The Effect of Language Training on Refugees' Economic Outcomes

Celine Li

Abstract

This paper investigates the impact of government-provided language training on the labor market outcomes of refugees in Germany. I use individual-level panel data from the IAB-SOEP-BAMF Refugee Survey spanning 2016 to 2021, and employ a Two-Way Fixed Effects (TWFE) research design to estimate the causal effects of language training. The analysis reveals that receiving language training is associated with an 11% increase in gross labor income and a 10.5% rise in the likelihood of employment. Thus, I provide empirical insights into the economic implications of language proficiency for refugee populations in Germany, contributing to the broader literature and policy understanding of migrants' economic integration influenced by language skills. My code and data are shared on GitHub.

1 Introduction

For new immigrants, the development of host-country language skills are important determinants for successful economic integration (Brell et al., 2020). This issue is particularly significant for refugees who, unlike economic migrants, must abruptly leave their home countries, often lacking the opportunity to invest in language skills before departure. It is in the host country's interest to support the process of economic assimilation and to realize immigrants' economic potential (Dustmann & Fabbri, 2003). One important instrument for this are language trainings, which help refugees to build the relevant linguistic skills to succeed in daily life and at the workplace. This motivates the research question: *What are the effects of language trainings on refugees' economic outcomes?*.

Germany is a suitable context to study the labor market integration of refugees due to the large influx of refugees and asylum seekers in recent years. In 2015 and 2016 alone, Germany received over one million first time asylum applications, mostly from people who fled their home countries Syria, Iraq and Afghanistan. As German is not a widely spoken language outside of Europe, like English is, those refugees often arrive with little to no prior knowledge about the German language. The German Federal Office for Migration and Refugees (BAMF) offers "integration courses", which contain a language training component. This course is obligatory for those who received their residence permit after 1 January 2005 and if they are "*unable to make themselves understood at a basic or adequate level in German*". Thus, the German government views language training as a crucial component to successful economic and social integration.

In this paper, I use longitudinal household survey data from the German Socio Economic Panel (SOEP) to estimate the causal impact of language training on refugees' economic outcomes. I build on a large body of literature that examines the effect of language proficiency on migrants' economic integration in predeominantly English-speaking countries. The challenge in isolating the causal effect lies in the endogenous choice of language acquisition and measurement error. Self-reporting of language skills biases the OLS estimator downward, while omitted variable bias (attributing more of the effect to language training) causes an upward bias (Schuss, 2018).

Previous studies aimed at overcoming this challenge by employing instruments such as linguistic distance between German and the origin country language (Isphording et al., 2014), the interaction of the native language with age-at-arrival (Miranda & Zhu, 2013) and parents' exposure to communication with their children in the host country's language (Budría et al., 2017). However, these approaches always hinge on the instrument satisfying the exclusion restriction. My approach relies instead on exploiting the within-individual variation from an actual language training policy. I contribute to the literature by quantifying the effect from *actual* language policy interventions, instead of a general notion of language proficiency, which is hard to measure. Thus, my research is relevant to policymakers who will have to decide about how to allocate resources.

The remaining paper is structured as follows. Section 2 introduces the data source and presents descriptive statistics, Section 3 the identfication strategy and Section 4 the results.

2 Data

I use data from the IAB-BAMF-SOEP Refugee Survey, which runs from 2016 to 2021. Commissioned by the German Federal Office for Migration and Refugees (BAMF), this special longitudinal study annually interviewed people who applied for asylum or temporary protection in Germany between 2013 and September 2022, regardless of whether their application was successful. The participants were selected from the Central Register of Foreigners (AZR) using a random sampling procedure and responses are voluntary. In addition, the respective household members are interviewed as well. For 2016, information on 4,465 adults is available, for 2017 on 5,595, for 2018 on 4,376, for 2019 on 3,906, for 2020 on 4,473 and on 2,636 in 2021. This is after excluding incomplete interviews and respondents younger than 18 years old, which gives us 25,451 person-year observations, corresponding to 10,077 distinct individuals. Each individual was asked about biographical information, their route to Germany, language proficiency, their employment status (coded as 0 or 1) as well as gross monthly income and household income.

Although the aim of the study is to survey the same persons each year, the panel is in fact unbalanced. Only a small fraction of people (n = 494) is observed across all six survey periods. Over half of the sample (59%) is observed only for one or two periods, while 41% of respondents are observed for three or more survey periods. This means that the average number of observation periods for any given individual is between 2 and 3 survey periods. The fact that we do not observe the same individuals across all periods of time is an important limitation of this data, but also reflects the difficult reality of sampling migrant populations, especially refugee populations. Nevertheless, as the survey continues and more waves are added this issue will gradually diminish.

In the interview questionnaire, respondents report whether they have participated in any of six different types of courses and if they did, the course start and end date. There are some differences between the courses, both in terms of scope and length. However, for the purposes of this study, I will pool them together, as all trainings contain a language component. I report the wording of the questions and participant numbers in the appendix A. The "BAMF integration course" is the largest by participation numbers (47.11% of full sample), followed by "other German language course" (25.3%). The BAMF integration course is obligatory for those who have received a residence permit after 1 January 2005 and are "unable to make yourself understood at a basic or adequate level in German". In such cases, the immigration office will determine the obligation to participate when your residence permit is issued. Each integration course consists of a language course as well as an orientation course, which discusses the German legal system, culture and values.

The distribution of treated individuals in the data is pictured in figure 1. As we can see, more and more individuals are treated over time. The "drop" in treated individuals in 2021 is most plausibly due to panel attrition, where the previously surveyed individuals couldn't be contacted for follow-up interviews anymore.

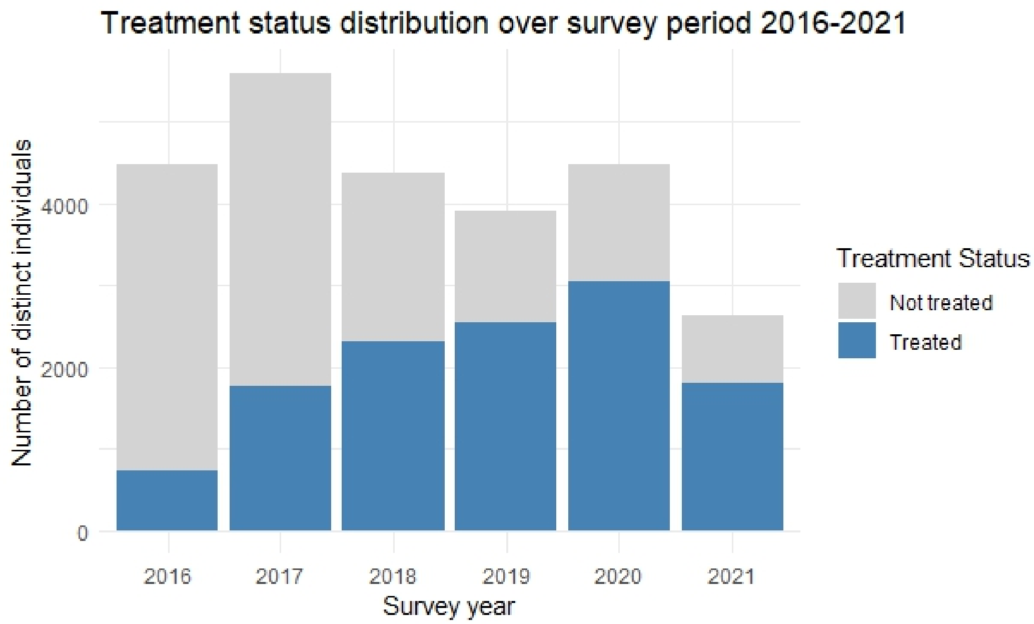

Figure 1: Treatment distribution

Treatment status is defined based on whether an individual has completed languge training in any year prior to the survey year. That is, $D_{it} = 1$ if an individual responded in year t to have participated in language training that ended in the previous year or prior, and 0 otherwise. If a respondent has participated in multiple language courses, I consider only the year of the earlier course. Once an individual has been treated, they are considered to be treated for the remaining survey periods. By defining treatment this way, I assume that treatment has a lasting effect and focus on capturing the longer-term effects of language training instead of immediate ones.

This leads to three types of respondent groups, with the number of distinct persons in brackets:

1. Never treated: Refugees who have never received treatment over the course of the survey period (n = 4,864),
2. Treated before survey: Refugees who report to have already participated in training before the survey period began, and for whom we therefore do not observe before-after differences (n = 2,360),
3. Treated during survey: Refugees who participated in language training over the course of the survey period and for whom we observe before-after outcomes (n = 2,853).

Summary statistics for all different treatment groups are reported in table 1, with the "Treated during survey" group separated by pre- and post treatment periods. The reason behind differentiating between *three* groups, instead of two (control and treatment) is that we do not observe individual-level variation between individuals who were never treated and those who were treated before the survey. These groups could exhibit very different characteristics. But we want to control for individual-invariant characteristics, such as innate ability or motivation, and so we need the third group for identification. This allows us to exploit the within-individual variation.

Treatment Group:	**Never treated**			**Treatment before survey**			**Treatment during survey**					
							Pre treatment			Post treatment		
Variable	Mean	Std. Dev	n	Mean	Std. Dev	n	Mean	Std. Dev	n	Mean	Std. Dev	n
Gender	0.49	0.5	8825	0.69	0.46	5221	0.64	0.48	4411	0.69	0.46	6994
Age	34	12	8824	34	11	5221	35	10	4411	37	10	6994
Married	0.89	0.31	6108	0.9	0.3	3324	0.91	0.28	3332	0.88	0.32	5157
Live in urban area	0.74	0.44	8825	0.74	0.44	5221	0.71	0.46	4411	0.74	0.44	6994
Years since immigration	3	2	8713	4.1	2	5217	1.9	1.1	4393	4.2	1.4	6971
Years of Education	8.4	1.7	7665	9.2	1.9	4677	8.8	1.9	4022	9.1	1.9	6460
Employed	0.15	0.35	8820	0.39	0.49	5219	0.1	0.31	4411	0.39	0.49	6994
Gross Labor Income	1230	980	1301	1469	975	2053	900	1017	461	1466	919	2757
Household Income	1490	865	7971	1565	884	4709	1283	690	4125	1720	1077	6350
Oral Ability German	2.8	1.1	8664	3.5	0.91	5221	2.8	0.86	4408	3.4	0.87	6987
Written Ability German	2.6	1.2	8666	3.4	1	5221	2.8	0.99	4408	3.4	0.97	6989
Reading Ability German	2.7	1.2	8666	3.5	0.99	5221	2.9	1	4409	3.5	0.95	6989

Table 1: Summary Statistics for full sample, split by groups

As we can see from the table by comparing before and after differences for the "treated during survey" group, there is an increase in employment probability from 10% to 39%, and an increase in gross labor income from 900€/month to 1,466€/month. The self-reported language scores for oral, written and reading ability for German also improve above average after receiving language training.

Comparing the "treated during survey" with the "Treated before survey: post treatment" group, we see that these groups exhibit very similar characteristics. This is intuitive, since both groups eventually participate in the language training but simply at different time periods. Comparing these characteristics with the "never treated" group on the other hand, shows some differences: Those who never participate in language training over the course of the survey appear to be more female, have fewer years of education, and a smaller probability of employment as well as lower reported German scores.

Finally, it is important to note here that although the sample of respondents is large, we do not have a complete set of information on one of the key outcome variables. There are a number of missing values (in the data coded as *NA*) for gross labor income, for example, the "Treated during survey: Pre treatment" group has a sample size of 461. This is due to the fact that many refugees, shortly after arriving in Germany, do not instantly earn a wage. This is an important caveat for the analysis that follows.

3 Analysis

I use a two-way fixed effects design and estimate the following equation:

$$Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + \tau D_{it} + u_{it}$$

where Y_{it} is two types of outcome of interest: First, the continuous outcome $\log(wage_{it})$ and second, the binary outcome of whether an individual is employed or not. Individual and time fixed effects are captured by α_i and γ_t, respectively, X_{it} are time-varying covariates, D_{it} is the language course treatment and τ is the causal effect of treatment. As described in the data section, treatment D_{it} is coded 1 if individual i has received treatment at any time before period t, and is 0 otherwise. Individual fixed effects account for an individual's innate ability, allowing us to isolate the effect of the policy, the language training. For covariates, I include age as a covariate as this may affect employment prospects, but I refrain from including other covariates to avoid post-treatment bias.

Goodman-Bacon, 2021 showed that with staggered treatment over time, the two-way fixed effects Differences in Differences (DiD) estimator τ is in fact a weighted average of all possible 2x2 DiD estimators. That is, the Average Treatment Effect (ATE) is simultaneously using different pairs and comparing them to each other: "Never treated" vs. "Treated before survey", "Treated before survey" vs. "Treated during survey" as well as "Treated during survey: pre-training" vs. "Treated during survey post-training". For the estimation, I consider three possible comparisons: The full sample for the ATE, the subset that only contains respondents who were never treated and who were treated over the course of the survey and finally, the subset that only contains respondents who were treated prior to the survey and during the survey. I omit the comparison between "never treated" and "treated before the survey" as there is no within-individual variation. Given eligibility requirements for the language training, one would be worried that those who undergo training have different characteristics compared to the group that did not, and so one is not a good control group for the other.

Instead, I focus my analysis primarily on two group-wise comparisons, between the "Never treated" group vs. "Treated during the survey" and the "Treated during the survey" group vs. "Treated before survey". The latter comparison is especially relevant to consider if we are concerned about selection bias, since it compares groups to each other which all eventually pursue the language training. To ensure that the comparison of the group "Never treated" with the group "Treated during survey" is valid, I check for the parallel pre-trends assumption to hold in the years 2016 to 2017, which are the earliest data available. If the group which will eventually be treated over the course of the survey exhibits a similar trend as the group that will stay untreated, then this gives us confidence in the causal interpretation of our estimates.

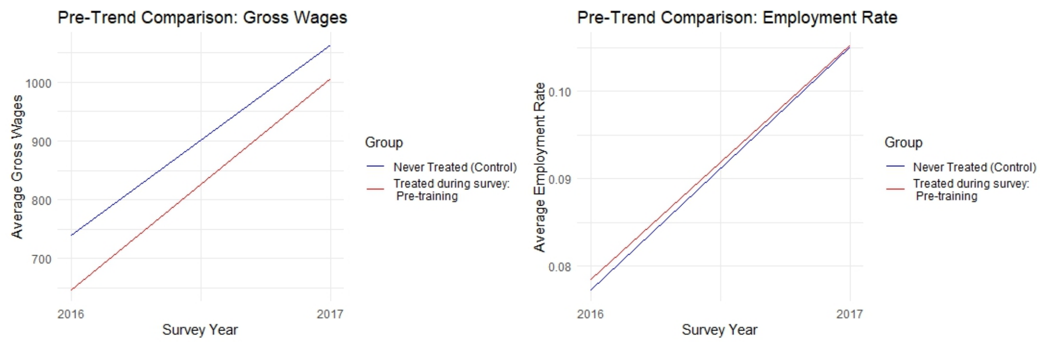

Figure 2: Parallel Pre-Trends between "Never Treated" and "Treated during survey", before training was received

As can be seen from figure 2, the parallel trends assumption appears to be met. While it would be ideal to demonstrate that the pre-trends assumption holds for more time periods than just 2016-2017, we observe individuals only once they entered the survey and are hence limited by the data in this regard.

4 Results

I first report the results for the estimating equation using the outcome variable "log gross income" in table 2. The three columns correspond to different samples over which I estimate the effect, which shed light on the variance decomposition of the DiD estimator τ: Column (1) is based on the full sample, column (2) uses the subset of the "never treated" group and the "treated during survey" group, and column (3) uses the subset of the "treated before survey" and the "treated during survey" group.

For the full sample reported in column (1), the causal effect of treatment is 0.0926, which is only significant at the 10% level. This effect seems to stem from the estimation in column (3), which is the comparison between the "Treated before the survey" with the "treated during the survey". Here, the causal effect of treatment is 11.02%, significant at the 5% level. For the estimation in column (2), the comparison between the "never treated" with the "treated during the survey", there is is no significant effect. We may attribute this to the low statistical power that the estimation in this subsample has, given the lower sample size of the "treated during survey: pre treatment" group for wages, as documented in the descriptive statistics table 1.

Dependent Variable:	Log Income (Gross)		
Model:	(1) Full	(2) Never + During	(3) Before + During
Variables			
Treatment	0.0926*	0.0558	0.1102**
	(0.0483)	(0.0531)	(0.0488)
Age Dummies	X	X	X
Fixed-effects			
pid	Yes	Yes	Yes
syear	Yes	Yes	Yes
Fit statistics			
Observations	6,421	4,406	5,152
R^2	0.85009	0.85324	0.83468
Within R^2	0.02476	0.02663	0.02827

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 2: OLS regression results for continuous gross labor income outcome

Given that the continuous variable "gross labor income" suffers from this limitation of a relatively small sample size of the pre-training observations of the "treated during survey" group, a more preferable measure of labor market outcome is that the binary outcome of whether an individual is employed or not, for which we have information on almost every respondent. Table 3 presents the results of an OLS estimation for the binary outcome variable of employment. Despite the shortcomings of the linear probability model, OLS is still the best linear unbiased estimator and allows us to interpret the coefficients in terms of the increase of probabilities.

Dependent Variable:	Employment		
Model:	(1) Full	(2) Never + During	(3) Before + During
Variables			
Treatment	0.1027***	0.1053***	0.0746***
	(0.0095)	(0.0098)	(0.0109)
Age Dummies	X	X	X
Fixed-effects			
pid	Yes	Yes	Yes
syear	Yes	Yes	Yes
Fit statistics			
Observations	25,443	20,224	16,624
R^2	0.65633	0.63080	0.61829
Within R^2	0.01917	0.02267	0.01539

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 3: OLS regression results for binary employment outcome

The estimation in column (1) shows that the average treatment effect of language training in previous years on employment 0.1027 for the full sample. Since this is a weighted average, this effect seems to be coming largely from estimation (2), which is on the "never treated" and "treated during survey" sample. Here, the likelihood of employment increases by 10.53% following language training as compared to having never taken the training. Given the validity of the pre-trends assumption, this can be interpreted as a causal effect: Those who receive language training are better off in terms of employment prospects than those who never received training at all. In estimation (3), we see that the effect is slightly smaller, at 7.46%. All coefficients are highly significant at the 1% level.

In the Appendix A, I report a third specification, which takes as dependent variable the absolute income level, where missing values for wages are coded as 0€ monthly income.

5 Discussion

The above results are very promising, however, my study has some notable limitations, which are important to highlight here. The first limitation relates to the limited data on individuals' labor market outcomes *before* they received language training. Due to the selection bias inherent in the way refugees sort into the different treatment groups, we cannot directly compare migrants who are treated vs. those who are untreated but instead rely on the within-individual variation before and after receiving the language training. Since the survey only started in 2016 and targeted individuals who arrived since 2013, the pool of people who we observe pre-treatment status is rather limited.

A second limitation of this study is that it pools different language trainings together, which may be sufficiently different. For policymakers, it is more instructive to know which of the trainings (pure language training, vocational German training, courses on culture and legal background) is driving the effect. Since most refugees in the data have participated in the BAMF integration course, the effect is most likely related to this, and others would have to be estimated using a larger sample size of participants.

Overall, my study provides further evidence in favor of the hypothesis that language proficiency is an important complementary good through which immigrants are able to apply home country human capital to the destination context (Chiswick & Miller, 2003). To the best of my knowledge, it is among the first to aim at quantifying the causal impact of government-sponsored language training for refugees in the context of Germany, using the data contained within the SOEP. Policymakers and stakeholders in the field of immigration and integration may find these results valuable for designing and implementing effective language training policies. With increasing survey waves, the data availability will improve and address some of the concerns on data limitations discussed in this paper.

References

Brell, C., Dustmann, C., & Preston, I. (2020). The Labor Market Integration of Refugee Migrants in High-Income Countries. *Journal of Economic Perspectives*, *34*(1), 94–121. https://doi.org/10.1257/jep.34.1.94

Budría, S., Martinez de Ibarreta, C., & Swedberg, P. (2017). The impact of host language proficiency across the immigrants' earning distribution in Spain. *IZA Journal of Development and Migration*, 7(1), 12. https://doi.org/10.1186/s40176-017-0094-2

Chiswick, B. R., & Miller, P. W. (2003). The complementarity of language and other human capital: Immigrant earnings in canada. *Economics of Education review*, *22*(5), 469–480.

Dustmann, C., & Fabbri, F. (2003). Language Proficiency and Labour Market Performance of Immigrants in the UK. *The Economic Journal*, *113*(489), 695–717. https://doi.org/10.1111/1468-0297.t01-1-00151

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277.

Isphording, I. E., Otten, S., & Sinning, M. (2014). The Effect of Language Deficiency on Immigrant Labor Market Outcomes in Germany.

Miranda, A., & Zhu, Y. (2013). English deficiency and the native–immigrant wage gap. *Economics Letters*, *118*(1), 38–41. https://doi.org/10.1016/j.econlet.2012.09.007

Schuss, E. (2018). The Impact of Language Skills on Immigrants' Labor Market Integration: A Brief Revision With a New Approach. *The B.E. Journal of Economic Analysis & Policy*, *18*(4), 20170280. https://doi.org/10.1515/bejeap-2017-0280

A Appendix

A.1 Replication Package

The cleaned data used in this paper and the replication code can be found on GitHub under the following link: https://github.com/celineli99/aqrd

A.2 Questions from the SOEP interviews

Section B017 Courses and Government Measures.

Have you attended an integration course organised by the German Federal Ministry for Migration and Refugees (BAMF)? When did the integration course start? When did it end?

Have you participated in an ESF BAMF course to learn vocational German? When did your ESF BAMF course to learn vocational German start?

Have you attended an entry course for German language skills organised by the German Federal Employment Agency (pursuant to Section 421 Social Code [SGB] III? When did the entry course for German language skills start? When did it end?

Have you attended the "Perspectives for Refugees" course organised by the German Federal Employment Agency [Bundesagentur für Arbeit]? When did the "Perspectives for Refugees" course start? When did it end?

Have you attended the "Perspectives for Young Refugees" course organised by the German Federal Employment Agency [Bundesagentur für Arbeit]? When did the "Perspectives for Young Refugees" course start? When did it end?

Have you attended any other German language courses? When did this other German language course start? When did it end?

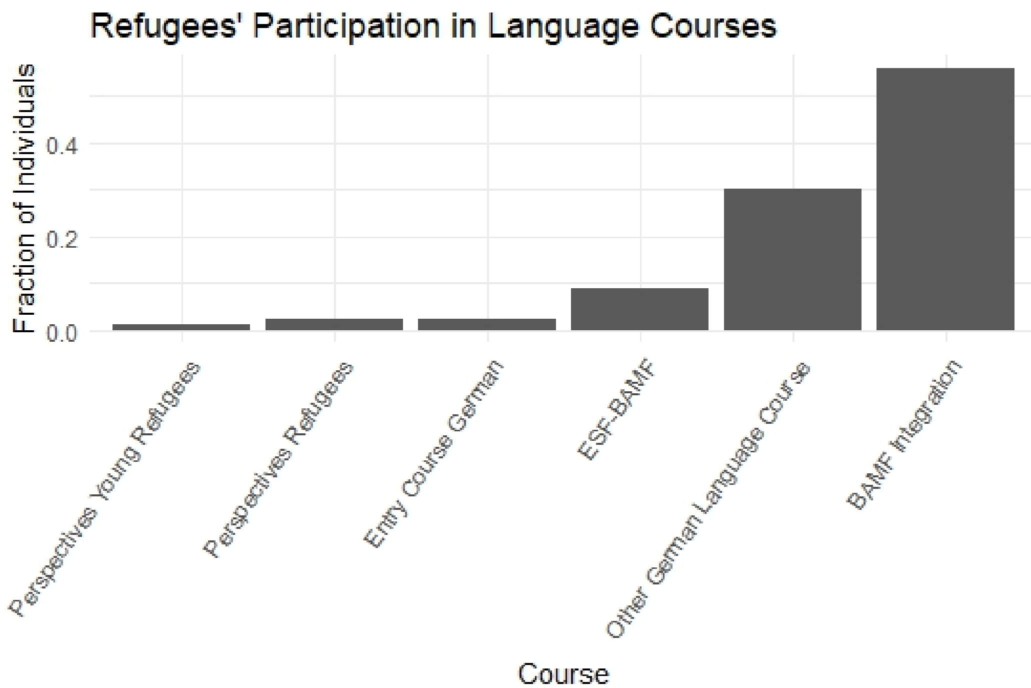

Figure 3: Number of participants (unique individuals) in language training offered by the Federal Government

A.3 Additional Regression on income level

In addition to estimating the effect of treatment on log wages and binary employment status, I also estimate the effect of treatment on absolute gross income (not logged income). For this, I coded the missing values in the gross income variables as zeroes, given that an individual only has missing values if they are registered as unemployed. This regression then estimates the effect of language training on absolute income, which is a mix between the effect on employment and the effect of wages. The former uses the entire the sample, while the latter only uses as data the subset of people who are already employed.

Omitting the log is suitable in this context, where the dependent variable (income level) is not sperad too far apart for this subsamplel of refugees. However, to be more rigorous, a more suitable way would be to use the Heckman selection procedure, where the continuous outcome wages are conditional on an individual being employed. I plan on doing this in my Master's dissertation which this paper will feed into.

Using this specification, we see that language training has a significant effect across all three estimations. The effect is largest for the comparison between the group that was never treated and the group that was treated during the survey. While significant, the increase of 137€/month is not very large in magnitude.

Dependent Variable:	Income Level (Gross)		
Model:	(1) Full	(2) Never + During	(3) Before + During
Variables			
Treatment	107.5***	137.5***	37.35**
	(16.56)	(16.59)	(18.77)
Age Dummies			
Fixed-effects			
pid	Yes	Yes	Yes
syear	Yes	Yes	Yes
Fit statistics			
Observations	25,450	20,229	16,626
R^2	0.67807	0.63587	0.65877
Within R^2	0.01693	0.02030	0.01475

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 4: Caption