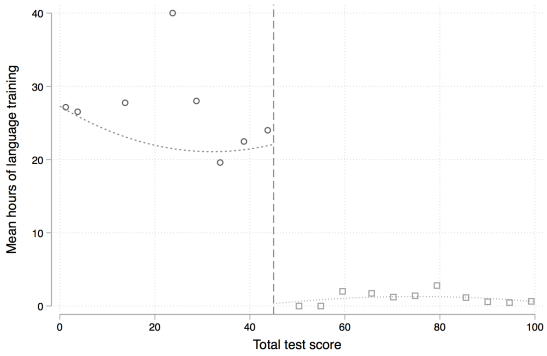
significantly higher impact on labor force participation for individuals with higher levels of education. Furthermore, the language classes do not appear to increase the probability of employment. The latter finding however needs to be interpreted with caution because our data do not allow to analyze long run effects.

We explore different mechanisms that explain the effect of language classes on labor force participation. We find little evidence for an objective improvement of language skills – which might be due to the basic level of the CAI French classes –, for a signaling effect of the diploma received after the end of the language classes, for an increase in the size of the network and for an information effect related to access to welfare benefits. Instead, we find negative behavioral effects that may derive from disappointment from the integration plan. Finally, the information on job search strategies that individuals derive from the time spent with their classmates and teachers during the classes appear to be an important channel that helps immigrants to integrate in the French labor market.

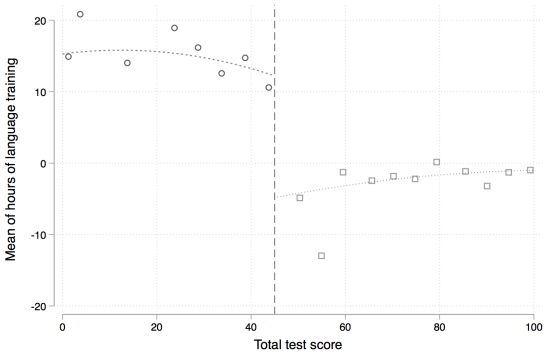
We believe that the insights we gain on the information mechanism amplify our understanding of language classes as a means of facilitating integration. We can speculate that this knowledge of the information channel is applicable beyond France, as long as the classes imply active interaction between immigrants and teachers for a sufficiently long time length. Instead, generalization of our findings on the lack of effects of the training on objective language skills and probability of employment should be taken with caution, given the basic level of the CAI French classes. For Finland, the reform analyzed by Sarvimäki and Hämäläinen (2016) allocated more resources toward language courses, in addition to improve communication between caseworkers and immigrants, and this was beneficial for the immigrants' earnings, even if it did not affect their employment prospects. In France, the language classes of the *Contrat d'Integration Republicaine* - the new integration plan introduced by the French government on the 1st of July 2016 - provide more advanced training than the CAI language classes. To analyze whether this change in the integration policy is beneficial to immigrants' integration, there is a need to collect new data on labor market outcomes for several years after the policy change.

Figure 1: First stage graph
(1) Number of assigned training hours

(A) Raw data

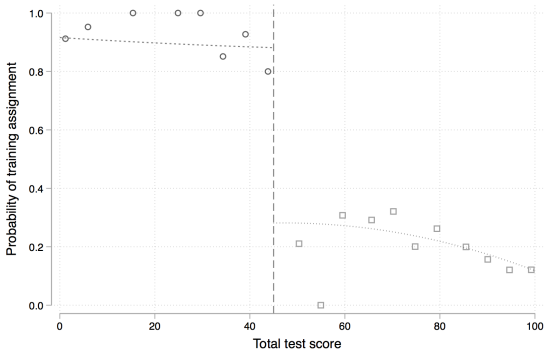


(B) Residuals from a regression on covariates

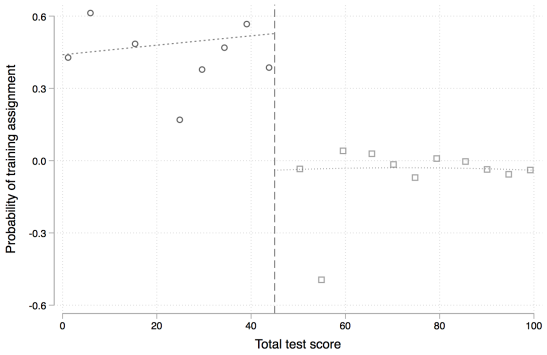


(2) Probability of being assigned training

(C) Raw data



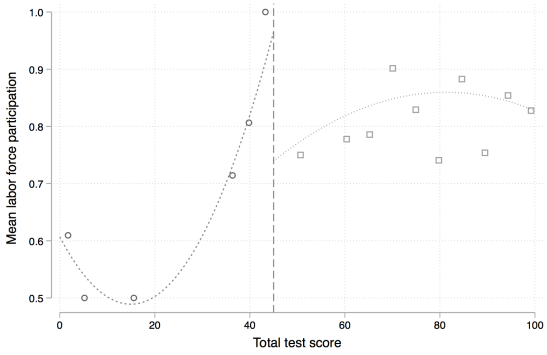
(D) Residuals from a regression on covariates



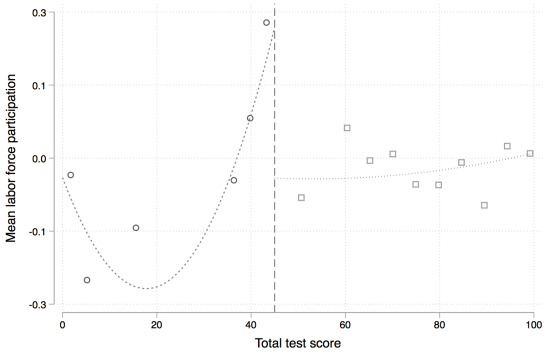
First stage plot. The left parts of this figure (panels A and C) show the relation on raw data, the right parts of the figure (panels B and D) plot residuals from a regression of "Number of hours of language training" or a dummy equal to 1 if the immigrant is assigned training, on a set of covariates (education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the "Ile-de-France" region (i.e. the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the immigrant has already spent in France, dummy variables indicating the reason to migrate, a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects). The variable "Total test score" refers to the language entry exam and is observed in interval of 5 points. The line shows the predicted values of a local quadratic smoother with rectangular kernel.

Figure 2: Reduced form graph

(a) Raw data



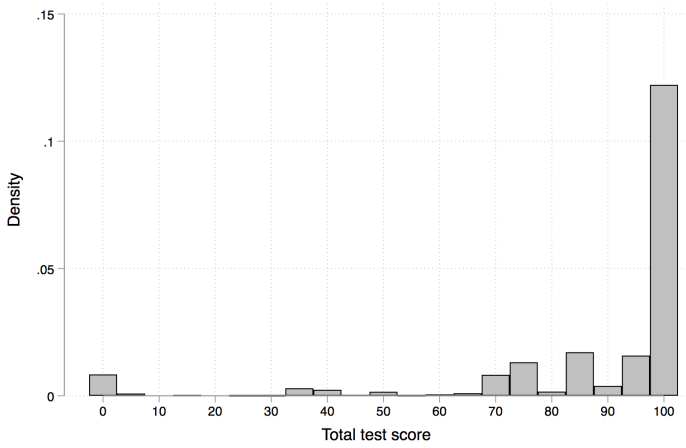
(b) Residuals from a regression on covariates



Reduced-form relationship. The left part of this figure shows the relation on raw data, the right part of the figure plots residuals from a regression of "Labor force participation" (observed in 2013) on a set of covariates from the baseline survey (education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the "Ile-de-France" region (i.e., the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the immigrant has already spent in France, dummy variables indicating the reason to migrate, a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects). The variable "Total test score" refers to the language entry exam and is observed at baseline in interval of 5 points. The line shows the predicted values of a local quadratic smoother with rectangular kernel.

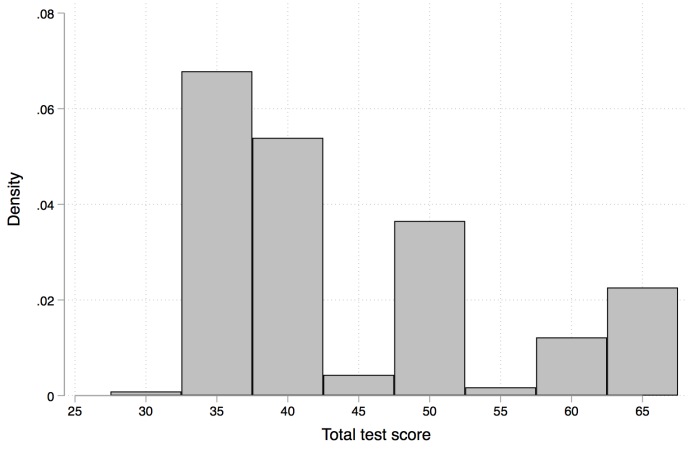
Figure 3:

Density plot over the whole sample



The y-axis represents the density, the x-axis the total test score (which refers to the language entry exam and is observed at baseline in interval of 5 points).

Density plot zoomed on the threshold



The y-axis represents the density, the x-axis the total test score (which refers to the language entry exam and is observed at baseline in interval of 5 points).

Table 1: Test of Balance of Pre-Treatment Characteristics, in 2010.

Test of Balance of Pre-Treatment Characteristics, in 2010.

Panel A: Demographic and household characteristics

	Education level	Age	Permanence in France	Ile-de-France	Married	Male	Number of children	Number of people in HH	Employed in 2010	Taking other French classes
I(test score initial language exam < 50)	-0.772	0.787	-2.071	-0.397	-0.048	0.152	0.013	-0.897	-0.205	0.010
	[2.381]	[4.094]	[1.863]	[0.243]	[0.125]	[0.210]	[0.378]	[0.593]	[0.202]	[0.136]
Polynomial of degree	2	2	2	2	2	2	2	2	2	2
Observations	4,969	5,299	5,298	5,307	4,231	5,307	5,307	5,307	5,303	5,304

Panel B: Types of migration and regions of origin

	Work migration	Refugee migration	Family migration	Europe	Asia	Maghreb	Sub-Sah. Africa	Africa (other)	America & Oceania	
I(test score initial language exam < 50)	-0.080	0.120	0.035	0.113	-0.256	0.315	0.019	-0.098	-0.093	
	[0.097]	[0.185]	[0.216]	[0.171]	[0.271]	[0.451]	[0.218]	[0.158]	[0.118]	
Polynomial of degree	2	2	2	2	2	2	2	2	2	
Observations	5,307	5,307	5,307	5,307	5,307	5,307	5,307	5,307	5,307	

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 2, interacted with I(test score initial language exam < 50).
*** p<0.01, ** p<0.05, * p<0.1.

Table 2: Assigned French Language Training and Labor Force Participation.

Assigned French Language Training and Labor Force Participation.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Local linear estimates				Parametric estimates			
	[35, 60]	[35, 60]	[30, 65]	[30, 65]				
Panel A: First stage regressions (depedent variable: number of assigned hours of French language training/100)								
I(test score initial language exam < 50)	3.317***	3.294***	3.154***	2.910***	1.591***	1.790***	2.665***	2.576***
	[0.507]	[0.655]	[0.492]	[0.370]	[0.312]	[0.287]	[0.499]	[0.466]
Panel B: Second stage regressions (dependent variable: labor force participation dummy)								
Hours of French classes (/100)	0.184**	0.266**	0.167**	0.190**	0.047	0.145**	0.174**	0.206**
	[0.080]	[0.117]	[0.076]	[0.075]	[0.062]	[0.072]	[0.077]	[0.098]
Polynomial of degree	1	1	1	1	1	1	2	2
Control variables	No	Yes	No	Yes	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	67	67	76	76	2.168	2.168	2.168	2,168

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category.), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects.
*** p<0.01, ** p<0.05, * p<0.1.

Table 3: Assigned French Language Training and LFP. Heteroneneity effects by education level.

Assigned French Language Training and Labor Force Participation. Heteroneneity effects by education level.

	(1)	(2)	(3)	(4)
	Local linear estimates		Parametric estimates	
	[35, 60]	[30, 65]		
Second stage regressions (dependent variable: labor force participation dummy)				
Hours of French classes (/100) (a_0)	0.095*	0.097**	0.110*	0.162*
	[0.054]	[0.039]	[0.066]	[0.088]
Hours of French classes (/100) X education level (a_1)	0.024*	0.024	0.008*	0.011*
	[0.014]	[0.015]	[0.004]	[0.006]
H_0: $a_0+12*a_1=0$ (p-value)	0.0064	0.0214	0.0419	0.0363
Polynomial of degree	1	1	1	2
Control variables	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes
Observations	67	76	2,168	2,168

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam. The median value of education level is 12 years.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects.
*** p<0.01, ** p<0.05, * p<0.1.

Table 4: Assigned French Language Training and LFP. Heterogeneity of effects by migration category.

Assigned French Language Training and Labor Force Participation. Heterogeneity of effects by migration category (labor migration, refugee migration, and family migration).

	(1)	(2)	(3)	(4)
	Local linear estimates		Parametric estimates	
	[35, 60]	[30, 65]		
Panel A: Second stage regressions (dependent variable: labor force participation dummy). Labor migration versus other types of migration (refugee and family migration)				
Hours of French classes (/100) (a_0)	0.265**	0.187**	0.143**	0.205**
	[0.119]	[0.082]	[0.071]	[0.097]
Hours of French classes (/100) X labor migration (a_1)	0.107	-0.017	0.076	0.095
	[0.943]	[0.495]	[0.052]	[0.070]
Labor migration	0.044	-0.026	-0.031**	-0.034**
	[0.635]	[0.246]	[0.015]	[0.016]
H_0: $a_0+a_1=0$ (p-value)	0.6851	0.7192	0.0246	0.0233
Panel B: Second stage regressions (dependent variable: labor force participation dummy). Refugee migration versus other types of migration (labor and family migration)				
Hours of French classes (/100) (a_0)	0.263**	0.197**	0.141*	0.204**
	[0.122]	[0.094]	[0.075]	[0.103]
Hours of French classes (/100) X refugee migration (a_1)	0.012	-0.033	0.025	0.009
	[0.139]	[0.124]	[0.057]	[0.064]
Refugee migration	-0.060	-0.052	0.005	0.009
	[0.391]	[0.429]	[0.047]	[0.048]
H_0: $a_0+a_1=0$ (p-value)	0.0674	0.0498	0.0167	0.0165
Panel C: Second stage regressions (dependent variable: labor force participation dummy). Family migration versus other types of migration (labor and refugee migration)				
Hours of French classes (/100) (a_0)	0.357*	0.175*	0.178***	0.234**
	[0.193]	[0.100]	[0.068]	[0.091]
Hours of French classes (/100) X family migration (a_1)	-0.097	0.012	-0.039	-0.033
	[0.178]	[0.129]	[0.046]	[0.050]
Family migration	-0.048	-0.027	0.018	0.016
	[0.327]	[0.221]	[0.022]	[0.022]
H_0: $a_0+a_1=0$ (p-value)	0.0338	0.0421	0.0657	0.0477
Polynomial of degree	1	1	1	2
Control variables	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes
Observations	67	76	2,168	2,168

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects.
*** p<0.01, ** p<0.05, * p<0.1.

Table 5: Assigned French Language Training and LFP. Heterogeneity of effects by gender and age.

Assigned French Language Training and Labor Force Participation. Heterogeneity of effects by gender and age.

	(1)	(2)	(3)	(4)
	Local linear estimates		Parametric estimates	
	[35, 60]	[30, 65]		
Panel A: Second stage regressions (dependent variable: labor force participation dummy). Women and Men				
Hours of French classes (/100) (a_0)	0.575**	0.266**	0.153**	0.213**
	[0.267]	[0.112]	[0.072]	[0.104]
Hours of French classes (/100) X female (a_1)	-0.401**	-0.142	-0.014	-0.012
	[0.188]	[0.102]	[0.041]	[0.045]
Female	-0.081	-0.203	-0.187***	-0.189***
	[0.251]	[0.155]	[0.030]	[0.030]
H_0: $a_0+a_1=0$ (p-value)	0.2238	0.1641	0.0846	0.0466
Panel B: Second stage regressions (dependent variable: labor force participation dummy). Above and below median age (32 years old)				
Hours of French classes (/100) (a_0)	0.349**	0.250**	0.168**	0.233**
	[0.158]	[0.104]	[0.075]	[0.098]
Hours of French classes (/100) X above median age (a_1)	-0.110	-0.097	-0.055	-0.059*
	[0.092]	[0.083]	[0.037]	[0.035]
H_0: $a_0+a_1=0$ (p-value)	0.0399	0.0426	0.0712	0.0545
Panel C: Second stage regressions (dependent variable: labor force participation dummy). Top quintile age (>=40 years old)				
Hours of French classes (/100) (a_0)	0.266**	0.195**	0.156**	0.214**
	[0.115]	[0.078]	[0.074]	[0.099]
Hours of French classes (/100) X top quintile age (a_1)	0.001	-0.030	-0.053	-0.053
	[0.092]	[0.089]	[0.043]	[0.046]
H_0: $a_0+a_1=0$ (p-value)	0.0574	0.0923	0.1264	0.0793
Polynomial of degree	1	1	1	2
Control variables	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes
Observations	67	76	2,168	2,168

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is female, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects.
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Assigned French Language Training and Other Measures of Labor Market Integration.

Assigned French Language Training and other measures of Labor Market Integration.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Local linear estimates				Parametric estimates		
	[35, 60]	[35, 60]	[30, 65]	[30, 65]				
Panel A: Dependent variable: employment (conditional on labor force participation). Second stage regressions.								
Hours of French classes (/100)	-0.041	-0.134**	-0.043	-0.069	-0.015	0.053	-0.087	0.007
	[0.113]	[0.062]	[0.110]	[0.069]	[0.068]	[0.041]	[0.105]	[0.068]
Observations	51	51	58	58	1,734	1,734	1,734	1,734
Panel B: Dependent variable: full-time work (conditional on employment). Second stage regressions.								
Hours of French classes (/100)	-0.126	0.557	-0.184	0.976	-0.317**	-0.197***	-0.263	-0.240**
	[0.173]	[0.347]	[0.181]	[0.703]	[0.143]	[0.067]	[0.200]	[0.120]
Observations	46	46	52	52	1,387	1,387	1,387	1,387
Panel C: Dependent variable: permanent contract (conditional on employment). Second stage regressions.								
Hours of French classes (/100)	-0.173	0.256	-0.120	0.271	-0.232*	-0.059	-0.116	0.084
	[0.150]	[0.232]	[0.125]	[0.339]	[0.118]	[0.065]	[0.132]	[0.117]
Observations	46	46	52	52	1.387	1,387	1,387	1,387
Panel D: Dependent variable: informal work (conditional on employment). Second stage regressions.								
Hours of French classes (/100)	-0.090	0.055	-0.143*	-0.010	0.029	0.004	-0.090	-0.053
	[0.089]	[0.161]	[0.081]	[0.169]	[0.065]	[0.031]	[0.074]	[0.055]
Observations	46	46	52	52	1,387	1,387	1,387	1,387
Panel E: Dependent variable: income per household individual. Second stage regressions.								
Hours of French classes (/100)	56.511	184.562**	-58.650	116.946	-60.575	-100.908**	-148.690*	-50.056
	[112.502]	[90.264]	[91.113]	[72.233]	[86.006]	[51.142]	[87.842]	[54.728]
Observations	61	61	69	69	2,040	2,040	2,040	2,040
Polynomial of degree	1	1	1	1	1	1	2	2
Control variables	No	Yes	No	Yes	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes	No	Yes	No	Yes

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects.
*** p<0.01, ** p<0.05, * p<0.1.

Table 7: Positive Selection out of Language Training: Labor Force Participation.

Positive Selection Out Of Language Training: Labor Force Participation.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Local linear estimates				Parametric estimates			
	[35, 60]	[35, 60]	[30, 65]	[30, 65]				
Panel A: By education level.								
Second stage regressions (dependent variable: labor force participation dummy)								
Hours of French classes (/100)	0.168**	0.215***	0.153**	0.144***	0.040	0.143**	0.160**	0.188**
	[0.068]	[0.069]	[0.067]	[0.043]	[0.063]	[0.071]	[0.071]	[0.095]
Panel B: By permanence in France.								
Second stage regressions (dependent variable: labor force participation dummy)								
Hours of French classes (/100)	0.198**	0.245**	0.209**	0.171*	0.063	0.153**	0.219**	0.213**
	[0.080]	[0.121]	[0.086]	[0.094]	[0.063]	[0.072]	[0.086]	[0.105]
Panel C: By language distance (Francophone vs. Non-Francophone).								
Second stage regressions (dependent variable: labor force participation dummy)								
Hours of French classes (/100)	0.174**	0.264**	0.151*	0.200***	0.039	0.143**	0.157**	0.199**
	[0.077]	[0.115]	[0.078]	[0.077]	[0.064]	[0.073]	[0.079]	[0.101]
Panel D: By employment in 2010.								
Second stage regressions (dependent variable: labor force participation dummy)								
Hours of French classes (/100)	0.181**	0.233**	0.194**	0.186**	0.060	0.150**	0.207**	0.208**
	[0.073]	[0.101]	[0.080]	[0.082]	[0.062]	[0.072]	[0.081]	[0.102]
Polynomial of degree	1	1	1	1	1	1	2	2
Control variables	No	Yes	No	Yes	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	67	67	76	76	2.168	2.168	2.168	2,168

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects.
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Assigned French Language Training and LFP: Excluding Test Result = 0.

Assigned French Language Training and Labor Force Participation: Excluding Test Result = 0.

	(1)	(2)	(3)	(4)
		Parametric estimates		
Second stage regressions (dependent variable: labor force participation dummy)				
Hours of French classes (/100)	0.085	0.178**	0.160**	0.181*
	[0.059]	[0.070]	[0.072]	[0.097]
Polynomial of degree	1	1	2	2
Control variables	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes
Observations	2,093	2,093	2,093	2,093

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects.
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Mechanisms. Improvement in Language Skills and Networks.

Mechanisms. Improvement in language skills and networks.

	(1)	(2) Parametric estimates	(3)	(4)	(5) Parametric estimates	(6)
Panel A: Improvement in language skills						
Second stage regressions. Dependent variable:		**oral skills**			**writing skills**	
Hours of French classes (/100) (a_0)	0.309	0.396	2.177**	-1.363	-1.240	-1.391
	[0.738]	[0.599]	[1.102]	[2.205]	[2.058]	[3.170]
Hours of French classes (/100) X language distance (a_1)			-2.011***			0.035
			[0.753]			[2.089]
Language distance			-0.370			-1.509***
			[0.409]			[0.430]
H_0: a_0+0.62*a_1=0 (p-value)			0.2092			0.5403
Observations	2,000	2,000	2,000	1,903	1,903	1,903
Second stage regressions. Dependent variable:		**subjective oral skills**			**subjective writing skills**	
Hours of French classes (/100) (a_0)	0.140	0.075	0.161	-0.109	-0.216*	-0.292**
	[0.098]	[0.104]	[0.144]	[0.103]	[0.117]	[0.141]
Hours of French classes (/100) X language distance (a_1)			-0.095			0.078
			[0.089]			[0.121]
Language distance			-0.041*			-0.118***
			[0.022]			[0.030]
H_0: a_0+0.62*a_1=0 (p-value)			0.3609			0.0336
Observations	2,136	2,136	2,136	2,142	2,142	2,142
Second stage regressions. Dependent variable:		**no other French classes**			**interview held in French**	
Hours of French classes (/100)	-0.001	0.002	-0.041	0.084	-0.015	0.112
	[0.031]	[0.029]	[0.044]	[0.099]	[0.079]	[0.108]
Hours of French classes (/100) X language distance (a_1)			0.048			-0.144**
			[0.037]			[0.067]
Language distance			-0.006			-0.100***
			[0.017]			[0.021]
H_0: a_0+0.62*a_1=0 (p-value)			0.7079			0.7868
Observations	2,142	2,142	2,142	2,142	2,142	2,142
Polynomial of degree	2	2	2	2	2	2
Control variables	No	Yes	Yes	No	Yes	Yes
Country of origin fixed effects	No	Yes	Yes	No	Yes	Yes
Panel B: Networks						
Second stage regressions. Dependent variable:	**new French friends in 2013**		**new international friends in 2013**		**new friends from same country of origin in 2013**	
Hours of French classes (/100)	0.013	-0.008	-0.036	-0.050	0.023	0.058
	[0.027]	[0.023]	[0.092]	[0.076]	[0.088]	[0.076]
Observations	1,642	1,642	1,642	1,642	1,642	1,642
Second stage regressions. Dependent variable:	**new French friends in 2011**		**new international friends in 2011**		**new friends from same country of origin in 2011**	
Hours of French classes (/100)	-0.072*	-0.069*	0.061	-0.024	0.011	0.093
	[0.040]	[0.039]	[0.096]	[0.070]	[0.087]	[0.062]
Observations	1,665	1,665	1,665	1,665	1,665	1,665
Polynomial of degree	2	2	2	2	2	2
Control variables	No	Yes	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes	No	Yes

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam. The median value of language distance is 0.62.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Mechanisms. Information Effect and Behavioral Effect.

Mechanisms. Information effect and behavioral effect.

	(1)	(2)
	Parametric estimates	
Panel A: Information effect		
Dependent variable: Job search trough "Pole emploi". Second stage regressions.		
Hours of French classes (/100)	0.159***	0.140***
	[0.039]	[0.027]
Observations	363	363
Dependent variable: Found his job through "Pole emploi". Second stage regressions.		
Hours of French classes (/100)	0.340	0.255*
	[0.214]	[0.146]
Observations	1,344	1,344
Dependent variable: Passed driver's license in France. Second stage regressions.		
Hours of French classes (/100)	0.179***	0.188***
	[0.066]	[0.064]
Observations	797	797
Dependent variable: Applied for recognition of the latest academic diploma. Second stage regressions.		
Hours of French classes (/100)	0.353*	0.640**
	[0.198]	[0.273]
Observations	665	665
Panel B: Behavioral effect		
Dependent variable: Feels "at home" in France. Second stage regressions.		
Hours of French classes (/100)	-0.153**	-0.149**
	[0.067]	[0.059]
Observations	2,157	2,157
Dependent variable: Interested in French politics. Second stage regressions.		
Hours of French classes (/100)	-0.206**	-0.145**
	[0.085]	[0.072]
Observations	2,147	2,147
Dependent variable: Interested in European politics. Second stage regressions.		
Hours of French classes (/100)	-0.162	-0.117
	[0.102]	[0.080]
Observations	2,141	2,141
Dependent variable: Interested in origin country politics. Second stage regressions.		
Hours of French classes (/100)	-0.289**	-0.191**
	[0.122]	[0.096]
Observations	2,148	2,148
Polynomial of degree	2	2
Control variables	No	Yes
Country of origin fixed effects	No	Yes

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50). The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region “Ile-de-France” (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

References

[1] Algan, Y., Dustman, C., Glitz, A. and A. Manning (2010) "The Economic Situation of First and Second Generation Immigrants in France, Germany and UK", *Economic Journal*, Vol. 120(542), pp. F4–F30.

[2] Alesina, A., Harnoss, J. and H. Rapoport (2016) "Birthplace diversity and economic prosperity", *Journal of Economic Growth*, Vol. 21(2), pp. 101-138.

[3] Angrist, J., Battistin, E. and D. Vuri (2016) "In a Small Moment: Class Size and Moral Hazard in the Mezzogiorno", *American Economic Journal: Applied Economics*, forthcoming.

[4] Angrist, J. and J.-S.Pischke (2009) "Mostly Harmless Econometrics: An Empiricist's Companion", Princeton University Press, 2009.

[5] Borjas, G. J. (1994) "The Economics of Immigration", *Journal of Economic Literature*, Vol. XXXII, pp. 1667-1717.

[6] Bleakley, H. and A. Chin (2004) "Language Skills and Earnings: Evidence from Childhood Immigrants", *Review of Economics and Statistics*, Vol. 86, pp. 481-496.

[7] Bleakley, H. and A. Chin (2010) "Age at Arrival, English Proficiency, and Social Assimilation Among US Immigrants", *American Economic Journal: Applied Economics*, Vol. 2(1), pp. 165–192.

[8] Borjas, G. J. (1999) "The economic analysis of immigration.", In *Handbook of labor economics*, Vol. 3, ed. Orley Ashenfelter and David Card, pp. 1697–1760. Amsterdam: Elsevier.

[9] Calonico, S., Cattaneo M. D. and R. Titiunik (2014) "Robust data-driven inference in the regression-discontinuity design", *The Stata Journal*, Vol. 14(4), pp. 909-946.

[10] Cameron, A. C., Gelbach, J. and D. L. Miller (2011) "Robust Inference with Multi-Way Clustering", *Journal of Business and Economic Statistics*, Vol. 29(2), pp. 238–249.

[11] Cameron, A. C. and D. L. Miller (2015) "A Practitioner's Guide to Cluster-Robust Inference", *Journal of Human Resources*, Vol. 50(2), pp. 317-373.

[12] Cappellari, L. and A. Di Paolo (2015) "Bilingual Schooling and Earnings: Evidence from a Language-in-Education Reform", *IZA Discussion Papers*, N. 9431, Institute for the Study of Labor (IZA).

[13] Cattaneo, M. D., Keele, L., Titiunik, R. and G. Vazquez-Bare (2015) "Identification in Regression Discontinuity Designs with Multiple Cutoffs", *Work in progress, University of Michigan*

[14] Chin, A., Daysal, N. M., and S. A. Imberman (2013) "Impact of bilingual education programs on limited English proficient students and their peers: Regression discontinuity evidence from Texas", *Journal of Public Economics*, Vol. 107, pp. 63-78.

[15] Chiswick, B. R. (1991) "Speaking, reading, and earnings among low-skilled immigrants" *Journal of Labor Economics*, Vol. 9(2), pp. 149-170.

[16] Chiswick, B. R., and P. W. Miller (1995) "The endogeneity between language and earnings: International analyses." *Journal of Labor Economics*, Vol. 13(2), pp. 246-288.

[17] Davidson, R., and J. MacKinnon (2010) "Wild Bootstrap Tests for IV Regression" *Journal of Business & Economic Statistics*, Vol. 28(1), pp. 128-144.

[18] Desmet, K., Ortuño, I. and R. Wacziarg (2012) "The Political Economy of Linguistic Cleavages", *Journal of Development Economics*, Vol. 97, pp. 322-338.

[19] Devillanova, C., Fasani, F. and T. Frattini (forthcoming) "Employment of Undocumented Immigrants and the Prospect of Legal Status: Evidence from an Amnesty Program", *Industrial and Labor Relations Review*

[20] Direction générale des étrangers en France, Département des statistiques, des études et de la documentation (2010) "Enquete Longitudinale sur l'Integration des Primo-Arrivants 2010 – Vague 1: Documentation et Guide d'utilisation"

[21] Direction générale des étrangers en France, Département des statistiques, des études et de la documentation (2011) "Enquete Longitudinale sur l'Integration des Primo-Arrivants 2011 – Vague 2: Documentation et Guide d'utilisation"

[22] Direction générale des étrangers en France, Département des statistiques, des études et de la documentation (2013) "Enquete Longitudinale sur l'Integration des Primo-Arrivants 2013 – Vague 3: Documentation et Guide d'utilisation"

[23] Dustmann, C. and F. Fabbri (2003) "Language proficiency and labour market performance of immigrants in the UK", *Economic Journal*, Vol. 113(489), pp. 695–717.

[24] Dustmann, C., Fasani, F. and B. Speciale (2017) "Illegal migration and consumption behavior of immigrant households", *Journal of the European Economic Association*, Vol. 15(3), pp. 654-691.

[25] Dustmann, C. and A. van Soest (2001) "Language fluency and earnings: Estimation with misclassified language indicators", *Review of Economics and Statistics*, Vol. 83(4), pp. 663–674.

[26] Fearon, J. D. (2003) "Ethnic and cultural diversity by country", *Journal of Economic Growth*, Vol. 8(2), pp. 195–222.

[27] Gathmann, C. (2015) "Naturalization and citizenship: Who benefits?", *IZA World of Labor, Institute for the Study of Labor (IZA)* Feburary, p. 125.

[28] Gelman, A. and G. Imbens (2016) "Why High-order Polynomials Should not be Used in Regression Discontinuity Designs", Mimeo

[29] Hinnerich, B. T. and P. Pettersson-Libdom (2014) "Democracy, Redistribution, and Political Participation: Evidence from Sweden 1919-1938", *Econometrica*, Vol. 82(3), pp. 961-993.

[30] Jacob, B. A. and L. Lefgren (2004) "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis", *The Review of Economics and Statistics*, Vol. 86(1), pp. 226-244.

[31] Kossoudji, S. A. and D. Cobb-Clark (2002) "Coming Out of the Shadows: Learning about Legal Status and Wages from the Legalized Population", *Journal of Labor Economics*, Vol. 20, pp. 598-628.

[32] Laitin, D. D. and R. Ramachandran (2016) "Language Policy and Human Development", *American Political Science Review*, 110(3), pp. 457-480.

[33] Landsberger, H. A. (1958) *Hawthorne Revisited. Management and the worker: its critics, and developments in human relations in industry.*, Ithaca, New York: Cornell University, 132 pp.

[34] Le Quentrec-Creven, G. (2014) "De la formation linguistique à l'intégration des nouveaux-migrants ?", Actes du XVIIe colloque international de l'AIDELF sur *Démographie et politiques sociales*, Ouagadougou; pp. 1-22.

[35] Lee, D. S., and D. Card (2008) "Regression discontinuity inference with specification error", *Journal of Econometrics*, Vol. 142, pp. 655-674.

[36] Lee, D. and T. Lemieux (2010) "Regression Discontinuity Designs in Economics", *Journal of Economic Literature*, Vol. 48(2), pp. 281-355.

[37] Manacorda, M. (2012) "The Cost of Grade Retention", *The Review of Economics and Statistics*, Vol. 94(2), pp. 596-606.

[38] Matsudaira, J. D. (2008) "Mandatory summer school and student achievement", *Journal of Econometrics*, Vol. 142, pp. 829-850.

[39] McCrary, J. (2008) "Manipulation of the running variable in the regression discontinuity design: A density test.", *Journal of Econometrics*, Vol. 142(2), pp. 698–714.

[40] McKenzie, D. and H. Rapoport (2010) "Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks.", *The Review of Economics and Statistics*, Vol. 92(4), pp. 811-821.

[41] Pinotti, P. (2017) "Clicking on Heaven's Door: The Effect of Immigrant Legalization on Crime", *American Economic Review*, Vol. 107(1), pp. 138-168.

[42] Sarvimäki, M. and K. Hämäläinen (2016) "Integrating Immigrants: The Impact of Restructuring ALMP", *Journal of Labor Economics*, Vol. 34(2), pp. 479-508.

[43] Service-Public.fr, Le site officiel de l'administration française (n.d.) *Travail*, Retrieved from https://www.service-public.fr/

[44] Urquiola, M. and E. Verhoogen (2009) "Class-Size Caps, Sorting, and the Regression-Discontinuity Design", *American Economic Review*, Vol. 99(1), pp. 179-215.

[45] van der Klaauw, W. (2002) "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach", *International Economic Review*, Vol. 43(4), pp. 1249–1287.

Appendix

Table A1: Descriptive statistics, sociodemographic characteristics by migration reason (in percentages)

	Economic migration	Family migration	Refugees
Gender			
Men	0.80	0.41	0.58
Women	0.20	0.59	0.42
Age in 2010			
18 to 29	0.22	0.46	0.46
30 to 39	0.50	0.37	0.34
40 to 49	0.21	0.12	0.12
50 or more	0.07	0.04	0.08
Age at arrival			
0 to 17	0.03	0.06	0.13
18 to 29	0.65	0.57	0.48
30 to 39	0.24	0.27	0.24
40 to 49	0.06	0.07	0.10
50 or more	0.01	0.02	0.05
Duration of stay in France in 2010			
Less than 2 years	0.09	0.60	0.30
From 2 to 4	0.10	0.12	0.51
From 5 to 9	0.59	0.22	0.16
10 years or more	0.22	0.07	0.04
Nationality			
Maghreb	0.20	0.46	0.01
Sub-Saharan Africa	0.51	0.21	0.23
Other areas in Africa	0.05	0.06	0.14
Asia	0.12	0.16	0.44
Europe (excluding France) and CIS	0.05	0.05	0.14
America and Oceania	0.07	0.06	0.03

Table A2: Descriptive statistics, labor market outcomes by gender (in percentages)

	All	Men	Women
Labor force participation			
Before migration	0.57	0.67	0.47
2010	0.69	0.88	0.51
2011	0.71	0.90	0.52
2013	0.77	0.94	0.62
Employment rate			
Before migration	0.52	0.63	0.43
2010	0.46	0.67	0.26
2011	0.54	0.74	0.35
2013	0.59	0.78	0.42
Unemployment rate			
Before migration	0.08	0.06	0.10
2010	0.33	0.24	0.48
2011	0.24	0.18	0.34
2013	0.21	0.15	0.29

Table A3: Positive Selection out of Training: Employment.

Positive Selection Out Of Language Training: Employment.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Local linear estimates				Parametric estimates			
	[35, 60]	[35, 60]	[30, 65]	[30, 65]				
PART I: By education level.								
Second stage regressions (dependent variable: employment dummy)								
Hours of French classes (/100)	-0.025	-0.153***	-0.013	-0.106*	-0.008	0.054	-0.057	0.020
	[0.117]	[0.058]	[0.126]	[0.062]	[0.071]	[0.042]	[0.123]	[0.079]
PART II: By permanence in France.								
Second stage regressions (dependent variable: employment dummy)								
Hours of French classes (/100)	-0.023	-0.137**	-0.009	-0.062	-0.008	0.051	-0.053	0.013
	[0.117]	[0.055]	[0.125]	[0.065]	[0.071]	[0.043]	[0.123]	[0.082]
PART III: By language distance (Francophone vs. Non-Francophone).								
Second stage regressions (dependent variable: employment dummy)								
Hours of French classes (/100)	-0.039	-0.141**	-0.038	-0.079	-0.014	0.055	-0.082	0.017
	[0.114]	[0.064]	[0.113]	[0.072]	[0.069]	[0.041]	[0.109]	[0.069]
PART IV: By employment in 2010.								
Second stage regressions (dependent variable: employment dummy)								
Hours of French classes (/100)	-0.051	-0.076	-0.038	-0.054	-0.010	0.052	-0.077	0.006
	[0.093]	[0.075]	[0.103]	[0.063]	[0.070]	[0.042]	[0.106]	[0.069]
Polynomial of degree	1	1	1	1	1	1	2	2
Control variables	No	Yes	No	Yes	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	51	51	58	58	1,734	1,734	1,734	1,734

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table A4: Positive Selection out of Training: Full-time Contract.

Positive Selection Out Of Language Training: Full-time contract.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Local linear estimates				Parametric estimates			
	[35, 60]	[35, 60]	[30, 65]	[30, 65]				
PART I: By education level.								
Second stage regressions (dependent variable: full-time contract dummy)								
Hours of French classes (/100)	-0.187	0.337	-0.247	0.552*	-0.349**	-0.214***	-0.344	-0.309**
	[0.160]	[0.213]	[0.178]	[0.286]	[0.148]	[0.071]	[0.217]	[0.139]
PART II: By permanence in France.								
Second stage regressions (dependent variable: full-time contract dummy)								
Hours of French classes (/100)	-0.187	0.519*	-0.242	0.918	-0.349**	-0.209***	-0.340	-0.288**
	[0.160]	[0.314]	[0.174]	[0.585]	[0.149]	[0.070]	[0.216]	[0.131]
PART III: By language distance (Francophone vs. Non-Francophone)								
Second stage regressions (dependent variable: full-time contract dummy)								
Hours of French classes (/100)	-0.126	0.635	-0.176	1.177	-0.316**	-0.201***	-0.255	-0.259**
	[0.174]	[0.418]	[0.182]	[0.991]	[0.144]	[0.069]	[0.202]	[0.129]
PART IV: By employment in 2010.								
Second stage regressions (dependent variable: full-time contract dummy)								
Hours of French classes (/100)	-0.170	0.487*	-0.185	0.592*	-0.322**	-0.197***	-0.278	-0.239**
	[0.162]	[0.283]	[0.166]	[0.340]	[0.144]	[0.068]	[0.196]	[0.115]
Polynomial of degree	1	1	1	1	1	1	2	2
Control variables	No	Yes	No	Yes	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	46	46	52	52	1,387	1,387	1,387	1,387

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Positive Selection out of Training: Permanent Contract.

Positive Selection Out Of Language Training: Permanent contract.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Local linear estimates				Parametric estimates			
	[35, 60]	[35, 60]	[30, 65]	[30, 65]				
PART I: By education level.								
Second stage regressions (dependent variable: permanent contract dummy)								
Hours of French classes (/100)	-0.148	0.323*	-0.117	0.220	-0.233*	-0.074	-0.118	0.049
	[0.132]	[0.192]	[0.121]	[0.226]	[0.122]	[0.069]	[0.142]	[0.124]
PART II: By permanence in France.								
Second stage regressions (dependent variable: permanent contract dummy)								
Hours of French classes (/100)	-0.140	0.343	-0.109	0.607	-0.231*	-0.063	-0.110	0.092
	[0.130]	[0.222]	[0.120]	[0.469]	[0.122]	[0.067]	[0.143]	[0.114]
PART III: By language distance (Francophone vs. Non-Francophone).								
Second stage regressions (dependent variable: permanent contract dummy)								
Hours of French classes (/100)	-0.162	0.263	-0.109	0.178	-0.228*	-0.060	-0.103	0.087
	[0.144]	[0.256]	[0.122]	[0.379]	[0.118]	[0.066]	[0.131]	[0.123]
PART IV: By employment in 2010.								
Second stage regressions (dependent variable: permanent contract dummy)								
Hours of French classes (/100)	-0.111	0.455	-0.103	0.168	-0.232*	-0.062	-0.114	0.067
	[0.117]	[0.299]	[0.115]	[0.206]	[0.120]	[0.066]	[0.131]	[0.112]
Polynomial of degree	1	1	1	1	1	1	2	2
Control variables	No	Yes	No	Yes	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	46	46	52	52	1,387	1,387	1,387	1,387

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Positive Selection out of Training: Informal Work.

Positive Selection Out Of Language Training: Informal work.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Local linear estimates				Parametric estimates			
	[35, 60]	[35, 60]	[30, 65]	[30, 65]				
PART I: By education level.								
Second stage regressions (dependent variable: informal work dummy)								
Hours of French classes (/100)	-0.119	0.046	-0.161*	0.009	0.026	0.003	-0.111	-0.060
	[0.079]	[0.096]	[0.084]	[0.124]	[0.066]	[0.031]	[0.081]	[0.059]
PART II: By permanence in France.								
Second stage regressions (dependent variable: informal work dummy)								
Hours of French classes (/100)	-0.122	-0.077	-0.163*	-0.235	0.026	-0.002	-0.113	-0.079
	[0.081]	[0.174]	[0.086]	[0.245]	[0.066]	[0.031]	[0.084]	[0.059]
PART III: By language distance (Francophone vs. Non-Francophone)								
Second stage regressions (dependent variable: informal work dummy)								
Hours of French classes (/100)	-0.094	0.102	-0.146*	0.057	0.029	0.005	-0.093	-0.054
	[0.087]	[0.180]	[0.082]	[0.199]	[0.065]	[0.031]	[0.076]	[0.057]
PART IV: By employment in 2010.								
Second stage regressions (dependent variable: informal work dummy)								
Hours of French classes (/100)	-0.130*	-0.019	-0.141*	-0.024	0.030	0.004	-0.088	-0.052
	[0.078]	[0.089]	[0.079]	[0.103]	[0.065]	[0.031]	[0.073]	[0.053]
Polynomial of degree	1	1	1	1	1	1	2	2
Control variables	No	Yes	No	Yes	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	46	46	52	52	1,387	1,387	1,387	1,387

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Positive Selection out of Training: Income per Household Individual.

Positive Selection Out Of Language Training: Income per household individual.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Local linear estimates				Parametric estimates			
	[35, 60]	[35, 60]	[30, 65]	[30, 65]				
PART I: By education level.								
Second stage regressions (dependent variable: income per household individual)								
Hours of French classes (/100)	31.891	186.510**	-47.931	129.194*	-59.113	-107.998**	-154.781	-52.557
	[107.860]	[89.266]	[99.547]	[72.587]	[90.052]	[53.815]	[100.749]	[62.384]
PART II: By permanence in France.								
Second stage regressions (dependent variable: income per household individual)								
Hours of French classes (/100)	28.434	185.638**	-49.246	119.001*	-59.864	-112.460**	-157.779	-63.267
	[113.369]	[90.195]	[105.225]	[70.742]	[91.030]	[53.180]	[106.443]	[63.685]
PART III: By language distance (Francophone vs. Non-Francophone)								
Second stage regressions (dependent variable: income per household individual)								
Hours of French classes (/100)	34.37	160.564**	-73.831	82.780	-68.049	-110.177**	-170.971*	-64.454
	[107.636]	[81.002]	[94.388]	[65.569]	[87.299]	[53.040]	[93.854]	[57.369]
PART VI: By employment in 2010.								
Second stage regressions (dependent variable: income per household individual)								
Hours of French classes (/100)	-6.226	120.512*	-47.564	101.850	-55.104	-114.367**	-145.568	-68.499
	[102.199]	[69.367]	[92.563]	[63.676]	[90.681]	[53.543]	[97.355]	[61.213]
Polynomial of degree	1	1	1	1	1	1	2	2
Control variables	No	Yes	No	Yes	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	61	61	69	69	2,040	2,040	2,040	2,040

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Assigned French Language Training and Other Measures of Labor Market Integration. Excluding Test Result = 0.

Assigned French Language Training and other measures of Labor Market Integration. Excluding Test Result = 0.

	(1)	(2)	(3)	(4)
		Parametric estimates		
Panel A: Dependent variable: employment (conditional on labor force participation). Second stage regressions				
Hours of French classes (/100)	-0.058	0.031	-0.052	0.036
	[0.065]	[0.039]	[0.103]	[0.069]
Observations	1,690	1,690	1,690	1,690
Panel B: Dependent variable: full-time work (conditional on employment). Second stage regressions				
Hours of French classes (/100)	-0.348**	-0.200***	-0.212	-0.235*
	[0.157]	[0.066]	[0.190]	[0.128]
Observations	1,357	1,357	1,357	1,357
Panel C: Dependent variable: permanent contract (conditional on employment). Second stage regressions				
Hours of French classes (/100)	-0.239*	-0.076	-0.037	0.181
	[0.124]	[0.073]	[0.127]	[0.123]
Observations	1,357	1,357	1,357	1,357
Panel D: Dependent variable: informal work (conditional on employment). Second stage regressions				
Hours of French classes (/100)	0.029	0.006	-0.142	-0.134**
	[0.058]	[0.028]	[0.091]	[0.064]
Observations	1,357	1,357	1,357	1,357
Panel E: Dependent variable: income per household individual. Second stage regressions				
Hours of French classes (/100)	-73.174	-120.737**	-151.557	-19.797
	[80.126]	[49.976]	[93.462]	[58.151]
Observations	1,967	1,967	1,967	1,967
Polynomial of degree	1	1	2	2
Control variables	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.
The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).
The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table A9: Robustness Checks Using Different Ways of Clustering the Standard Errors: Assigned French Language Training and Labor Force Participation.

Assigned French Language Training and Labor Force Participation.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Local linear estimates				Parametric estimates			
	[35, 60]	[35, 60]	[30, 65]	[30, 65]				
Panel A: Clustering by country of origin times entry test score								
H_0: Hours of French classes=0 (p-value)	0.021	0.023	0.029	0.011	0.449	0.044	0.023	0.035
Panel B: Two-way clustering à la Cameron et al. (2011), by country of origin and entry test score								
H_0: Hours of French classes=0 (p-value)	0.000	0.002	0.000	0.005	0.528	0.042	0.000	0.005
Panel C: One-way clustering, by country of origin								
H_0: Hours of French classes=0 (p-value)	0.033	0.017	0.046	0.037	0.504	0.134	0.026	0.086
Panel D: One-way clustering, by entry test score								
H_0: Hours of French classes=0 (p-value)	0.000	0.005	0.000	0.000	0.482	0.000	0.000	0.000
Panel E: Wild bootstrap, by entry test score								
H_0: Hours of French classes=0 (p-value)	0.067	0.470	0.081	0.176	0.624	0.025	0.018	0.012
Panel F: Robust standard errors								
H_0: Hours of French classes=0 (p-value)	0.029	0.058	0.040	0.074	0.398	0.020	0.043	0.031
Polynomial of degree	1	1	1	1	1	1	2	2
Control variables	No	Yes	No	Yes	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	67	67	76	76	2,168	2,168	2,168	2,168

The dependent variable is a dummy equal to 1 if the immigrant participates in the labor force. This table presents p-values for the null hypothesis H_0: Hours of French classes=0, using different ways of clustering the standard errors. The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50). The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects.