

Dahlberg, Matz; Egebark, Johan; Vikman, Ulrika

Working Paper

Long-run integration of refugees: RCT evidence from a Swedish early intervention program

Working Paper, No. 2023:23

Provided in Cooperation with:

IFAU - Institute for Evaluation of Labour Market and Education Policy, Uppsala

Suggested Citation: Dahlberg, Matz; Egebark, Johan; Vikman, Ulrika (2023) : Long-run integration of refugees: RCT evidence from a Swedish early intervention program, Working Paper, No. 2023:23, Institute for Evaluation of Labour Market and Education Policy (IFAU), Uppsala

This Version is available at:

<https://hdl.handle.net/10419/297013>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Long-run integration of refugees: RCT evidence from a Swedish early intervention program

Matz Dahlberg
Johan Egebark
Ulrika Vikman

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Long-run integration of refugees: RCT evidence from a Swedish early intervention program^a

by

Matz Dahlberg^b, Johan Egebark^c and Ulrika Vikman^d

December 05, 2023

Abstract

This study uses a randomized control trial (RCT) to evaluate a new program for increased labor market integration of refugees. The program has immediate and substantial short-run effects on employment, corresponding to around 15 percentage points. The effect lasts for three years but eventually fades out, as the control group catches up and reaches the long-run employment level of about 50 percent. We show that the program boosts language skills in the short run, and that this channel explains an increasing share of the effect on employment. Using survey data, we finally measure if the program affects integration in other dimensions, such as psychological, social, political, and navigational integration. Our findings suggest that faster labor market integration in the short run does not lead to increased general integration in the long run.

Keywords: Refugee immigration; Multidimensional integration; Randomized control trial; Field experiment; Labor market program; Employment

JEL-codes: C93, J08, J15

^a This work has benefitted from comments from Olof Åslund, Spencer Bastani, Maria Hemström Hemmingsson, Göran Holmqvist, and participants at seminars and workshops at IFAU and Uppsala University. We are very grateful to Peter de Jonge at the Swedish Public Employment Service in Gothenburg for helping out with the survey interviews. Ulrika Vikman acknowledges financial support from FORTE (Reg. No. 2016-07123).

^b Uppsala University, CESifo, and IFAU. matz.dahlberg@ibf.uu.se.

^c Swedish PES and UCLS. johan.hartvig-egbark@arbetsformedlingen.se.

^d IFAU, UCLS and UCFS. ulrika.vikman@ifau.uu.se

Table of contents

1	Introduction	3
2	The early intervention program and the RCT design	4
2.1	Immediate and intensive assistance	4
2.2	The RCT design	5
2.3	Comparing the program to the baseline services at the PES	6
3	Effects of the early intervention program on labor market integration	8
3.1	Estimation methods	8
3.2	Data	8
3.3	Effects on employment	9
3.4	Effects of actual program participation	17
3.5	Comparing employment rates to the rest of Sweden	19
3.6	Effects on language skills and the mediating role of language	21
4	Long-run effects on social and economic integration	25
4.1	Data collection	25
4.2	Results	27
4.3	How should we understand the survey results?	30
5	Marginal value of public funds (MVPF)	31
6	Conclusion	33
A	Appendix	38
A.1	Additional results	38
A.2	Sensitivity analysis: Definition of employment	47
A.3	The questions in the survey	51

1 Introduction

In Dahlberg et al. (2023) we designed a randomized control trial (RCT) to study the short-run impact of a new and ambitious integration program in Sweden. We found that the program had positive, and substantial, short-run (one year after treatment) effects on refugees' labor market outcomes. It raised employment during the first year by 15–20 percentage point, implying that the job-finding rate in the treatment group almost doubled. Our conclusion was that early and intensive assistance, targeted at low-educated and newly arrived refugees, is an effective way to improve labor market integration.

In this follow-up paper, we examine the long-run effects of the program. Since the literature so far have focused on documenting the short-run impacts of integration policies, a key question that needs to be answered is whether the estimated positive effects in previous studies are only temporary or whether they last more permanently (see discussion in Foged et al., 2022a). In this new study, we seek to start filling this gap in the literature. We contribute to the literature in two specific ways. First, we study if the large impact of the program on employment in the short run is persistent. To this end we use improved measures of labor market attachment, in combination with a significantly longer time-horizon where we follow treatment and control individuals for up to four years after the treatment ends.

Second, we study the effect of the program on multidimensional integration. It is sometimes assumed, in the public debate, that faster labor market integration of refugees in the shorter term will lead to increased general integration in the longer run. One of the main aims of our study is to provide insights on whether this is correct or not. We use survey data to capture the causal impact of the program on psychological, social, political, linguistic, and navigational integration. The questions used in the survey was first developed by Harder et al. (2018) with the goal to provide scholars with a survey-based general measure of immigrant integration. Since we conduct the survey more than three years after the end of the program, we are able to capture the link between early labor market entries and refugee integration in other dimensions.

The evaluation is based on a field experiment where 140 low educated and newly arrived refugees in the City of Gothenburg in Sweden were randomly assigned to a treatment and control group (with a 50 percent share in each group). While the individuals in the treatment group were invited to participate in the new integration program, and 63 percent accepted, the non-compliance in the treatment group and the control group attended the baseline services at the Swedish Public Employment Service (PES). The program introduces highly intensive assistance shortly after the residence permit is granted, and combines three major components: language training, work practice, and job search assistance. The (full time) language training starts immediately after the refugees enroll in the program and lasts for three months. This is followed by six months of supervised work practice. Finally, after the work practice ends, the participants receive help from professional caseworkers with finding a job.

We reach three main conclusions. First, we find that the early intervention program has long-lasting effects on employment. It takes almost four years for the control group in our experiment to reach the long-run employment share of around 45 percent. In sharp contrast, the treated refugees, as a result of the intervention, reaches this employment share directly after exiting the program. Our

findings are well in line with recent research showing that the initial conditions that immigrants face in their host countries have important and long-lasting impacts on their future economic and social integration (Åslund and Rooth, 2007; Braun and Dwenger, 2020; Bansak et al., 2018; Aksoy et al., 2023; Fasani et al., 2021; Bratu et al., 2021). We show that men benefit more from the program than women. While men display treatment effects that are large and stable in the long run, the impact for women, while being large in magnitude just after the program, vanishes quickly. Finally, we show that the control group in our experiment experiences the same employment growth as newly arrived low-skilled refugees in the rest of Sweden, indicating that our results have strong external validity.

Second, we present convincing evidence that the program boosts language skills in the short run, adding to a small literature studying whether more language training has an actual impact on host-country language proficiency. Using the decomposition analysis suggested by Heckman et al. (2013), we next show that the increased language skills in the treatment group explain a substantial share of the effect on employment, and that this mechanism grows in importance over time – explaining around 40 percent of the employment effect up to three years after the end of the program.

Third, using the multidimensional integration index developed by Harder et al. (2018) we find little evidence of any impact of the program on general integration. We find that the (male) participants in our program perceive themselves as more politically integrated than those who did not participate, but for the other dimensions there are no statistically significant differences. One potential reason for the small differences between the treated and the control group is that the refugees in our experiment (both treated and non-treated individuals) perceive themselves as surprisingly well integrated. Overall, these findings indicate that faster labor market integration in the short run may not automatically result in broader (self-perceived) integration in the longer run.

The structure of the paper is as follows: In the next section (Section 2), we describe the early intervention program and the design of the experiment, which generates the sample of individuals to be used in the two sets of analyses. In the first set of analyses, we examine the long-run effects on labor market outcomes (Section 3). In the second set of analyses, we study the long-run effects on integration in the other dimensions (Section 4). Finally, in Section 5 we compare individual gains of the program with the public net costs, before we conclude in Section 6.

2 The early intervention program and the RCT design

2.1 Immediate and intensive assistance

This study evaluates the long-term impact of a new and ambitious integration program that was developed and implemented in the city of Gothenburg in Sweden.¹ The program introduced highly intensive assistance immediately after the residence permit is granted, and combined three major components: language training, work practice, and job search assistance. The language training started immediately after the refugees entered the program, and lasted for three months. This was followed by six months of supervised work practice. After nine months of language training and work practice, professional PES caseworkers helped the participants searching for jobs.

¹ See Dahlberg et al., 2023 for a detailed description of the program.

Language training

All newly arrived refugees who register as unemployed at the Swedish PES are offered Swedish language classes, so-called Swedish for Immigrants (SFI).² The SFI typically (in the baseline services at the PES) amounts to 15 hours of teaching per week, throughout a two-year integration period. One of the aims of the new integration program was to introduce more intensive language training at an early stage. Hence, all refugees who entered the program were expected to take additional hours of SFI classes. Moreover, in addition to the extended SFI training, the new integration program included other more general courses as well, which were given in Swedish.³ The extended SFI language training together with the additional general courses added up to (close to) 40 hours of classroom training per week.

Work practice

After the initial intensive language training, the participants continue to the next stage as they start work practice. The work practice consisted of three days a week for six months (with the remaining two days per week dedicated to SFI and additional general courses). Each participant was appointed 1–3 supervisors, who had the main responsibility for the intern. When allocating the participants to a workplace, two aspects were considered. First, in order to challenge the participants to practice Swedish, everyone was assigned supervisors who did not speak their native language. Second, to get the participants used to commuting, they were assigned a workplace outside their own residential area.

Job search assistance

Finally, as the work practice ended, efforts began in terms of finding potential employers. The job search assistance, which was performed by professional caseworkers at the PES, focused on helping the refugees find suitable vacant jobs matching their level of skills, and helping with applications, CV-writing, interview training etc.

2.2 The RCT design

In total, four waves of refugees have started the program, one per year during 2016–20.⁴ Each wave accepted the same number of participants, around 50 individuals. Our study follows the wave where treatment was randomized, the one starting in 2017.

The eligibility criterion that was used in the program is comparably strict. First, participation required recently having received a Swedish residence permit. Second, only those with less than high school education (from the country of origin) were eligible. Based on these criteria, the PES in Gothenburg identified 140 potential participants in April 2017 (note that this constituted the

²Swedish for Immigrants consist of three different tracks. Track 1 includes four levels of courses (A–D) and is offered to illiterates. Track 2 includes three levels (B–D). Track 3 includes two levels (C and D) and is offered to those with some higher education.

³This includes civic orientation classes, one course teaching workplace rules, one course teaching how to be service-minded, and introductory courses in IT and mathematics. All extra courses were given in Swedish and they ran throughout the whole program period, in parallel with the other activities.

⁴The different waves started in October 2016, May 2017, April 2018 and January 2020, respectively.

universe of eligible individuals in Gothenburg at the time of the start of the program). Out of the 140 potential participants, 70 individuals were randomly drawn to a treatment group and 70 were randomly drawn to a control group. Since previous studies show that the effects of active labor market programs differ between men and women (Card et al., 2018), we block-randomized based on gender to get better precision. This makes the share of men and women in the treatment group identical to the shares in the target population. The randomization was carried out at the end of April 2017. In the analysis below, we therefore denote April 2017 as month zero.

All 70 individuals in the treatment group were summoned to an information meeting, which took place at the PES in Gothenburg at the end of April 2017 (less than one week after the randomization). All 70 individuals who were summoned showed up at the meeting, which means that everyone in the treatment group were given a detailed description of the program by the representatives from the three founders of the program. At the end of the information meeting everyone in the treatment group were offered to participate. 44 individuals accepted the offer and started the program.⁵ In the analysis, we denote May 2017 as the first month of the program, since the individuals in the treatment group may react already based on the information provided during this month. Note that all 70 individuals in the control group, and all the individuals in the treatment group who turned down the offer, continued within the baseline services provided by the PES in Gothenburg.

2.3 Comparing the program to the baseline services at the PES

Figure 1 compares activities across treated and non-treated individuals during 12 months after the randomization. The purpose here is to show how the objectives of the program, as described above, match actual realizations as observed in the data. We follow the two groups both before and after the point when we did the randomization (which we denote by zero on the horizontal axis). Note here that treatment starts in the second month, as the first month was dedicated to preparations.

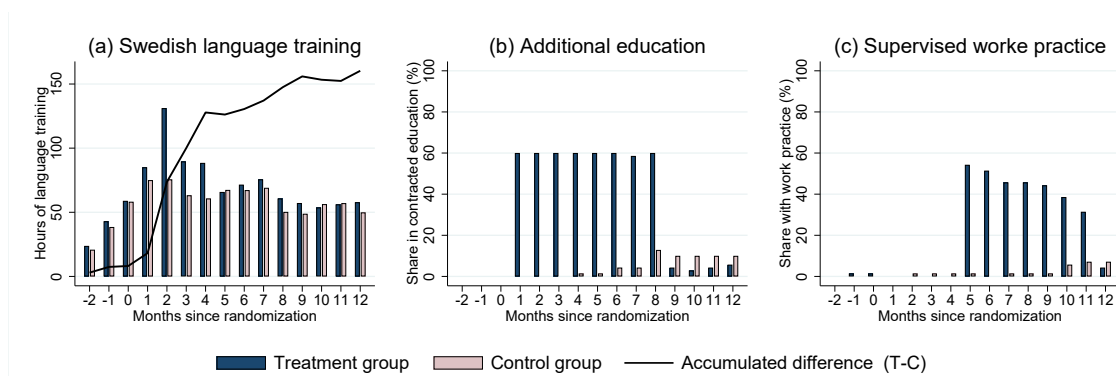


Figure 1: Timing and intensity of treatment.

Notes: The figure shows the extent to which treated and control individuals participate in different activities during the first year of the program.

Source: Dahlberg et al. (2023).

⁵More men than women accepted to participate: 30 out of 43 invited men accepted, compared to 14 out of 27 invited women.

Figures 1 (a) and (b) document differences in terms of language training. We first conclude, from Figure 1 (a), that the treatment group received substantially more Swedish training at the beginning of the program period. The difference corresponds to roughly one extra week (30–50 hours) of training per month, which means the program added one month (160 hours) of extra Swedish classes (green/solid line). Next, in Figure 1 (b), we show that, in contrast to the control group, the treated individuals participated in additional, more general, courses as well. Since these courses were taught in Swedish they also add language training. Figure 1 (c) shows what happened during the six months that followed. While there is almost no (or at least very little) supervised work practice in the control group, treated individuals exhibit a large increase just after the initial full-time language training has ended. We also note that the difference between treated and non-treated individuals drops to zero after one year, which is expected given that the work practice is not supposed to last more than six months. To sum up, there is convincing evidence that the treatment group received substantially more assistance, in terms of language training and work practice, relative to the baseline. The intensified assistance starts just after the participants have entered the program, and hence constitutes an immediate injection of additional help.

The last stage of the program contains job search assistance. We cannot provide direct evidence that the treatment group received more support relative to the baseline as this activity is not registered in our data. However, we have verified that the treatment group displays a sharp increase in the number of contacts with PES caseworkers just after the work practice ended, and since there is no such increase for the control group we believe that the program added intensity at this final stage as well.

While we have presented a detailed description of the new integration program that we evaluate, we know less about the non-treated individuals. As we saw above, they got less language training, and virtually no one in the control group participated in the type of work practice offered to the treatment group. However, we want to rule out that the control group were exposed to other types of activities with a similar program intensity as the treatment group. In Dahlberg et al. (2023) we show that only a small fraction of the control group participated in intensive activities during the year when the program ran. There was a minimal increase from 1.43 percent in month 3 to 4.29 percent in month 9. This is in sharp contrast to what we observe in the treatment group. Instead, non-treated individuals ended up in low-intensive activities, such as own job search or different kinds of preparatory measures.

Finally, in Dahlberg et al. (2023) we conclude that there are no differences across treated and non-treated individuals in terms of economic incentives. The individuals that participate in the new early intervention program that we evaluate were not paid any additional money – they got the same cash transfer, a SEK 7,000 benefit, from the government as the individuals in the control group. Hence, all individuals in our sample face the same (strong) incentive to exit unemployment as fast as possible.

3 Effects of the early intervention program on labor market integration⁶

As highlighted in Foged et al. (2022a), the previous literature is silent on the question whether integration policies have permanent effects. In this study we follow the targeted refugees for five years, which means that we provide novel and important insights about the long-lasting impact of integration programs. After explaining our estimation methods in section 3.1 and providing details on the data and outcome definitions in section 3.2, we proceed to present the results related to labor market integration. These results can be found in sections 3.3 through 3.6 below.

3.1 Estimation methods

We measure successful labor market integration via (stable) employment. When evaluating the employment effects of the new program we are interested in two different types of effect estimates: the effect of being *offered* the possibility to participate in the program and the effect of *actually* participating in the program. For the first type of estimates, we estimate intention-to-treat (ITT) effects (which in this case is equal to the average treatment effect of being offered the possibility to participate in the program). Since the procedure with summons, information meetings, and invitations to participate in different programs is the procedure typically used by the Swedish PES, we argue that the ITT analyses produce the most policy-relevant effect estimates. The ITT-results are obtained by running a linear OLS-model controlling for gender (since we block-randomized on that variable).⁶

Even though the ITT-estimates might be the most policy-relevant estimates, it is still of interest for policy-makers to know if, and to what extent, *actual participation* in the program is beneficial for the participants or not. For the second type of estimates we therefore estimate the *effects of actually participating* in the program. To obtain the average treatment effects on the treated (ATT), we instrument actual participation with treatment status. Since the IV-estimates are given by scaling the ITT-estimates with the compliance rate, we get effects from *actual* participation that are larger than the effects of *being offered* to participate with a factor of 70/44.⁷

3.2 Data

Our study of employment effects benefits from comprehensive administrative data collected by the Swedish Public Employment Service, Statistics Sweden, the City of Gothenburg, and the Swedish National Board of Health and Welfare. The data cover the period 2017-2021. The PES records contain daily information on each job seeker's unemployment status, including information on active

⁶This means that we rely on asymptotic properties for inference. One potential issue with this in our case is that we have an original sample of 140 individuals. Since the assignment mechanism for allocating individuals into treatment and control groups in our stratified randomized experiment is known and controlled by us, we can however also apply Fisher's approach for calculating exact p-values. Basing the inference on exact p-values instead of relying on the asymptotic properties of the estimators is an alternative approach that we use in a sensitivity analysis, see Appendix table A2. Since the two approaches provide very similar results, we use inference based on asymptotic properties throughout the main text.

⁷Since we have a case of one-sided compliance (i.e., while those that were randomly assigned to the control group cannot participate in the new program, those that were randomly assigned to the treatment group can choose if they want to participate or not), the local average treatment effects (LATE) equals the ATT.

labor market programs, part-time and temporary work, and all types of subsidized employment. In addition, the registers contain individual characteristics, such as, e.g., age, gender, education, country of origin, date of entry to Sweden, and treatment status. The records from Statistics Sweden contain monthly information on each individual's labor income.⁸ From the City of Gothenburg we have detailed information on education administered by the municipality. First, we have information on the number of hours that a refugee participates in SFI, and we have information on the grades from these classes. Second, we get participation in the extra education mentioned in figure 2 (b). Finally, the records held by the Swedish National Board of Health and Welfare contain information on individual level social assistance (SA) payments per month. We use the SA payments in our cost-benefit analysis.

To construct outcome measures, we link the PES records to the records from Statistics Sweden. Our employment outcome is defined as follows. An individual is employed in a given month if he or she (1) has positive labor income that month, and (2) is registered as either having non-subsidized work or New Start Job (NSJ) that month.⁹ The information on labor income comes from the records held by Statistics Sweden, whereas the information on the type of employment comes from the PES registers. We are able to follow our target group until 2021, almost 4 years after the end of the program.

In Dahlberg et al. (2023) we use observable characteristics to show that the treatment- and control groups are identical on average, which means that the randomization worked as expected. Moreover, we show that 90 percent of the individuals in our sample have no formal education above elementary school (from their country of origin), and that the average number of days since the residence permit was granted is 220–230. This confirms that we target individuals with a very weak initial position in the labor market (see also Figure 9).

3.3 Effects on employment

Mean outcomes and Intention-to-Treat (ITT) effects

Figure 2 provides the month-by-month results by estimating the effects on employment; while the left panel provides the raw (mean) outcomes, the right panel provides the estimated ITT effects (i.e., the change in the probability of being employed in a given month from being offered to participate in the labor market program).

First, it can be noted that soon after the end of the program (in April 2018), a large share of the individuals in the treatment group (more than 40 percent) are employed (see the left panel). The corresponding figure for the individuals in the control group is closer to 30 percent. The share employed in the treatment group hovers around 45–50 percent over the full follow-up horizon (up until December 2021, almost four years after the end of the program), indicating that the employment level for this group might be at a stable long-run level. The long-run trend for the individuals

⁸Since 2019, the Swedish Tax Authority collects information on labor income on a monthly basis. Note that we also have information on income from self-employment.

⁹As described in Dahlberg et al. (2023), all individuals in our sample are eligible for New Start Job, but they need to find an employer willing to employ them. With this subsidy, the employer has total wage costs of 15–16,000 SEK per month which roughly corresponds to the average minimum wage level in the U.S. The subsidy is given for one year at a time for a maximum of two years and is typically used for job seekers who have been unemployed for more than 6 months.

in the control group hovers at a somewhat lower level, around 30 percent in the first year after the program and between approximately 30 and 40 percent in the next two years, landing at 45-50 percent at the end of the period.

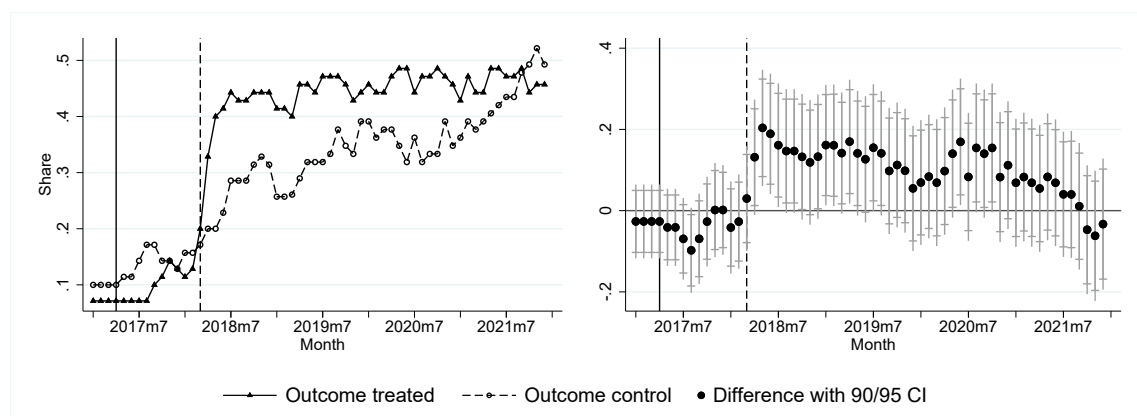


Figure 2: Mean outcomes and ITT estimates for employment

Note: Employment is defined as having positive earnings from non-subsidised work or a New Start Job during a month. Outcomes in the left figure and ITT effects in right figure with 90/95 confidence intervals from OLS estimations controlling for sex. See Appendix Table A1 for the exact employment levels, point estimates and p -values.

The results in the left panel are interesting for several reasons. First, even with the new and enhanced data, the results are similar (albeit showing an even larger share employed) to the short run-effects found in Dahlberg et al. (2023). Second, it is encouraging to see that the initial positive effects from the program on employment not only lasts for several years (and that it seems to stabilize at a relatively high level), but that it is also stable over the Covid-19 pandemic shock in 2020.¹⁰ Third, it is interesting to see that the share employed in the control group also increases over time (with a slight dip when the pandemic hits), closing the gap to the individuals in the treatment group during the last months of the follow-up period. An important insight from these results is hence that the employment shares in both groups seem to stabilize around 50 percent in the long run, but thanks to the new program the treated refugees reached the long-run employment level almost immediately after the end of the program – four years earlier than both the control group and a similarly defined comparison group for the rest of Sweden; c.f. Figure 9.

Turning to the ITT-estimates, the right panel in Figure 2 shows that the estimated treatment effects hover between 10 and 20 percentage points for the first three to three and a half years after the end of the program. This must be considered a large treatment effect given the average employment rate in the control group. The estimated treatment effects are significantly different from zero in many, but not all, months during these first years after the end of the program. From mid-2021 and onward, the point estimates get closer to zero and becomes clearly insignificant. What is interesting, and important, to note here is that when the estimated treatment effect becomes

¹⁰During the pandemic year 2020, the overall employment rate decreased for the first time since 2010, and the decrease was largest among foreign-born individuals (the employment rate decreased with 1,9 percentage points for this group, compared to a decrease of 0.8 percentage points for native-born individuals. According to Statistics Sweden, the larger decrease for foreign-born is probably due to the fact that they to a larger extent were employed in sectors that were more negatively hit by the pandemic (Public Statistics of Sweden (Sveriges Officiella Statistik, 2019).

smaller and/or insignificant during the follow-up period, it is due to the fact that the employment rate in the control group increases, not that the employment rate in the treatment group decreases.

One way to reduce the uncertainty in the month-by-month estimates in Figure 2 could be to estimate employment effects over more aggregated time periods. For this reason, we have estimated the number of months each individual has been employed in each quarter from the start of the program until the last quarter of 2021 (remember that the program ended at the end of the first quarter in 2018).¹¹ The ITT-estimates, presented in the right panel in Figure 3, generally confirm the main results in Figure 2, but with higher precision: there is an immediate and positive effect after the end of the program (the individuals in the treatment group are employed half a month more than the individuals in the control group already in the second quarter of 2018), the estimated effects are stable for a long time (there is a tendency of a catching-up effect until the pandemic hits in the second quarter of 2020, when the point estimate once again increases and becomes statistically significant), and, then, the individuals in the control group catch up and eventually close the gap by the end of the period. The magnitude of the estimated effects during the first three years after the end of the program are large relative to the average number of months employed in the control group (the individuals in the control group are on average employed approximately 0.5-1 months per quarter over this time period; see the left panel in Figure 3).

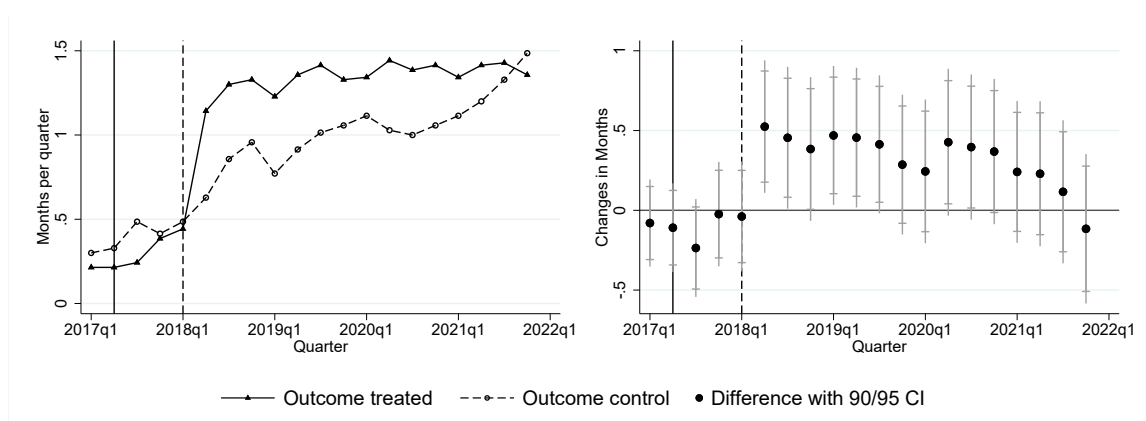


Figure 3: Effects on months employed per quarter

Note: Employment is defined as having positive earnings from non-subsidised work or a New Start Job during a month. Outcomes in the left figure and ITT effects 90/95 confidence intervals in right figure from OLS estimations controlling for sex. See Appendix Table A2 for the exact employment levels, point estimates and *p*-values, both based on asymptotic approximations and on Fisher’s exact tests.

Causal heterogeneity analysis: men vs. women

Given that we block-randomize based on gender in our experimental design, we are able to estimate separate causal effects for men and women. Even though we get fairly small sets of observations, it is still of interest to examine if the pattern is similar across the two groups. Conducting sepa-

¹¹We have defined employment as any positive labor earnings. As a sensitivity check, we also consider defining employment as labor earnings above a specific threshold. Our results remain consistent, as discussed in Section A.2 of the Appendix, where we also examine earnings effects and control for baseline characteristics.

rate analyses for men and women, it is clear that the positive treatment effects found in Figures 2 and 3 to a large extent are driven by male participants (c.f. the top panel in Figure 4, where the blue/triangle lines provide the employment rates for the male participants and the red/circle lines provide the employment rates for the female participants). The average number of months employed per quarter and the estimated treatment effect is on average somewhat larger than the baseline estimates (c.f. the lower left panel in Figure 4).¹² For women, both the number of months employed per quarter and the estimated treatment effects are lower than for men, and there are more uncertainty in the point estimates (c.f. the lower right panel in Figure 4).¹³ When examining how many months per quarter the male participants are employed, we get a very similar pattern as in the baseline analysis: for most of the follow-up period, the males in the treatment group are employed approximately half a month more per quarter than the males in the control group; these point estimates have a surprisingly good precision given the fairly small sample size.

¹²It is also worth noting that the average employment rate in the male treatment group stays around 60 percent for almost the whole period after the end of the program; see the upper panel in Appendix Figure A1.

¹³The same is true for the employment rates by month for women; see Appendix Figure A1.

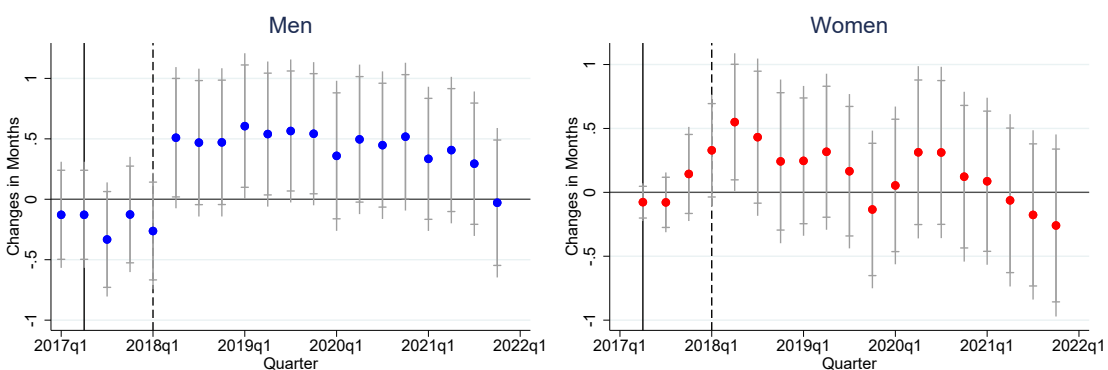
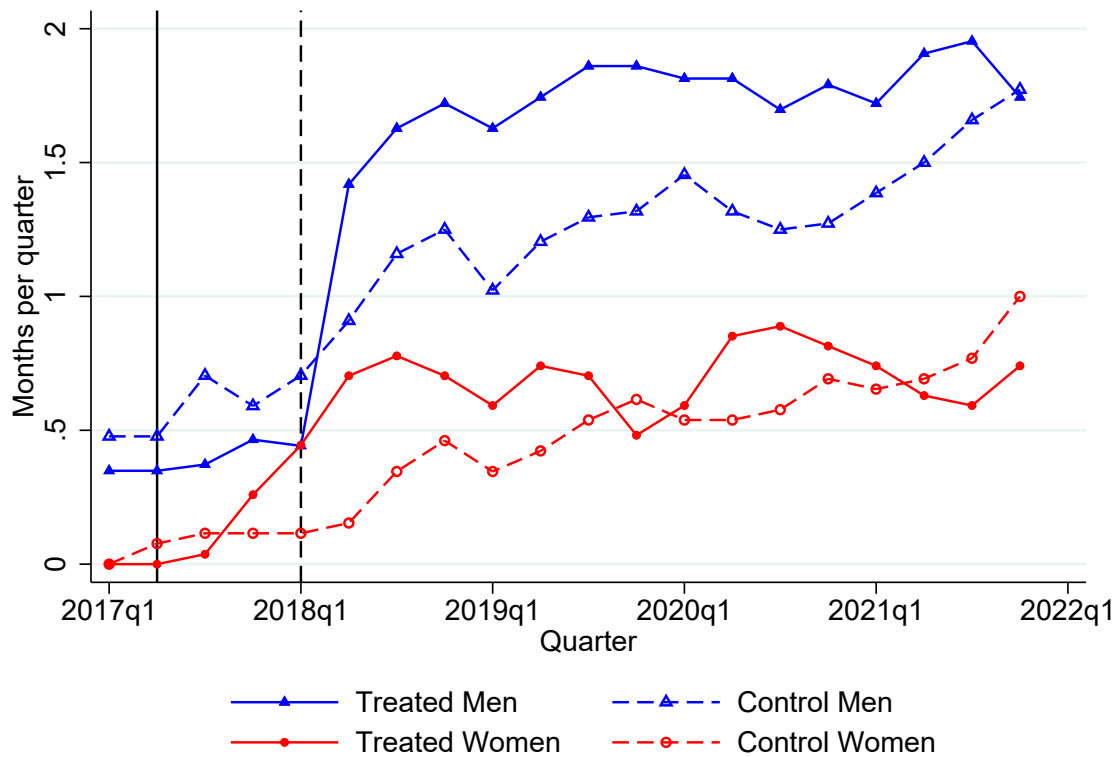


Figure 4: Effects on months employed per quarter, by sex

Note: Employment is defined as having positive earnings from non-subsided work or a New Start Job during a month. Outcomes in upper figure and ITT effects with 90/95 confidence intervals in bottom figures from OLS estimation separated by sex. See Appendix Table A3 for the exact employment levels, point estimates and p -values.

Descriptive heterogeneity analysis: The role of age and level of education

From earlier literature we know that age and level of education matter for how fast and how well immigrants integrate into the labor market, and it has been shown that older and less educated individuals have a harder time to integrate. For these reasons, we are interested in examining if and how the new labor market program that we are evaluating correlates with these type of variables. One hypothesis is that the intensive language training and the ambitious, supervised, work practice can be extra beneficial for (relatively) older and (relatively) less educated individuals¹⁴.

When plotting the average number of months per quarter by young/old (where the cut-off age is 34; see Figure 5) and by low-/high-educated (where the cut-off is nine years of schooling; see Figure 6), interesting descriptive patterns emerge. Starting with age, there are indications that the new program was very beneficial for the older individuals in the sample (older than 34 years of age). There is a substantial and clear difference in employment levels across the treatment and control group during almost the whole follow-up period. The older individuals in the treatment group are on average employed one month per quarter over the whole follow-up period, whereas the older individuals in the control group are employed less than half a month per quarter (c.f. the red/circle lines in Figure 5). It is only by the very end (three years after the end of the program) that the older individuals in the control group catch up with those in the treatment group. Even though the figure makes clear that younger individuals have an easier time integrating into the labor market, it is also clear that the program is less beneficial for this group; the employment rates are very similar for treated and untreated young individuals during the entire follow-up horizon (c.f. the blue/triangle lines in Figure 5).

Turning to the level of education, there are indications that the new program was very beneficial for those with very low or no education (i.e. less than nine years of education). The program seems to have the same large impact for low educated as for older individuals; the low-educated individuals in the treatment group are on average employed 1-1.5 months per quarter over the follow-up period, whereas the same figure for the low-educated individuals in the control group is 0.5-1 (c.f. the red/circle lines in Figure 6). The employment rates for the low-educated individuals in the treatment group is even close to that of the high-educated for a long time during the follow-up period. The program is less beneficial for those with at least nine years of schooling; the employment rates are fairly similar for treated and untreated individuals with relatively higher education for the entire follow-up horizon (c.f. the blue/triangle lines in Figure 6).

¹⁴Remember that we have a sample of individuals where everybody has a low level of education, implying that the relatively low-educated in our sample has very low or no education, and even includes illiterates.

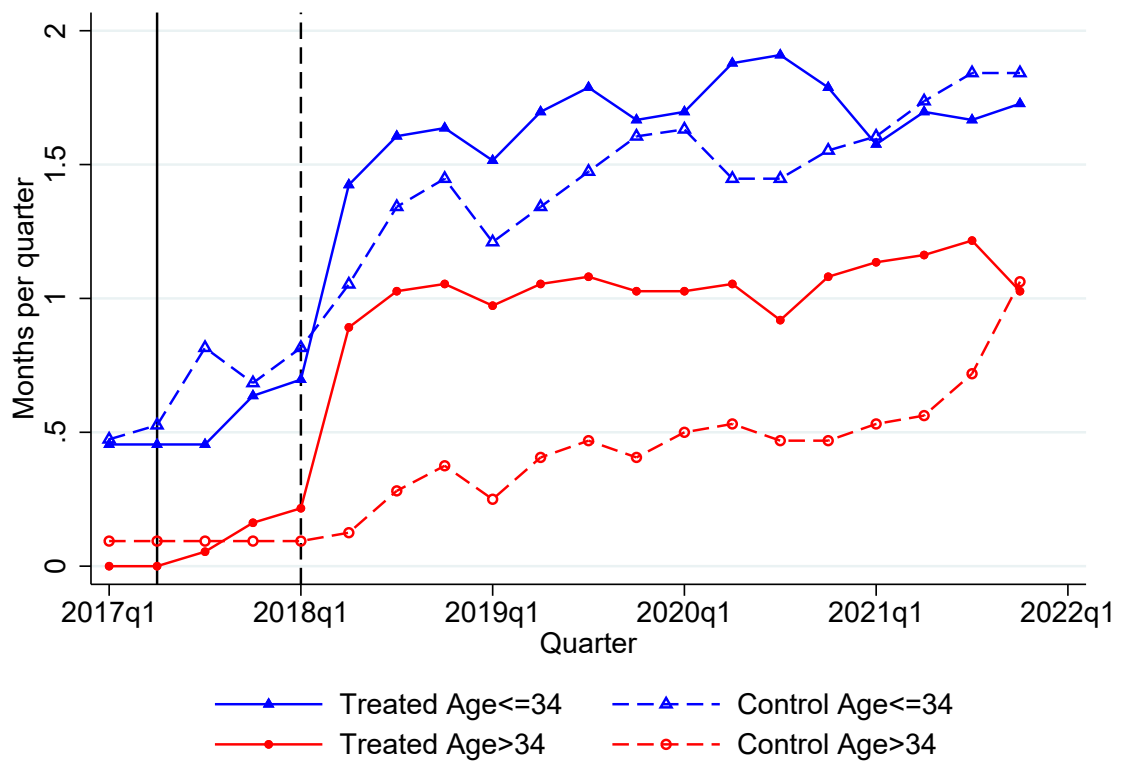


Figure 5: Months employed per quarter, by age

Note: Employment is defined as having positive earnings from non-subsidised work or New Start Job during a month.

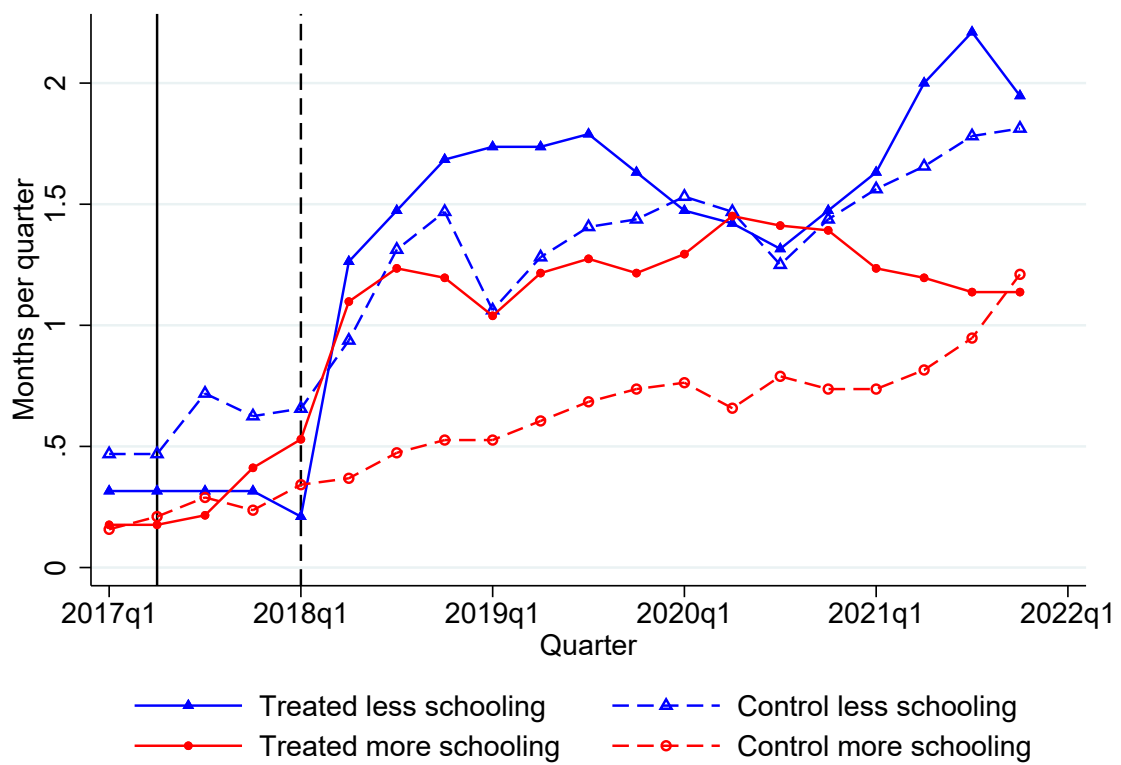


Figure 6: Months employed per quarter, by education

Note: Employment is defined as having positive earnings from non-subsided work or a New Start Job during a month. Less schooling is defined as less than nine years and more schooling is defined as nine or more years of education.

3.4 Effects of actual program participation

Even though we consider the ITT-estimates as the most policy-relevant estimates, it is still relevant to study the magnitude of the *effects of actually participating* in the program. To obtain the average treatment effects on the treated (ATT), we instrument actual participation with treatment status.¹⁵ As is clear from Figure 7, actual participation in the program has a positive effect on our output measures. For those who participated in the program employment increased by between 0.5 and 1 month per quarter, and this effect lasted for up to three years after the end of the program. Once again it is important to stress that when the point estimates tend towards zero, it is because the control group catches up with the treatment group, not because employment in the treatment group decreases. The employment rate among those that actually attended the program is around 60% (not shown), which is a *very* high level for this low-educated group (c.f. from the figures for Sweden as a whole in Figure 9). Actual participation seems to be especially beneficial for the male participants; they are employed almost one month more per quarter compared to men in the control group, and this difference seems to be stable for a long time after the end of the program (see the left bottom panel in Figure 8).

¹⁵As described earlier, the IV-estimates are given by scaling the ITT-estimates by 70/44. The first stage is given by the first columns in Table A4 in Appendix.

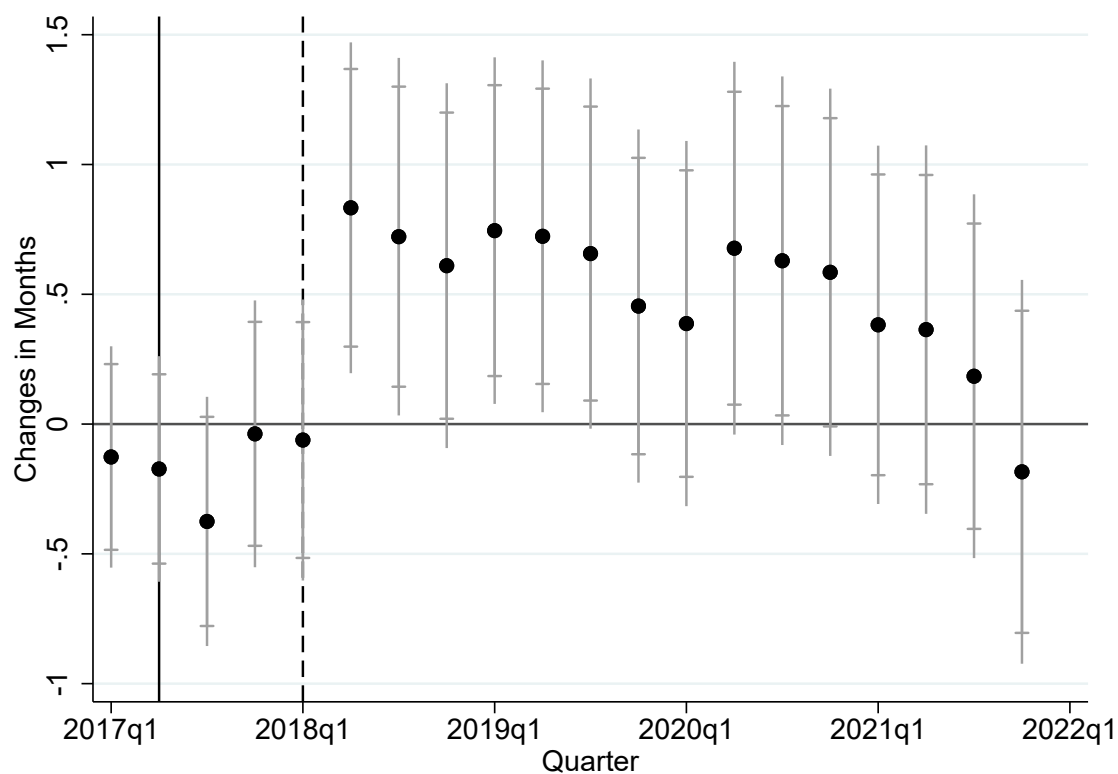


Figure 7: IV estimates on months employed by quarter

Note: Employment is defined as having positive earnings from non-subsidised work or a New Start Job during a month. Estimated IV effects controlling for sex with 90/95 confidence intervals are shown in figure. See Table A4 for estimation on first stage of treatment on participating and Table A5 for the exact employment levels, point estimates and *p*-values, in Appendix.

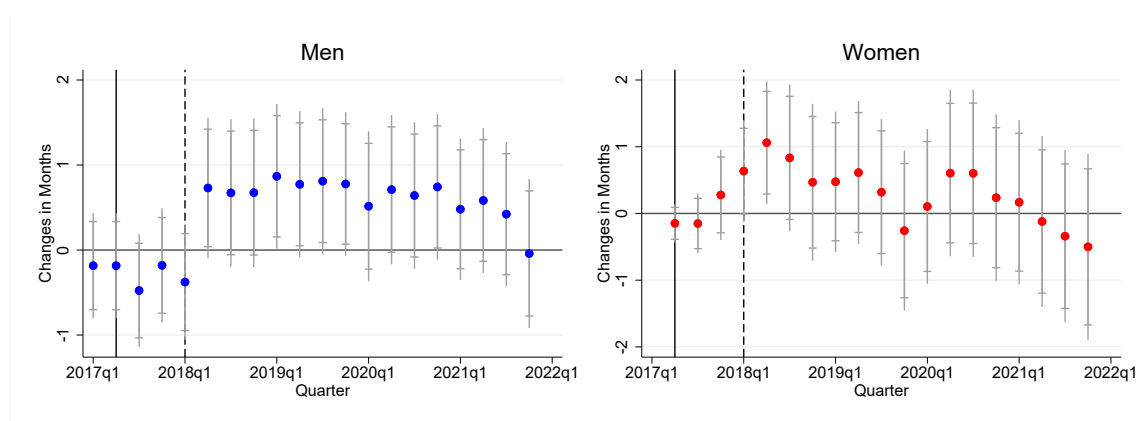


Figure 8: IV estimates on months employed by quarter, by sex

Note: Employment is defined as having positive earnings from non-subsidised work or a New Start Job during a month. Estimated IV effects separated by sex with 90/95 confidence intervals are shown in figures. See Table A4 for estimation on first stage of treatment on participating and Table A5 for the exact employment levels, point estimates and *p*-values, in Appendix.

3.5 Comparing employment rates to the rest of Sweden

How unique is the case we study? To get a sense of that we construct groups of newly arrived, low-educated refugees in other parts of Sweden. Our goal here is to use the eligibility criteria that was used in the program to create identical groups of refugees that we can use for comparisons.¹⁶ Figure 9 shows employment for the newly arrived refugees in our experiment in Gothenburg (black lines; solid with circles for the treatment group and dotted with circles for the control group) and for identical refugees (newly arrived and low-educated) in the groups we compare with: the average employment rates are displayed, separately, for the two other large cities in Sweden, Stockholm (red line with diamonds) and Malmö (green line with triangles), and the rest of Sweden (excluding the three big cities; blue line with squares).

A few things can be noted from Figure 9. First, the employment rates are rather similar across all regions just before and during the year that our program runs (the solid vertical line indicates the start of the program and the dotted vertical line the end of the program). This indicates that the group of refugees that we target in our study is very similar in terms of initial conditions to refugees in other parts of Sweden. One implication of this is that the problematic labor market situation for low-educated refugees in Gothenburg, that spurred the instigation of the new program, is not a unique case, but rather a common situation, indicating that the need for effective integration programs is not isolated to the Gothenburg region.

Second, the treatment effect is clearly visible also compared to the rest of Sweden. After the end of the integration program, the average employment rate in our control group closely follows that of the refugees in Stockholm.¹⁷ Given this similarity in employment trends, it is quite likely that the refugees in Stockholm would have benefited from participating in a similar early intervention program.¹⁸

Finally, since the development of employment between the control group in Gothenburg and the refugees in Stockholm is so similar, this can be used to inform us on the existence of displacement effects in our experiment. One concern could be that employment in our treatment group crowds out employment in the control group. Judging from the employment patterns it seems unlikely that the employment effects that we estimate is explained by worse employment outcomes in the control group.

¹⁶ To get as close as possible to our target group, we use data covering the universe of immigrants in Sweden and apply the following selection restrictions: we select immigrants (i) who are classified as refugees, including those in need of protection (i.e., the type of immigrants we have in our experimental sample), (ii) that are between 23 and 64 years old (which is the age span we have in our sample) and with no education above compulsory school (similar to our sample), (iii) that were registered at the PES but not involved in any subsidized work or high-intensity activities the day when we conducted the randomization (April 22, 2017), (iv) that did not have any work disability, (iv) that received their residence permit at most one year before the randomization date, and (v) that had been registered at PES for at most 250 days.

¹⁷ The average employment rates for the rest of Sweden and, in particular Malmö, are lower during the whole follow-up period. From Figure 2 in Foged et al. (2022a) in their analysis of Danish data, it is also clear that this is not unique for Sweden.

¹⁸ It is quite likely that the other regions in Sweden would have benefited from that as well, but since the counterfactual development is different, this conclusion is less clear.

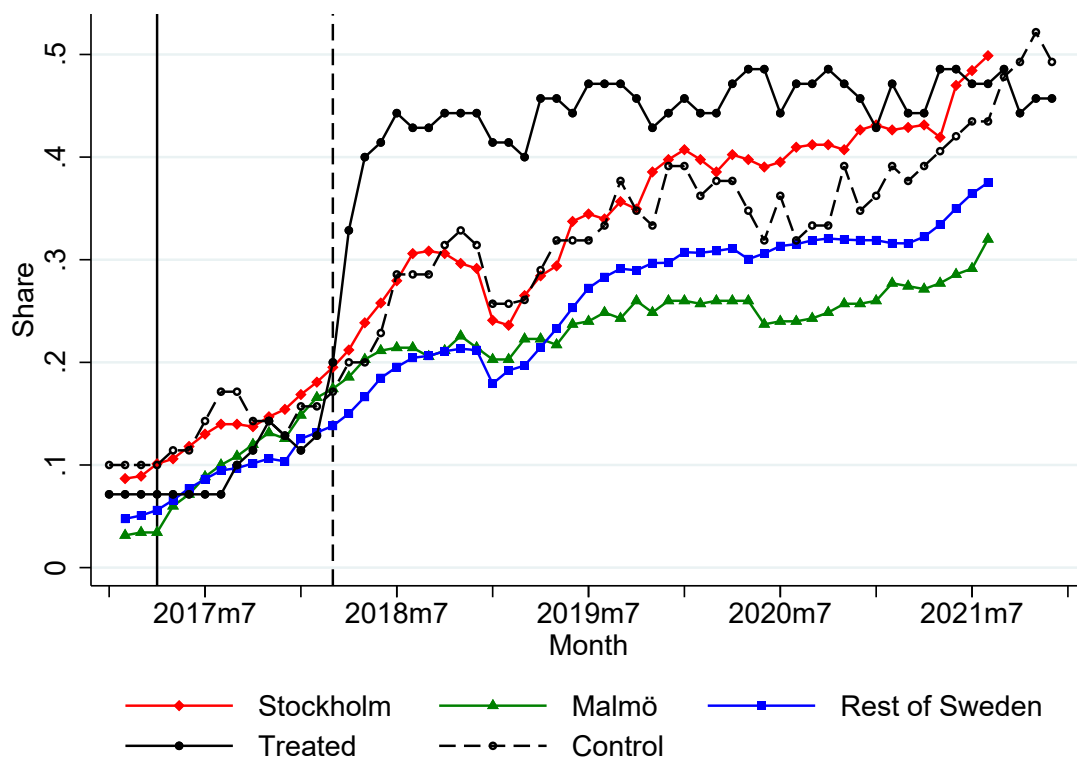


Figure 9: Employment rates compared to similar groups in other parts of Sweden

Note: Employment is defined as having positive earnings from non-subsidised work or a New Start Job during a month. Sample outside of Gothenbourg is defined as explained in footnote 16.

3.6 Effects on language skills and the mediating role of language

One of the main components of the program we are evaluating is the early and intensive language training. In Dahlberg et al. (2023) we examined the impact of this training and used the decomposition analysis suggested by Heckman et al. (2013), to study how much of the estimated effect on employment that was mediated via the improved language skills for the short run outcomes we had access to. In the analysis in this paper, we first exploit the fact that we block-randomized based on gender and estimate the effect for women and men separately. We also examine if the effect varies depending on age and level of education. Second, we use the decomposition analysis to see if the effect of language changes in the long-run.

Effects of the program on language acquisition

Recent studies on immigrant integration have had a special focus on the effects of *language training* (Sarvimäki and Hämäläinen, 2016; Lochmann et al., 2019; Foged et al., 2022b; Heller and Slungaard Mumma, 2023; Pont-Grau et al., 2023). Despite this focus in the literature, it is still not clear to what extent *language proficiency* is important for integration.¹⁹ Since we have access to administrative data on documented language skills, we can shed light on this question, adding to a small literature studying whether more language training has any actual impact on host-country language proficiency (see, e.g., Foged and van der Werf, 2023; Pont-Grau et al., 2023). The data come from municipal records and show all the grades that a recent immigrant gets in the SFI language courses. The SFI grades are set on the basis of pre-determined goals, and whether an individual meets these goals or not is largely decided based on the results on standardized tests.

Figure 10 shows that language training has a strong impact on refugees' documented language skills. The share with a grade from the SFI courses increases sharply in the treatment group relative to the control group following the initial three months of intensive training. By the end of the program, in the beginning of 2018, the share with grades in the treatment group is almost 60 percent, compared to 30–40 percent in the control group. In Dahlberg et al. (2023), we provide estimates of the effect of the program on two measures of language skills, i.e. the probability of having any grade from SFI and the number of grades. Irrespective of which measure we use, we document a large and statistically significant effect of participating in the program on language skills. The probability of having an SFI grade increases by around 27 percent, and the number of grades increases by around 0.4. The two first columns of Table 1 repeats this exercise, and hence provide the same estimates as in Dahlberg et al. (2023) for the full sample.

Since we are interested in potential gender differences we next add separate effect estimates for men and women (see column 3–6 in Table 1). When estimating the effect on language skills by gender, it is clear that the intensive language courses are extra beneficial for men; one year after randomization, close to 70 percent of the men in the treatment group have received a grade in SFI (c.f. the solid blue/triangle line in Figure 11). The corresponding figure for men in the control group is approximately 40 percent (c.f. the dotted blue/triangle line in Figure 11). The language courses are also beneficial for women, but to a lower degree (c.f. the red/circle lines in Figure

¹⁹Language training classes could (in theory) have a positive impact on labor market integration by increasing the participants social skills, network size, cultural understanding, level of job search activity etc. Hence, it is unclear what role language skills have for the transition to employment.

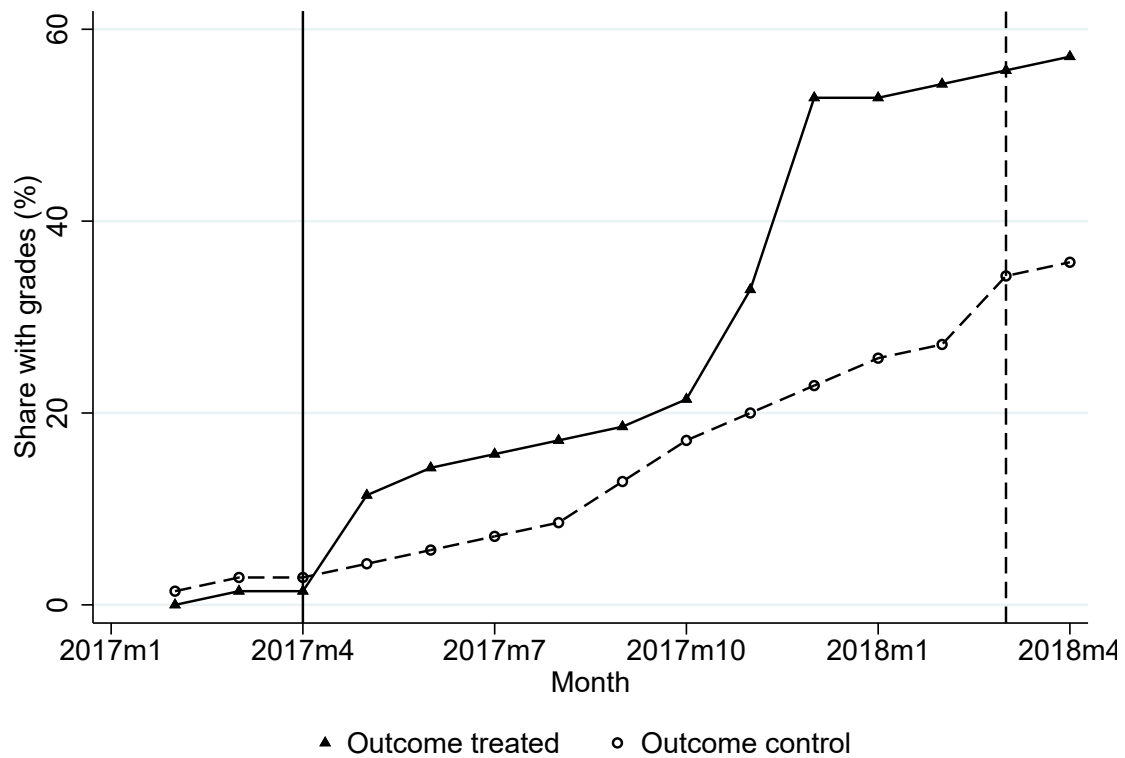


Figure 10: Documented language skills

Note: Based on calculations on data from Swedish for immigrants (SFI) registers over started and completed SFI courses.
Source: Dahlberg et al. (2023)

11). These gender differences are confirmed by the estimates in the last columns of Table 1, which provides estimates of the effect of the program on the probability of having any grade from SFI (column 1 for men and column 3 for women) and the number of grades (column 2 for men and column 4 for women).

Before turning to the mediating role of language, it is of interest to provide descriptive evidence on obtained language skills when dividing the sample along the two dimensions examined in the descriptive heterogeneity analysis. From the left panel in Figure 12, we note that the language courses in the new program seem to be extra beneficial for young individuals (c.f. the blue/triangle lines). Dividing the sample by education level (the right panel in Figure 12), there are indications that the language courses in the program were helpful in terms of improved language skills for both the relatively low-educated (blue/triangle lines) and the relatively high-educated (red/circle lines).

Table 1: Estimated program effect on language skills, nine months after randomization.

	All		Men		Women	
	(1)	(2)	(3)	(4)	(5)	(6)
	Any grade	No. of grades	Any grade	No. of grades	Any grade	No. of grades
Treatment	0.274 (0.001)	0.375 (0.001)	0.332 (0.002)	0.404 (0.007)	0.178 (0.156)	0.326 (0.056)
Women	-0.181 (0.027)	-0.233 (0.042)				
Constant	0.324 (0.000)	0.401 (0.000)	0.295 (0.000)	0.386 (0.000)	0.192 (0.034)	0.192 (0.112)
<i>N</i>	140	140	87	87	53	53

Notes: The estimates represent the effect of treatment on language skills measured nine months after randomization. *p*-values in parentheses from OLS estimations.

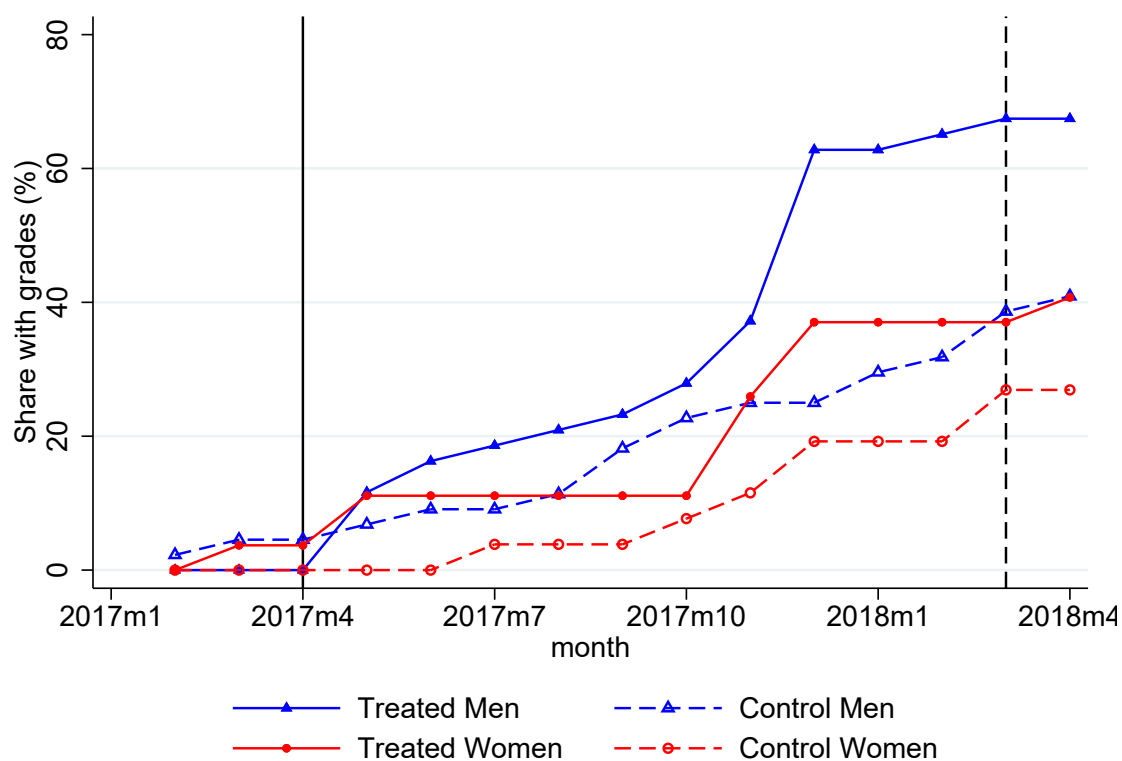


Figure 11: Documented language skills, by gender

Note: Based on calculations on data from Swedish For Immigrants (SFI) registers over started and completed SFI courses.

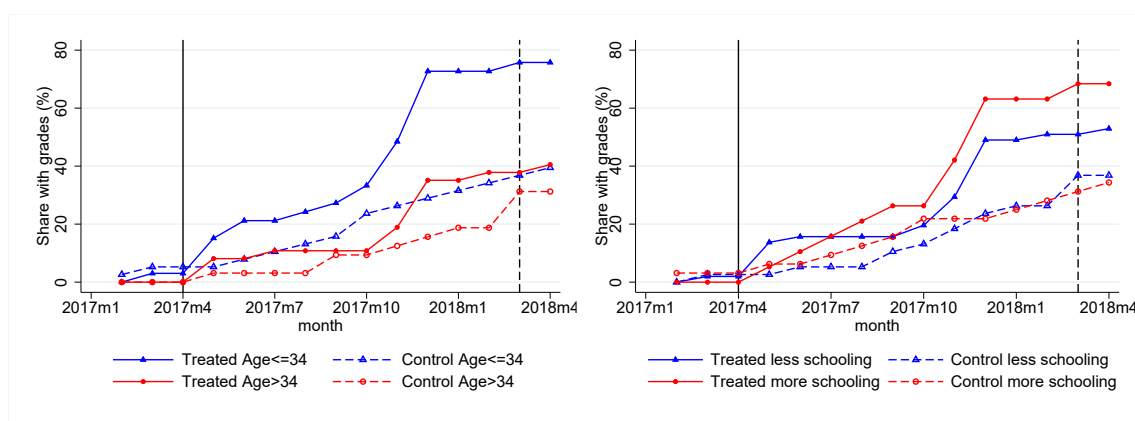


Figure 12: Documented language skills, by age and education

Note: Based on calculations on data from Swedish For Immigrants (SFI) registers over started and completed SFI courses. Less schooling is defined as less than 9 years and more schooling is defined as 9 or more years of education.

Mediation results: The role of improved language skills

Next, we examine how much of the effect on employment that is explained by the increased Swedish language skills. For the decomposition analysis, we follow Heckman et al. (2013) who examined how the Perry Preschool program affected mediators (personality skills) and what role these mediators had on later-in-life outcomes. We conduct the analysis in two steps. In the first step, performed in the section above, we examine if the new integration program affected the participants' documented language skills, and in the second step we decompose the estimated effects of the new program on the probability to be employed into a component attributable to such improvements in language skills.²⁰ A causal interpretation of the language mediator hinges on two assumptions. First, all unobserved factors must be uncorrelated with both program participation and the mediator (in our case increased language skills due to the intervention). Second, all unobserved factors should be orthogonal to the link between the mediator and the probability to be employed. The first assumption is likely to hold since we use an RCT design. However, randomization does not guarantee that the second assumption is fulfilled, and hence we should be careful making too strong claims about causality.

It is clear from table 2, column (6), that the language proficiency mechanism seems to grow in importance over time. In the short run (first year), increased language skills following program treatment explains around 7 percent of the effect on employment (11 percent if we use number of grades as our measure). In the longer run (from year two), 40 percent of the employment effect is explained by increased language skills in the treatment group. The importance of improved

²⁰More precisely, in a first step, we estimate the equation $Language_i = \gamma + \delta_1 Treatment_i + \delta_2 Female_i + \varepsilon_i$, where $Language_i$ measures documented language skills (any grade or number of grades), $Treatment_i$ denotes treatment status for individual i , ε_i is the error term and γ and $\delta_1 - \delta_2$ are parameters to be estimated. In a second step, we estimate the equation $Y_i = \alpha + \beta_1 Treatment_i + \beta_2 Language_i + \beta_3 Female_i + \varepsilon_i$, where Y_i is the employment status for individual i in each of the months 12–25, respectively, $Female_i$ is a gender dummy, ε_i is the error term, and α and $\beta_1 - \beta_3$ are parameters to be estimated. Using the estimated coefficients from these equations we can then calculate how large share of the estimated total effect ($\beta_1 + \hat{\beta}_2 \times \hat{\delta}_1$) that can be attributed to increased language skills ($\frac{\hat{\beta}_2 \times \hat{\delta}_1}{\beta_1 + \hat{\beta}_2 \times \hat{\delta}_1}$), where $\hat{\beta}_2 \times \hat{\delta}_1$ is the contribution from improved language skills. This way of decomposing the effect is also in line with Grönqvist et al. (2020).

language skills hence seems to grow over time, but since the assumption that all unobserved factors should be orthogonal to the link between the mediator and the probability to be employed can't be guaranteed, these results must be considered as offering only suggestive evidence.²¹

Table 2: Decomposition of treatment effect on number of months employment, 1-3 years after program.

	(1)	(2)	(3)	(4)	(5)	(6)
Any grade	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\delta}_1$	$\hat{\beta}_2 * \hat{\delta}_1$	$\hat{\beta}_1 + \hat{\beta}_2 * \hat{\delta}_1$	$\frac{\hat{\beta}_2 * \hat{\delta}_1}{\hat{\beta}_1 + \hat{\beta}_2 * \hat{\delta}_1}$
First year	1.701	.4816	.274	.132	1.833	.07198
Second year	.8161	1.868	.274	.5119	1.328	.3855
Third year	.8127	2.011	.274	.551	1.364	.4041
No. of grades	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\delta}_1$	$\hat{\beta}_2 * \hat{\delta}_1$	$\hat{\beta}_1 + \hat{\beta}_2 * \hat{\delta}_1$	$\frac{\hat{\beta}_2 * \hat{\delta}_1}{\hat{\beta}_1 + \hat{\beta}_2 * \hat{\delta}_1}$
First year	1.63	.5432	.3748	.2036	1.833	.111
Second year	.7945	1.422	.3748	.533	1.328	.4015
Third year	.858	1.345	.3748	.5041	1.362	.3701

Note: For the decomposition analysis we follow Heckman et al. (2013) as described in footnote 20. Number of months employed are summarized from April-March each year.

4 Long-run effects on social and economic integration

In section 3 we found that the program managed to get low-educated newly arrived refugees integrated on the labor market (in terms of employment), which was also the main aim of the program. The fast labor market entry in combination with a stable long-run position on the labor market might have increased the probability that the individuals in the treatment group can feel more integrated in other dimensions as well.²² In this section, we will examine if the program had any effect on other dimensions of integration than on the labor market outcomes. To accomplish this, we adopt the multidimensional integration index developed by Harder et al. (2018), slightly modified to the Swedish context. The survey interviews were conducted in the fall of 2021.

4.1 Data collection

The goal with the integration index developed by Harder et al. (2018) is to provide scholars with a “pragmatic, survey-based measure of immigrant integration”. The authors defined integration as “the degree to which immigrants have the knowledge and capacity to build a successful, fulfilling life in the host society”. By using their index, our results can also be compared to other studies using the same measure.

The integration index builds on six different dimensions. The questions for *Psychological integration* captures the immigrants feeling of connection to the host country while *Political in-*

²¹The residual contains all the other mechanisms that may explain the large treatment effect we find, such as improved social skills and increased networks following, e.g., the supervised work practice.

²²There are also components of the program (all the different courses providing not only knowledge about the Swedish language but also knowledge about the Swedish society and the inclusive part of the supervise work practice), that might have increased the probability that the individuals in the treatment group can feel more integrated in different dimensions.

tegration aims to capture the respondents understanding of political issues and engagement in discussions about them. The measure for *Social integration* captures social ties and connections to natives in the host country, whereas the goal with *Linguistic integration* is to measure the immigrants' assessments of their ability to read and speak the dominant language, in our case Swedish. To measure *Navigational integration* the questions investigate the respondents' ability to manage basic needs in the host country, such as visiting a doctor or searching for a job. The last dimension in the integration index concerns *Economic integration* to capture employment, income and satisfaction with the current economic situation.

The answer to each of the different questions are given scores that are then summarized into an index between 0 and 1. The index can be calculated separately for each of the six dimension as well as for the six dimensions combined into an overall integration index (the so called IPL-index). Even though we base our questions on the multidimensional integration index developed by Harder et al. (2018), we slightly adapt the questions to the Swedish setting. There are two forms of the index, a short form (IPL-12) and a long-form (IPL-24). To increase the response rate in our group of low-educated refugees that are fairly new to Sweden, we chose to adopt the index with the fewest set of questions (which is the IPL-12 index). However, we excluded the questions related to economic integration since we have access to register data to capture this dimension. The final questions we use and how we adapted them to our setting are described in Appendix section A.3.

The interviews

Given our sample of low-educated, recently arrived refugees, we chose, in order to minimize the attrition rate, not to send out a survey by mail. Rather, we chose to ask the questions in person (with interpreter at hand if necessary; the languages covered were Arabic, Somali, Farsi, Tigrinya)²³. It was a bit of detective work to find all individuals in the treatment and control groups, but in the end we reached an overall response rate of 80 percent (85 percent response rate in treatment group and 75 percent response rate in control group; c.f. Table 3). From Table 3 it is also clear that very few of those that we managed to get in contact with chose not to participate in the survey (six persons in total, two in the treatment group and four in the control group). The majority of the non-responses were individuals that we could not get in contact with (no answer when calling them; 21 in total, eight in the treatment group and 13 in the control group). For six individuals updated contact information were missing in the PES registers, one of them had emigrated.

Those that answered the survey had had their residence permit a bit longer, had permanent residence permit to a larger extent, and when it comes to education; were more in the middle part of the education distribution (with some formal education, but not more than ten years of schooling) (c.f. Table 4).

²³The interviews were conducted by a caseworker at the PES in Gothenburg.

Table 3: Survey response rate

	Treatment	Control	Overall
Yes	57 (85%)	50 (75%)	107 (80%)
No/No answer'	10	17	27
No	2	4	6
No answer	8	13	21
Missing/Emigrated	3	3	6

Table 4: Selection to survey

	Survey	No Survey	Diff	P-value
Treated	0.533	0.394	0.139	0.166
No formal education	0.140	0.212	-0.072	0.324
Up to ten years of school	0.813	0.636	0.177	0.035
Upper secondary school or more	0.047	0.152	-0.105	0.041
Time since residence permit	234.336	209.727	24.609	0.133
Women	0.364	0.424	-0.060	0.539
Age	36.533	37.788	-1.255	0.532
Born in Syria	0.542	0.576	-0.034	0.736
Born in Eritrea	0.178	0.121	0.056	0.449
Born in Somalia	0.112	0.030	0.082	0.159
Born in rest of the world	0.168	0.273	-0.105	0.186
Time since coming to Sweden	688.058	604.242	83.815	0.367
Share with permanent residence	0.533	0.364	0.169	0.091
Observations	107	33	140	140

Note: P-values from separate OLS regressions.

4.2 Results

As discussed in Appendix section A.3, we exclude the questions concerning economic integration from our questionnaire. For economic integration, the index is instead built on register information on number of months with employment and number of months without social assistance in the fall of 2021 (when the interviews were conducted). Following Harder et al. (2018) we calculate both an overall index, IPL-12, and a separate index for each of the six dimensions.

The distribution of the overall index is shown in Figure 13. The grey (darker) bars show the distribution for the individuals in the treatment group, while the orange (lighter) bars show the distribution for the individuals in the control group. To make it easier to see if the distributions differ, kernel density estimations are also included (blue/solid for the treatment group and pink/dashed for the control group). In addition to this, the p-value from the non-parametric Wilcoxon rank-sum test, which were used to examine if the distributions are significantly different from each other, are shown at the bottom of the figure. For the overall integration index, it is clear that we do not see any differences between the treatment group and the control group.

From the earlier analysis of the effects on employment, we found much clearer effects for

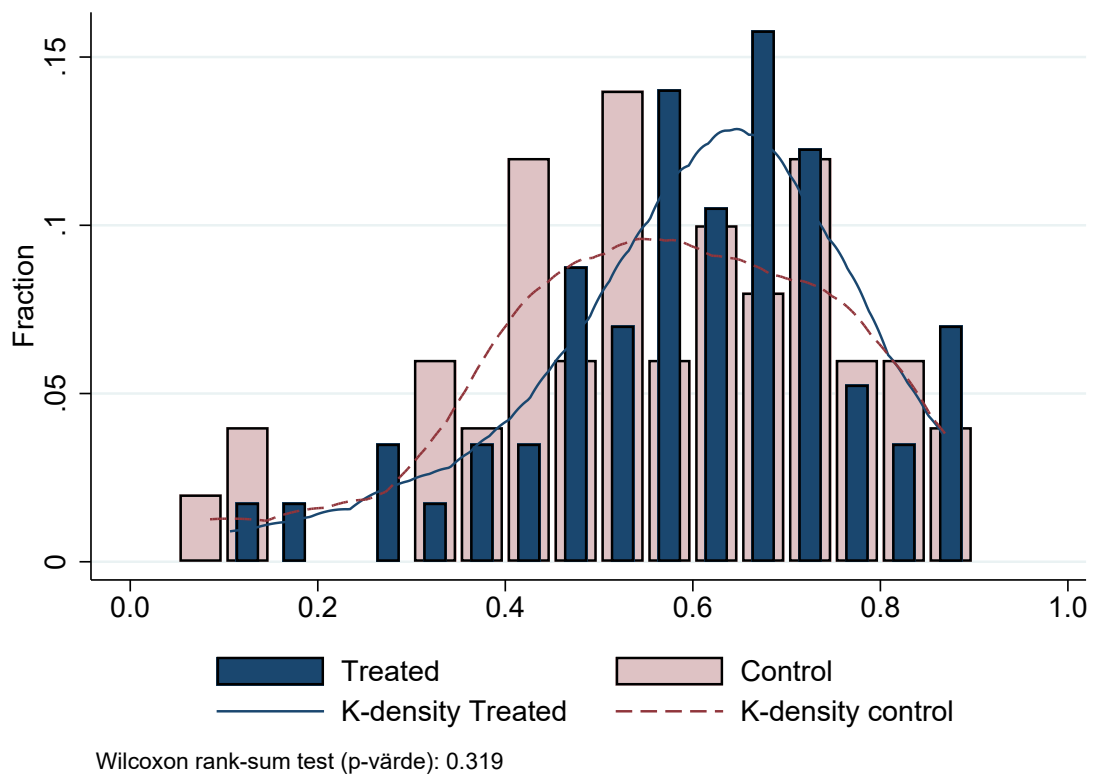


Figure 13: Distribution of IPL-12 index

Note: Each bar shows the fraction within a range of 0.05 on the horizontal axis. K-densities are estimated using the Epanechnikov kernel. The p-value at the bottom is from the non-parametric Wilcoxon rank-sum test, which were used to examine if the distributions are significantly different from each other.

men than for women. To examine if this is the case also for the IPL-10 index, we present the distribution of the index by gender (see Figure 14). Irrespective of sex, we find no statistically significant differences in the integration index between the treatment and the control groups.

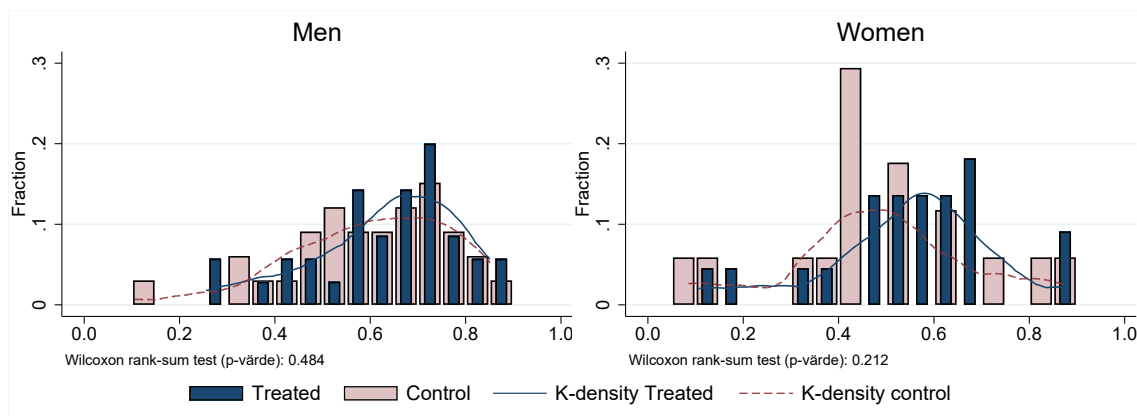


Figure 14: Distribution of IPL-12 index, by sex

Note: Each bar shows the fraction that lies within a range of 0.05 on the horizontal axis. K-densities are estimated using the Epanechnikov kernel. The p-value at the bottom is from the non-parametric Wilcoxon rank-sum test, which were used to examine if the distributions are significantly different from each other.

Even if we do not see any differences between the individuals in the treatment group (i.e., individuals offered to participate in the program) and individuals in the control group in the overall index, the program may have affected integration in some of the six social integration dimensions. In addition, since we think that the program mainly affected social integration for those who *actually* participated in the program (i.e., those that actually got the extra language courses and experienced the well-structured supervised work practice), we run IV-estimations (where we instrument actual participation with treatment status) for the overall index and for each dimension separately.

From the results, shown in Table 5²⁴, it is clear that program participants seem to be more integrated in one social dimension; the political (point estimate 0.13, p-value 0.04). It is also clear that this result is driven by the males (see the analyses by gender in Table A7 in the Appendix and in the histograms over the separate indices in Figure A3 in the Appendix). Even if some caution should be taken (since we study many outcomes and some may be significant by chance) the result for political integration is in line with the results in Heller and Slungaard Mumma (2023). They study English language training for immigrants in Massachusetts and find not only increased earning due to the program but also that voter registration doubled.

²⁴The ITT estimates are shown in Table A6 in the Appendix.

Table 5: IV estimations for integration dimensions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	IPL-12	Psychological	Political	Social	Linguistic	Navigational	Economic
	b/p	b/p	b/p	b/p	b/p	b/p	b/p
Participants	0.056 (0.22)	-0.025 (0.68)	0.130 (0.04)	-0.042 (0.53)	0.014 (0.81)	0.102 (0.18)	0.136 (0.19)
Women	-0.097 (0.00)	-0.001 (0.98)	-0.037 (0.43)	-0.087 (0.07)	-0.091 (0.04)	-0.167 (0.00)	-0.181 (0.02)
Constant	0.541 (0.00)	0.775 (0.00)	0.223 (0.02)	0.429 (0.00)	0.539 (0.00)	0.740 (0.00)	0.483 (0.00)
r2	.12	.00062	.031	.0094	.048	.13	.08
N	107	107	106	106	107	107	107

Note: P-values from 2SLS estimations in parentheses. First stage is shown in Table A4.

4.3 How should we understand the survey results?

Successful labor market integration is often seen as key for successful integration also in other dimensions. Even if we have found that the program accelerated labor market entry, we do not see any clear effects in other dimensions of integration between the treatment and the control group. One explanation could be that when we conducted the survey in the fall of 2021, the control group had caught up with the treatment group in terms of employment rates. At the same time, it is somewhat surprising that having been established for a longer time on the labor market is not translated into more social integration. Here we will provide some potential explanations.

In some of the dimensions we find that the individuals in the control group consider themselves as rather integrated (prime examples here are psychological and navigational integration; c.f. the estimate for the constant in Table 5 and the distributions of the indices in Figure A2 in the Appendix). The scope for the program to increase integration in these dimensions is therefore limited.

The non-significant results for linguistic integration is especially interesting given the short-run positive treatment effect on Swedish language skills that we presented earlier (see Figure 10 and Table 1). One explanation for the diverging results is that individuals in the control group have closed the gap over time (the short-run effects on language skills were estimated nine months after randomization, in December 2017, and the survey questions were asked in the fall of 2021). Another potential explanation is that the individuals in the treatment and the control groups respectively, in their self-evaluation of their Swedish language skills, may have different reference points to which they compare their knowledge in Swedish, as well as in other dimensions. Descriptive support for the latter explanation is given by the fact that individuals in the treatment group made use of an interpreter when the survey questions were asked to a lower extent than what the individuals in the control group did.²⁵

Another explanation can be the time dimension. Even though the survey was conducted more than three years after the end of the program, it might still be too short a time for any potential effects from the program on social integration to have visualized.

²⁵We gathered information on whether the respondents made use of an interpreter or not when conducting the interviews. The correlation between treatment status and making use of an interpreter is negative.

A final explanation can be our limited sample size, making it difficult to rule out that a potential impact on some of the dimensions can be hard to detect statistically.

5 Marginal value of public funds (MVPF)

In our earlier paper (Dahlberg et al., 2023), we calculated the additional cost of the program compared to the cost of the integration program to be 1.2 million SEK or 27,000 SEK per participant. This is cheaper than many of the other Swedish labor market programs (such as labor market training) but the important question is how these costs, and potentially other costs, relate to the benefits of the program. In order to answer this question, we will use the approach developed in Hendren and Sprung-Keyser (2020) and calculate the marginal value of public funds (MVPF).

The MVPF is given by comparing the benefits accruing to the individuals, i.e. their willingness to pay (WTP), with the public net costs of the program: $MVPF = WTP / \text{Net Cost}$. Since the program was initiated by the municipality of Gothenburg, we will calculate the cost and benefits separately for the municipality and for the rest of the public sector (which is mainly the Swedish Public Employment Service). It is important to stress that in this simple cost benefit analysis we will not be able to capture all potential cost and benefits of the program. We will, for instance, not put any value on integration in any social dimensions (such as increased political integration) or any increased well-being, or of getting a job (above the pure earnings benefits). The MVPF calculations will only focus on financial benefits and costs that we have access to, but we will discuss what we are lacking and how that may affect the calculations. We are able to capture cost and benefits up until December 2021. To see how they vary over time, we present them by year in Table 6.

Starting with the individuals' benefits of the program, we follow earlier literature and study the participants' changes in income. Preferably, we would have liked to use disposable income for these calculations, but what we have access to is income from earnings and social assistance. These are probably the main income sources that varies due to the program.

During the year of the program, there is a lock-in effect and program participants have on average 17,000 SEK less income than non-participants (c.f. the first column in the first panel of Table 6). However, for the four years following the end of the program, except for the last year that we are able to study, program participants have higher incomes as an effect of the program. All in all, this adds up to a total of 15,942 SEK in extra income over the whole time period (c.f. the last column in Table 6).

From the municipality's perspective, the program is also beneficial. During the first year following randomization, the total cost amounts to almost 30,000 SEK per participant (as seen in the first column of the second panel in Table 6). This cost is primarily driven by the direct expenses related to intensive training.²⁶ However, over the four years that follow, the municipality benefits from reduced expenditures on social assistance as well as increased tax revenues. The municipal tax system in Sweden is based on pure income tax, and this positive effect results from higher employment rates among individuals in the treatment group. In total, the municipality experiences reduced social assistance costs of 16,636 SEK per participant and increased income tax revenues of 10,208 SEK per participant (as depicted in the last column of the second panel in Table 6).

²⁶These calculations also include costs for the housing company AB Framtiden.

Finally, we examine the program's effects on the rest of the public sector. While there were some costs incurred by the Public Employment Service (PES) due to increased job search assistance during the program, these costs are outweighed by the subsequent benefits. After the program's completion, the public sector benefits from increased employer contributions and higher regional tax revenues, which exceed the initial expenses by far.

Based on these calculations and the factors we've considered, the program is self-financing, providing advantages to both the participating individuals and different segments of the public economy.

Table 6: Estimated changes in earnings and costs, during and up to four years after program.

Individual Benefits (WTP)	Years after program					
	During	1	2	3	4	Total
Earnings	-25688	17789	37642	23272	-4680	48335
Tax(32.6%)	8374	-5799	-12271	-7587	1526	-15757
Social assistance	-281	-2597	1857	-7632	-7983	-16636
Total	-17595	9393	27228	8054	-11137	15942
Changes in Municipal Costs(+) and benefits(-)						
	During	1	2	3	4	Total
Municipal cost	23840	0	0	0	0	23840
Social assistance	-281	-2597	1857	-7632	-7983	-16636
Municipal tax (21.12%)	5425	-3757	-7950	-4915	988	-10208
Total	28984	-6354	-6093	-12547	-6994	-3004
Changes in rest of Public Costs(+) and benefits(-)						
	During	1	2	3	4	Total
PES cost	3190	0	0	0	0	3190
Regional tax (11.48%)	2949	-2042	-4321	-2672	537	-5549
Employer contribution (31.42%)	8071	-5589	-11827	-7312	1471	-15187
Total	14210	-7631	-16148	-9984	2008	-17546
Total changes in Public Costs(+) and benefits(-)						
	During	1	2	3	4	Total
Total costs	43195	-13986	-22242	-22531	-4986	-20550

Note: Changes in earnings and social assistance are based on IV estimations presented in Table A8. Changes in taxes and employer contributions are calculated using the estimated effect on earnings. All years are from April-March, except for the fourth year that only includes nine months, March-December.

The calculations in Table 6 are based on variables we observe. An important question is then what aspects we might have overlooked. There are other income sources for the group we are studying, but some of these (such as child allowances) remain unaffected by the program. The primary income sources we lack data on, that may change due to the program, are the integration benefit, activity support, and unemployment insurance (UI). The integration benefit is the payment that newly arrived refugees forfeit if they secure employment within the two-year integration period. Activity support is provided to unemployed individuals participating in programs at PES, excluding the integration period. To qualify for UI, an unemployed person must have been a member of a UI-fund for at least 12 months and meet a work requirement equivalent to working at least

half-time for six months.²⁷

While the individuals in our treatment group were predominantly employed through New Start Jobs after the program, often for two years (see Figure A4 in Appendix where we present the results for non-subsidized employment and New Start Jobs separately), the individuals in the control group were more likely to participate in a costlier Job Creation Scheme, albeit for a shorter period (see Figure A5 in Appendix).²⁸ As the control group tended to remain unemployed to a greater extent, they became eligible for activity support when participating in PES programs. On the other hand, individuals in our treatment group, became eligible for unemployment insurance (UI), in case of job loss after two years of New Start Job employment. This likely explains the observed patterns in the fourth year, where participants, on average, experienced lower earnings and received less social assistance.

On average, participating individuals will receive reduced integration benefits and activity support due to the program, but they will receive more unemployment insurance (UI). It's important to note that alterations in these benefits would also affect the net program cost, leading to only minor adjustments in the calculated MVPF.

However, it's also worth noting that the control group, which tends to remain unemployed at the PES and require more social assistance, will necessitate additional administrative efforts and job search assistance—factors we haven't accounted for. This omission likely results in an overestimation of the net costs, as it can be considered a reduced cost for participating individuals.

In conclusion, considering the four-year follow-up period, while we may not be able to account for all potential changes in the public net costs resulting from the program, the majority of these changes are likely to involve shifting costs among various forms of subsidies and benefits. Importantly, these changes are unlikely to alter the fundamental conclusion that the program effectively self-finances and accelerates the employment process for newly arrived refugees.

6 Conclusion

This paper has examined the long-run effects of a new labor market program instigated in Gothenburg in Sweden to combat the low employment rates among newly arrived, low-educated refugees. The program uses an early intervention strategy that contains three main components; it starts with three months of intensive, full-time, language courses, followed by six months of supervised work practice organized by the Swedish Public Employment Service together with the largest real estate company in Gothenburg, and finalized by job search assistance in which both the PES and the housing company assists.

Using a randomized controlled trial design, we evaluate the long-run effects of the program on labor market integration (measured via employment outcomes) and on multidimensional integration (psychological, linguistics, economic, political, social, and navigational integration). The follow-up horizon is almost four years after the end of the program, and we use detailed register data as well as survey information gathered via telephone interviews (the questions used in our

²⁷Note that a New Start Job employment give access to UI, while employment within a job creation scheme does not.

²⁸The primary Job Creation Scheme used for refugees, known as *Extratjänster*, had the potential for a second-year extension. However, in early 2019, *Extratjänster* was discontinued by a political majority, which accounts for the observed shorter durations of Job Creation Scheme participation in our control group.

survey was developed by (Harder et al., 2018)).

We reach three main types of conclusions. The first type of conclusions is related to the labor market effects of the program. We find that the positive and immediate effect of the program on employment found in Dahlberg et al. (2023) lasts in the longer run. The individuals in the treatment group have a higher employment rate than the individuals in the control group for more than three years after the end of the program. It is only at the end of the follow-up period, after approximately four years, that the control group catches up. Since estimates of the long-run effects of integration programs have been missing in the earlier literature (Foged et al., 2022a), these results add new and important information.

In addition, we find that the employment rates and the estimated effects are sizeable. The employment rate in the treatment group is around 45 percent for most of the follow-up period, which is 10-20 percentage points higher than that in the control group (until the very end of the period). Furthermore, the individuals in the treatment group are employed up to 1.5 months per quarter, which is approximately half a month more per quarter than the individuals in the control group. These magnitudes must be considered large. The point estimates are statistically significant in most months after the end of the program. It is important to note that when the effect estimates get closer to zero and become insignificant, it is because the control group catches up, not that employment in the treatment group decreases.

In terms of heterogeneity, we show that the positive employment effects of the program are mainly driven by the male participants. The employment rate among the male participants in the treatment group is around 60 percent for most of the follow-up horizon. This finding is a bit surprising since earlier literature has shown that women often do not participate in labor market programs that resembles real jobs, but when they do they benefit from doing so (Cheung, 2018; Bratu et al., 2023). In our case, they do not. Why? One reason may be that the jobs focused on in the Gothenburg program (jobs related to indoor and outdoor maintenance of housing properties) are less attractive to women. Among those individuals that were offered to participate in the new labor market program, it was also a lower share of women than men that accepted the offer to participate.

We also find, in a descriptive heterogeneity analysis, clear indications that the program seems to be extra beneficial (in terms of employment rates) for the relatively older (above 34 years of age) and the relatively low-educated (less than nine years of schooling). Since these are groups that have been identified as extra vulnerable on the labor market, these results are important to bear in mind when constructing future programs targeting these groups.

We argue that our labor market results have strong external validity. When constructing a group of low-educated refugees for the whole of Sweden that is similar to the group of refugees in our experiment, we find that the employment rate in the control group closely follows that of the rest of Sweden (and especially Stockholm), indicating that the program would have had similar positive effects in other cities, had it been implemented there. In addition, we see no obstacles to implementing the program in other cities (neither in Sweden nor in most other Western countries). As argued in our earlier paper (Dahlberg et al., 2023), the risk of displacement effects appear to be small. This view is strengthened by our comparison with the development in the rest of Sweden: in doing that comparison, we see no signs of any displacement effects.

The second type of conclusions is related to the language effects of the program. Since an important component of the intervention we study is intensive language training, an obvious and interesting question is to what extent this led to an improvement in the participants' knowledge of Swedish. We find that the program has positive and statistically significant effects on documented language skills (both in terms of the probability of taking any grade and in terms of number of grades taken). This paper is hence among the first to show that language training for immigrants has an effect on actual language skills. When we then, in a decomposition analysis, examine the role of language skills for the employment outcomes, we find that the short-run improvement in language skills over time explains an increasing share of the variation in the estimated treatment effect on employment – explaining around 40 percent of the employment effect up to three years after the end of the program.

The third type of conclusions is related to multidimensional integration. Using the multidimensional integration index developed by Harder et al. (2018), we find that most of our study population perceive themselves as relatively well integrated in most of the studied dimensions (especially in terms of psychological, societal, and linguistic integration and somewhat less in terms of social and political integration). When we examine whether the intervention had any effect on the multidimensional integration, we find that the participants in the intervention perceive themselves as more politically integrated than those who did not participate in the intervention. This result is also driven by the male participants. For the other dimensions, however, there are no statistically significant differences. These findings indicate that faster labor market integration in the short run may not automatically result in broader (self-perceived) integration in the longer run.

Finally, when conducting a cost-benefit analysis, in which we compare the program to the baseline service using the framework developed by Hendren and Sprung-Keyser (2020), we conclude that the program is self-financing, offering benefits to both the participating individuals and various segments of the public economy. Although we lack data on costs related to different subsidies and benefits, we argue that this mainly involves a shifting of costs that will not alter the overall conclusion. Instead, we believe that our assessment underestimates the reduction in costs, as it does not account for the reduced administrative burden on the municipality and the Swedish Public Employment Service.

References

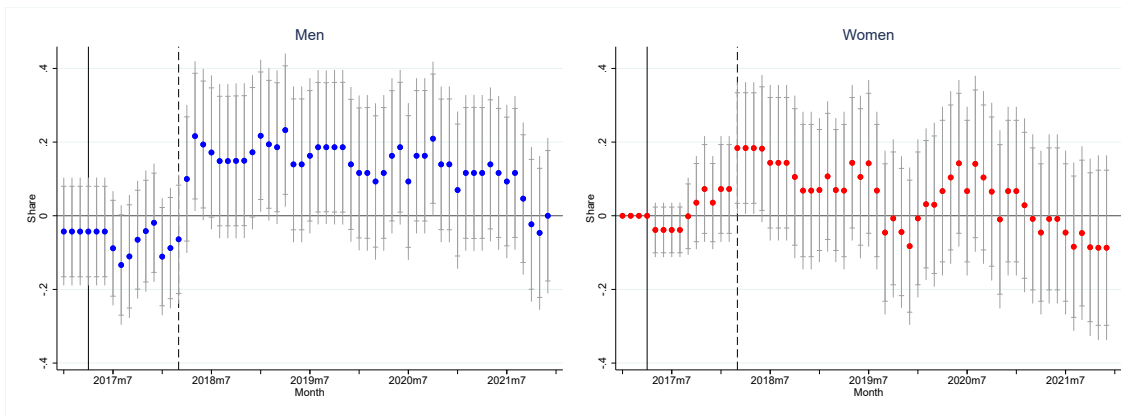
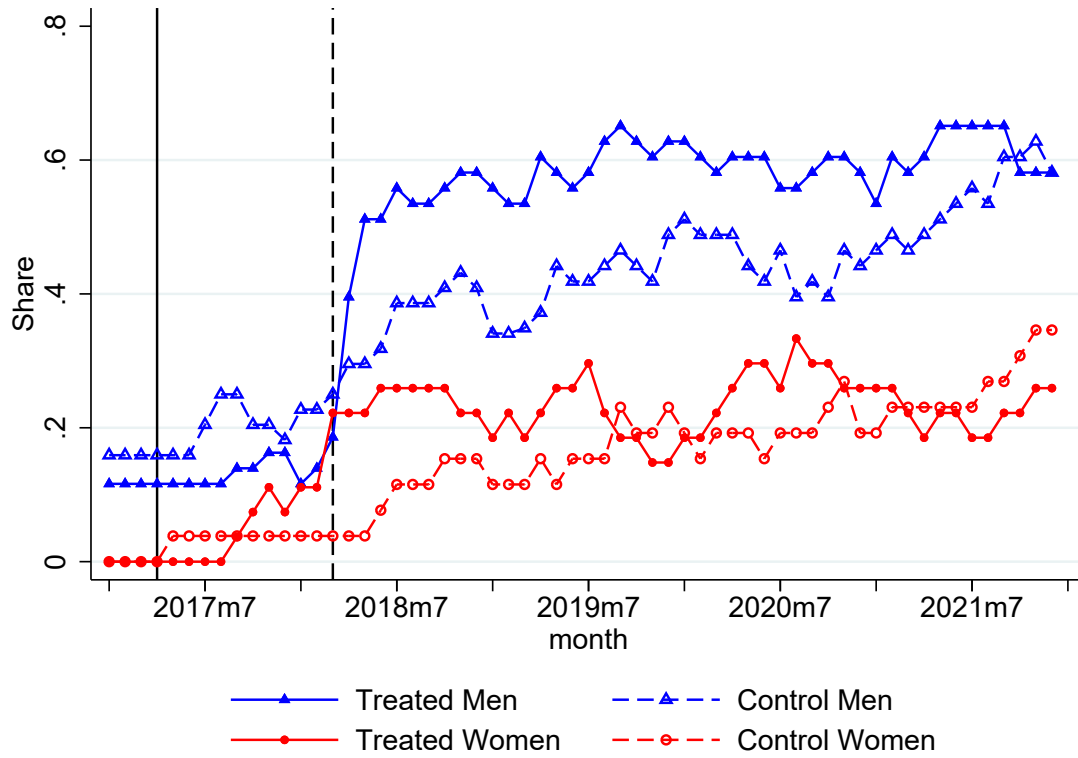
- Aksoy, C. G., Poutvaara, P., and Schikora, F. (2023). First time around: Local conditions and multi-dimensional integration of refugees. *Journal of Urban Economics*, 137:103588.
- Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., and Weinstein, J. (2018). Improving refugee integration through data-driven algorithmic assignment. *Science*, 359(6373):325–329. Publisher: American Association for the Advancement of Science.
- Bratu, C., Dahlberg, M., and Valeyatheepillay, M. (2021). Age at Arrival and Residential Integration. SSRN Scholarly Paper ID 3885424, Social Science Research Network, Rochester, NY.
- Bratu, C., Martén, L., and Ottosson, L. (2023). Individualized benefits and access to active labor market programs boost refugee women’s economic integration. Swedish Institute for Social Research (SOFI) Working Paper 3/2023.
- Braun, S. T. and Dwenger, N. (2020). Settlement location shapes the integration of forced migrants: Evidence from post-war Germany. *Explorations in Economic History*, 77:101330.
- Card, D., Kluve, J., and Weber, A. (2018). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association*, 16(3):894–931.
- Cheung, M. (2018). Hur skapar vi en mer jämställd arbetsmarknadsetablering? Arbetsförmedlingens Working Paper 2018:2.
- Dahlberg, M., Egebark, J., Özcan, G., and Vikman, U. (2023). Labor market integration of refugees: RCT evidence from an early intervention program in Sweden. *Journal of Economic Behavior and Organization*, forthcoming.
- Fasani, F., Frattini, T., and Minale, L. (2021). Lift the Ban? Initial Employment Restrictions and Refugee Labour Market Outcomes. *Journal of the European Economic Association*, 19(5):2803–2854.
- Foged, M., Hasager, L., and Peri, G. (2022a). Comparing the effects of policies for the labor market integration of refugees. Working Paper 30534, National Bureau of Economic Research.
- Foged, M., Hasager, L., Peri, G., Arendt, J. N., and Bolvig, I. (2022b). Language Training and Refugees’ Integration. *The Review of Economics and Statistics*, pages 1–41.
- Foged, M. and van der Werf, C. (2023). Access to language training and the local integration of refugees. *Labour Economics*, 84:102366.
- Grönqvist, H., Nilsson, J. P., and Robling, P. O. (2020). Understanding how low levels of early lead exposure affect children’s life trajectories. *Journal of Political Economy*, 128:3376 – 3433.
- Harder, N., Figueroa, L., Gillum, R. M., Hangartner, D., Laitin, D. D., and Hainmueller, J. (2018). Multidimensional measure of immigrant integration. *PNAS*, 115(45):11483–11488.

- Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–86.
- Heller, B. H. and Slungaard Mumma, K. (2023). Immigrant Integration in the United States: The Role of Adult English Language Training. *American Economic Journal: Economic Policy*.
- Hendren, N. and Sprung-Keyser, B. (2020). A Unified Welfare Analysis of Government Policies. *The Quarterly Journal of Economics*, 135(3):1209–1318.
- Lochmann, A., Rapoport, H., and Speciale, B. (2019). The effect of language training on immigrants' economic integration: Empirical evidence from France. *European Economic Review*, 113:265–296.
- Pont-Grau, A., Lei, Y.-H., Lim, J. Z. E., and Xia, X. (2023). The effect of language training on immigrants' integration: Does the duration of training matter? *Journal of Economic Behavior & Organization*, 212:160–198.
- Public Statistics of Sweden (Sveriges Officiella Statistik) (2019). Statistiska Meddelanden AM 12 SM 2101.
- Sarvimäki, M. and Hämäläinen, K. (2016). Integrating Immigrants: The Impact of Restructuring Active Labor Market Programs. *Journal of Labor Economics*, 34(2):479–508.
- Åslund, O. and Rooth, D. (2007). Do When and Where Matter? Initial Labour Market Conditions and Immigrant Earnings. *The Economic Journal*, 117(518):422–448.

A Appendix

A.1 Additional results

Figures



Notes: Employment is defined as having positive earnings from non-subsidised work or a New Start Job during a month. Outcomes in top figure and ITT effects in bottom figures with 90/95 confidence intervals from OLS estimation.

Figure A1: Mean outcomes and ITT for employment, by sex.

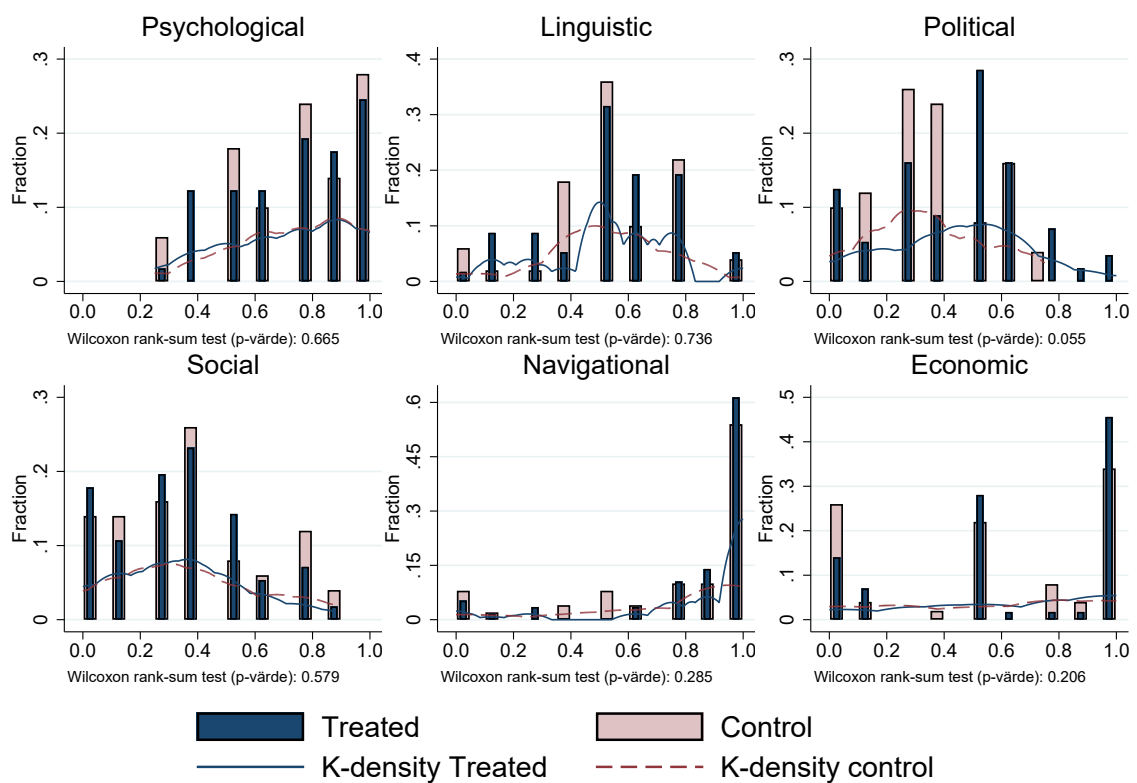


Figure A2: Distribution of all dimensions

Note: Each bar shows the fraction that lies within a range of 0.05 on the horizontal axis. K-densities are estimated using the Epanechnikov kernel. The p-value at the bottom is from the non-parametric Wilcoxon rank-sum test, which is used to examine if the distributions are significantly different from each other.

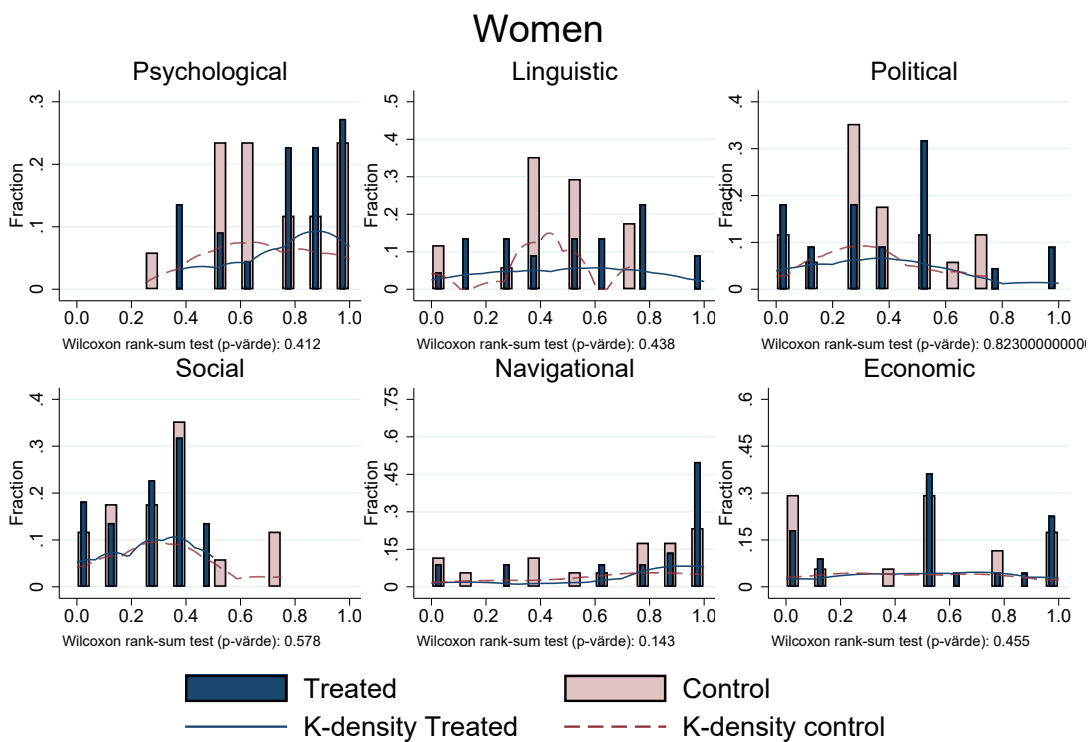
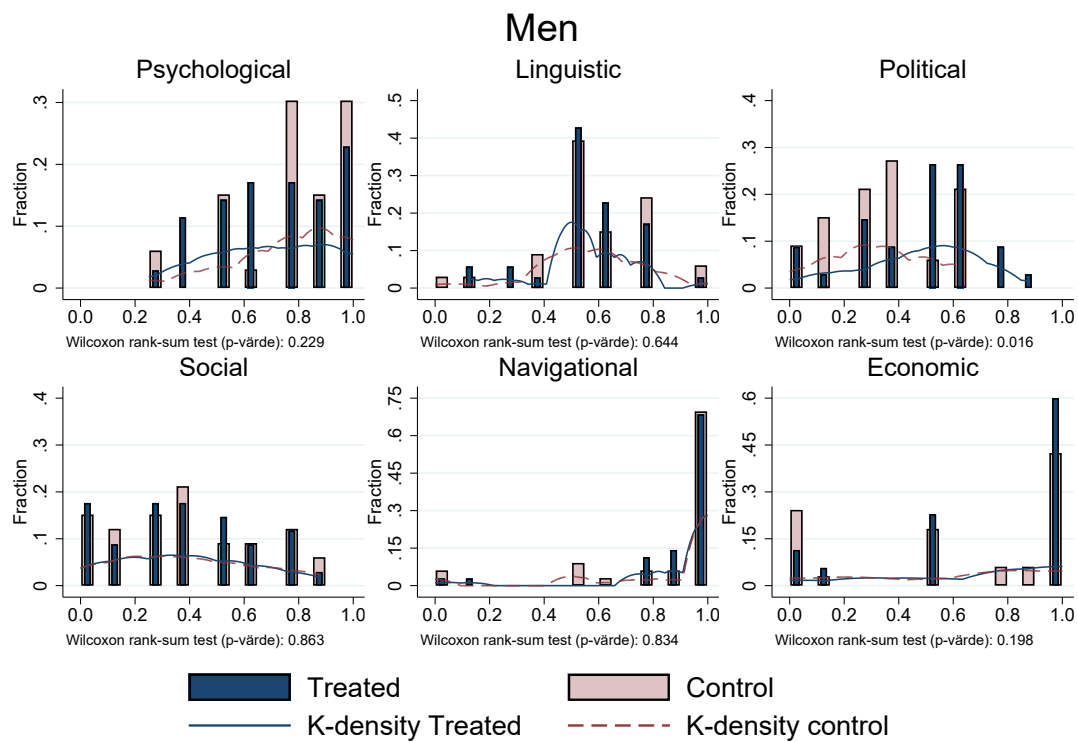
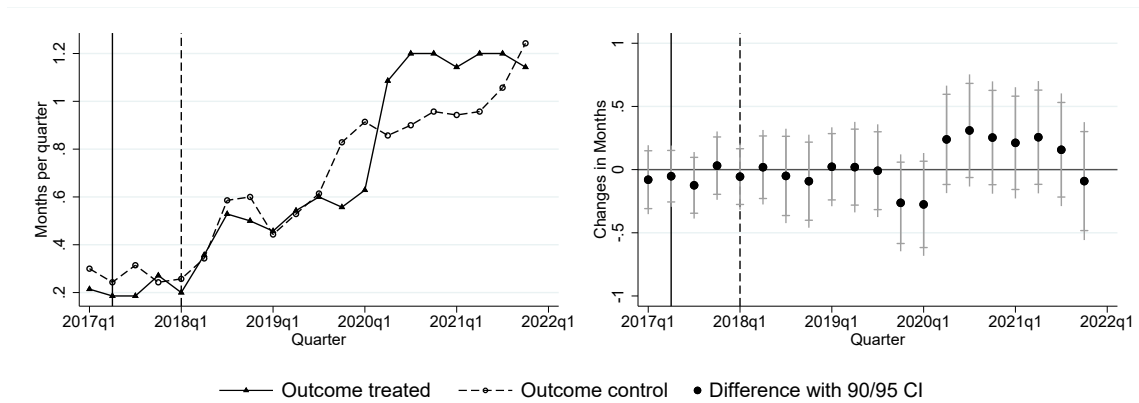
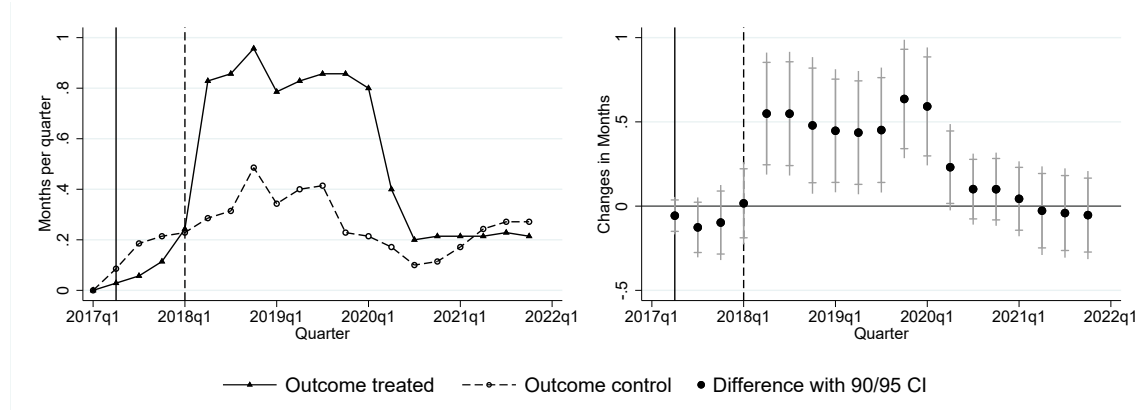


Figure A3: Distribution of all dimensions, by sex.

Note: Each bar shows the fraction that lies within a range of 0.05 on the horizontal axis. K-densities are estimated using the Epanechnikov kernel. The p-value at the bottom is from the non-parametric Wilcoxon rank-sum test, which is used to examine if the distributions are significantly different from each other.



(a) Non-subsidized employment



(b) New Start Jobs

Figure A4: Mean outcomes and ITT estimates for non-subsidized employment and New Start Jobs, separately

Notes: Outcomes in left figures and ITT effects in right figures with 90/95 confidence intervals from OLS estimation controlling for sex.

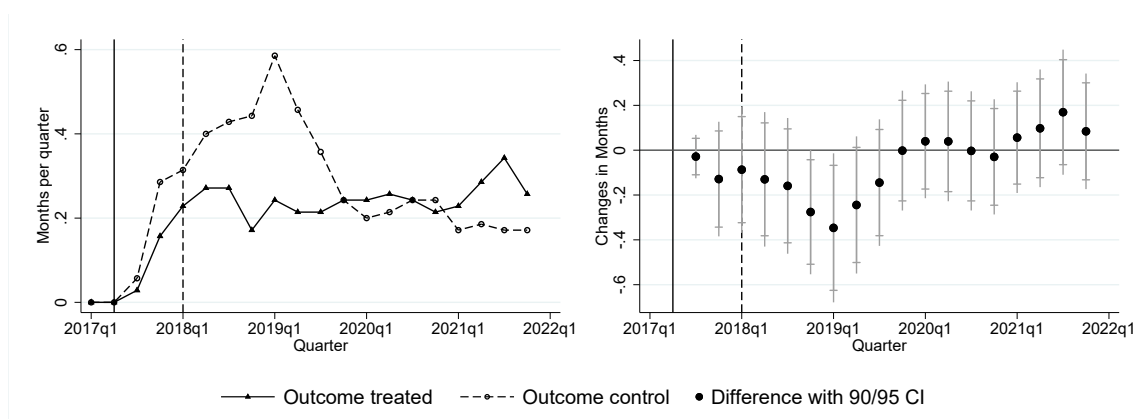


Figure A5: Mean outcomes and ITT estimates for Job Creation Schemes

Notes: Outcomes in left figure and ITT effects in right figure with 90/95 confidence intervals from OLS estimation controlling for sex.

Tables

Table A1: Point estimate and P-values (Asymptotic) for Employment outcome by month.

Month	Y0	Y1	ITT	P-value	Month	Y0	Y1	ITT	P-value
2017m1	0.100	0.071	-0.027	0.567	2019m7	0.319	0.471	0.155	0.054
2017m2	0.100	0.071	-0.027	0.567	2019m8	0.333	0.471	0.141	0.073
2017m3	0.100	0.071	-0.027	0.567	2019m9	0.377	0.471	0.098	0.219
2017m4	0.100	0.071	-0.027	0.567	2019m10	0.348	0.457	0.112	0.154
2017m5	0.114	0.071	-0.041	0.397	2019m11	0.333	0.429	0.098	0.209
2017m6	0.114	0.071	-0.041	0.397	2019m12	0.391	0.443	0.055	0.487
2017m7	0.143	0.071	-0.069	0.177	2020m1	0.391	0.457	0.069	0.379
2017m8	0.171	0.071	-0.098	0.070	2020m2	0.362	0.443	0.084	0.282
2017m9	0.171	0.100	-0.069	0.225	2020m3	0.377	0.443	0.069	0.388
2017m10	0.143	0.114	-0.027	0.634	2020m4	0.377	0.471	0.097	0.225
2017m11	0.143	0.143	0.002	0.979	2020m5	0.348	0.486	0.140	0.083
2017m12	0.129	0.129	0.002	0.977	2020m6	0.319	0.486	0.169	0.035
2018m1	0.157	0.114	-0.041	0.475	2020m7	0.362	0.443	0.083	0.303
2018m2	0.157	0.129	-0.027	0.648	2020m8	0.319	0.471	0.154	0.058
2018m3	0.171	0.200	0.030	0.652	2020m9	0.333	0.471	0.140	0.083
2018m4	0.200	0.329	0.132	0.071	2020m10	0.333	0.486	0.154	0.058
2018m5	0.200	0.400	0.204	0.006	2020m11	0.391	0.471	0.083	0.314
2018m6	0.229	0.414	0.189	0.013	2020m12	0.348	0.457	0.112	0.165
2018m7	0.286	0.443	0.161	0.040	2021m1	0.362	0.429	0.069	0.395
2018m8	0.286	0.429	0.147	0.061	2021m2	0.391	0.471	0.083	0.308
2018m9	0.286	0.429	0.147	0.061	2021m3	0.377	0.443	0.069	0.395
2018m10	0.314	0.443	0.133	0.095	2021m4	0.391	0.443	0.055	0.494
2018m11	0.329	0.443	0.119	0.131	2021m5	0.406	0.486	0.083	0.299
2018m12	0.314	0.443	0.133	0.091	2021m6	0.420	0.486	0.069	0.389
2019m1	0.257	0.414	0.161	0.035	2021m7	0.435	0.471	0.040	0.610
2019m2	0.257	0.414	0.161	0.037	2021m8	0.435	0.471	0.040	0.617
2019m3	0.261	0.400	0.142	0.064	2021m9	0.478	0.486	0.011	0.892
2019m4	0.290	0.457	0.170	0.031	2021m10	0.493	0.443	-0.047	0.563
2019m5	0.319	0.457	0.141	0.073	2021m11	0.522	0.457	-0.062	0.450
2019m6	0.319	0.443	0.127	0.113	2021m12	0.493	0.457	-0.033	0.688

Note: Employment is defined as having positive earnings from non-subsidised work or a New Start Job during a month. ITT effects and p-values from OLS estimations controlling for sex.

Table A2: Point estimate and P-values (Fisher and Asymptotic) for Employment outcome by quarter.

	Y0	Y1	ITT	Asymptotic	Fisher
2017q1	0.300	0.214	-0.080	0.567	0.759
2017q2	0.329	0.214	-0.109	0.444	0.557
2017q3	0.486	0.243	-0.236	0.133	0.146
2017q4	0.414	0.386	-0.024	0.887	0.925
2018q1	0.486	0.443	-0.039	0.826	0.866
2018q2	0.629	1.143	0.525	0.014	0.019
2018q3	0.857	1.300	0.455	0.047	0.062
2018q4	0.957	1.329	0.384	0.097	0.124
2019q1	0.771	1.229	0.469	0.036	0.052
2019q2	0.914	1.357	0.456	0.043	0.058
2019q3	1.014	1.414	0.414	0.063	0.085
2019q4	1.057	1.329	0.286	0.202	0.254
2020q1	1.114	1.343	0.244	0.290	0.349
2020q2	1.029	1.443	0.427	0.071	0.091
2020q3	1.000	1.386	0.396	0.090	0.112
2020q4	1.057	1F.414	0.368	0.115	0.141
2021q1	1.114	1.343	0.241	0.290	0.348
2021q2	1.200	1.414	0.229	0.325	0.396
2021q3	1.329	1.429	0.116	0.613	0.710
2021q4	1.486	1.357	-0.116	0.628	0.633

Note: Employment is defined as having positive earnings from non-subsidised work or a New Start Job during a month. ITT effects and p-values from OLS estimations and Fisher's exact tests controlling for sex.

Table A3: Point estimate and P-values (Asymptotic) for Employment outcome by quarter and by sex.

	Men				Women			
	Y0	Y1	ITT	P-value	Y0	Y1	ITT	P-value
2017q1	0.477	0.349	-0.128	0.568	0.000	0.000	.	.
2017q2	0.477	0.349	-0.128	0.568	0.077	0.000	-0.077	0.313
2017q3	0.705	0.372	-0.332	0.171	0.115	0.037	-0.078	0.514
2017q4	0.591	0.465	-0.126	0.606	0.115	0.259	0.144	0.447
2018q1	0.705	0.442	-0.263	0.288	0.115	0.444	0.329	0.145
2018q2	0.909	1.419	0.510	0.091	0.154	0.704	0.550	0.051
2018q3	1.159	1.628	0.469	0.137	0.346	0.778	0.432	0.175
2018q4	1.250	1.721	0.471	0.136	0.462	0.704	0.242	0.462
2019q1	1.023	1.628	0.605	0.052	0.346	0.593	0.246	0.414
2019q2	1.205	1.744	0.540	0.082	0.423	0.741	0.318	0.313
2019q3	1.295	1.860	0.565	0.065	0.538	0.704	0.165	0.594
2019q4	1.318	1.860	0.542	0.076	0.615	0.481	-0.134	0.672
2020q1	1.455	1.814	0.359	0.260	0.538	0.593	0.054	0.864
2020q2	1.318	1.814	0.496	0.120	0.538	0.852	0.313	0.366
2020q3	1.250	1.698	0.448	0.154	0.577	0.889	0.312	0.366
2020q4	1.273	1.791	0.518	0.100	0.692	0.815	0.123	0.719
2021q1	1.386	1.721	0.335	0.275	0.654	0.741	0.087	0.795
2021q2	1.500	1.907	0.407	0.192	0.692	0.630	-0.063	0.856
2021q3	1.659	1.953	0.294	0.337	0.769	0.593	-0.177	0.603
2021q4	1.773	1.744	-0.029	0.928	1.000	0.741	-0.259	0.479

Note: Employment is defined as having positive earnings from non-subsidised work or a New Start Job during a month. ITT effects and p-values from OLS estimations.

Table A4: First stage in IV estimations

	All		Survey sample			
	(1) All b/p	(2) Men b/p	(3) Women b/p	(4) All b/p	(5) Men b/p	(6) Women b/p
Treatment	0.630 (0.00)	0.698 (0.00)	0.519 (0.00)	0.704 (0.00)	0.743 (0.00)	0.636 (0.00)
Women	-0.090 (0.13)			-0.058 (0.39)		
r2	.47	.54	.35	.53	.58	.43
F	60	99	27	58	93	28
N	140	87	53	107	68	39

Note: Point estimates and p-values are obtained through OLS estimations to assess the effect of treatment on participating status.

Table A5: IV estimations for Employment outcome by quarter.

	All		Men		Women	
	IV	se	IV	se	IV	se
2017q1	-0.127	0.218	-0.184	0.315		
2017q2	-0.173	0.222	-0.184	0.315	-0.148	0.145
2017q3	-0.375	0.245	-0.477	0.338	-0.151	0.229
2017q4	-0.038	0.262	-0.180	0.342	0.277	0.345
2018q1	-0.061	0.276	-0.377	0.347	0.635	0.390
2018q2	0.833	0.325	0.730	0.420	1.060	0.468
2018q3	0.722	0.351	0.672	0.442	0.832	0.561
2018q4	0.610	0.359	0.675	0.446	0.467	0.599
2019q1	0.745	0.341	0.867	0.433	0.475	0.538
2019q2	0.723	0.346	0.773	0.439	0.613	0.546
2019q3	0.657	0.344	0.810	0.439	0.319	0.559
2019q4	0.455	0.347	0.777	0.431	-0.258	0.612
2020q1	0.387	0.359	0.515	0.450	0.104	0.592
2020q2	0.678	0.366	0.711	0.448	0.604	0.635
2020q3	0.629	0.362	0.642	0.439	0.602	0.639
2020q4	0.585	0.361	0.742	0.437	0.236	0.638
2021q1	0.382	0.352	0.480	0.425	0.168	0.628
2021q2	0.364	0.362	0.583	0.435	-0.121	0.653
2021q3	0.184	0.358	0.422	0.432	-0.341	0.658
2021q4	-0.184	0.377	-0.041	0.447	-0.500	0.711

Note: Employment is defined as having positive earnings from non-subsidised work or a New Start Job during a month. IV effects and standard errors from 2SLS estimations.

Table A6: ITT estimations for different dimensions of integration index.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	IPL-12	Psychological	Political	Social	Linguistic	Navigational	Economic
Treated	0.040 (0.24)	-0.018 (0.69)	0.091 (0.05)	-0.029 (0.53)	0.010 (0.81)	0.072 (0.20)	0.096 (0.20)
Women	-0.100 (0.00)	0.000 (0.99)	-0.044 (0.35)	-0.084 (0.08)	-0.091 (0.04)	-0.173 (0.00)	-0.189 (0.02)
Constant	0.598 (0.00)	0.750 (0.00)	0.355 (0.00)	0.386 (0.00)	0.554 (0.00)	0.844 (0.00)	0.622 (0.00)
r2	.083	.0016	.044	.034	.04	.09	.066
N	107	107	106	106	107	107	107

Note: P-values from OLS estimations in parentheses.

Table A7: IV estimations for different dimensions of integration index, by sex.

Men	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	IPL-12 b/p	Psychological b/p	Political b/p	Social b/p	Linguistic b/p	Navigational b/p	Economic b/p
Treated	0.031 (0.42)	-0.062 (0.25)	0.126 (0.02)	-0.015 (0.82)	-0.026 (0.59)	0.040 (0.51)	0.104 (0.28)
Constant	0.602 (0.00)	0.773 (0.00)	0.337 (0.00)	0.379 (0.00)	0.572 (0.00)	0.860 (0.00)	0.617 (0.00)
r2	.0097	.02	.083	.00081	.0045	.0066	.017
N	68	68	67	67	68	68	68
Women	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	IPL-12 b/p	Psychological b/p	Political b/p	Social b/p	Linguistic b/p	Navigational b/p	Economic b/p
Treated	0.054 (0.40)	0.061 (0.40)	0.029 (0.73)	-0.055 (0.38)	0.074 (0.39)	0.127 (0.26)	0.082 (0.49)
Constant	0.489 (0.00)	0.706 (0.00)	0.346 (0.00)	0.316 (0.00)	0.426 (0.00)	0.640 (0.00)	0.441 (0.00)
r2	.02	.019	.0032	.021	.02	.035	.013
N	39	39	39	39	39	39	39

Note: P-values from 2SLS estimations in parentheses. First stage is shown in Table A4.

Table A8: IV estimations for MVPF calculations

Earnings	Years after program				
	(1) During	(2) First	(3) Second	(4) Third	(5) Fourth
Participated	-25688.2 ⁺ (13576.9)	17789.0 (27082.8)	37642.1 (33451.3)	23272.4 (37579.3)	-4680.4 (30279.5)
Women	-12360.5 (8882.3)	-44593.9* (17718.2)	-86135.0* (21856.1)	-78564.1* (24553.2)	-81944.5* (19783.7)
Constant	62218.8* (19047.0)	87323.5* (37994.4)	114263.9* (47063.0)	134037.7* (52870.7)	157935.3* (42600.6)
Observations	140	140	139	139	139
Social assistance	Years after program				
	(1) During	(2) First	(3) Second	(4) Third	(5) Fourth
Participated	-281.0 (2932.0)	-2597.1 (3819.1)	1856.7 (7031.4)	-7631.8 (8593.5)	-7982.6 (7318.5)
Women	1986.3 (1918.2)	5365.3* (2498.5)	11160.0* (4594.1)	10297.8 ⁺ (5614.7)	8146.0 ⁺ (4781.7)
Constant	7766.7 ⁺ (4113.3)	6947.8 (5357.8)	5770.3 (9892.5)	23593.8 ⁺ (12090.3)	21560.2* (10296.4)
Observations	140	140	139	139	139

Note: In each panel, earnings and social assistance are summarized over 1 year (April–March), except for the last year when it is April–December. Standard errors in parentheses. ⁺ $p < 0.10$, * $p < 0.05$

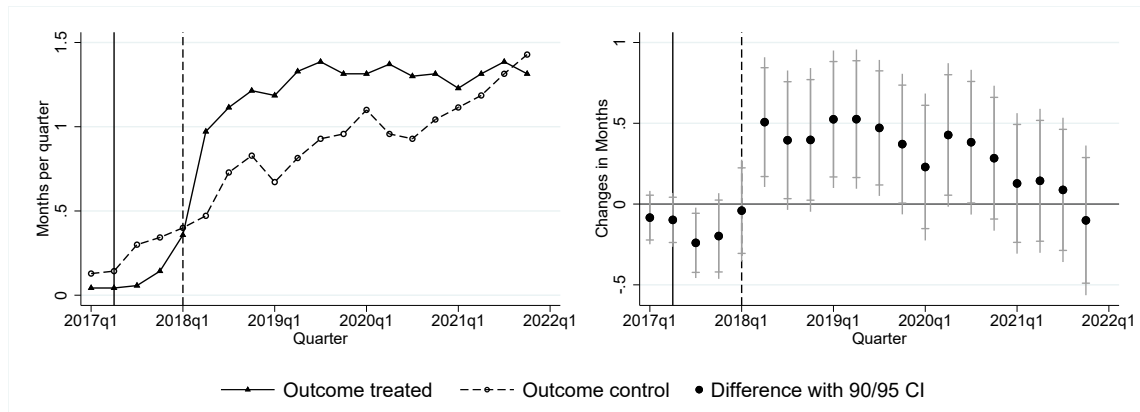
A.2 Sensitivity analysis: Definition of employment

In the baseline analysis, employment was defined as having *any* labor earnings above zero in a specific month. However, this is a quite generous measure as it includes individuals who have worked only a few hours. To check how sensitive the results are to the use of more conservative definitions of employment, we define employment as having monthly earnings above one monthly income base amount, above two monthly income base amounts, and above three monthly income base amounts, respectively, see Table A9 for the exact amounts.²⁹ As is clear from Figure A6, the different definitions of employment produces very similar patterns and results as in the baseline analysis (c.f. Figure 3).

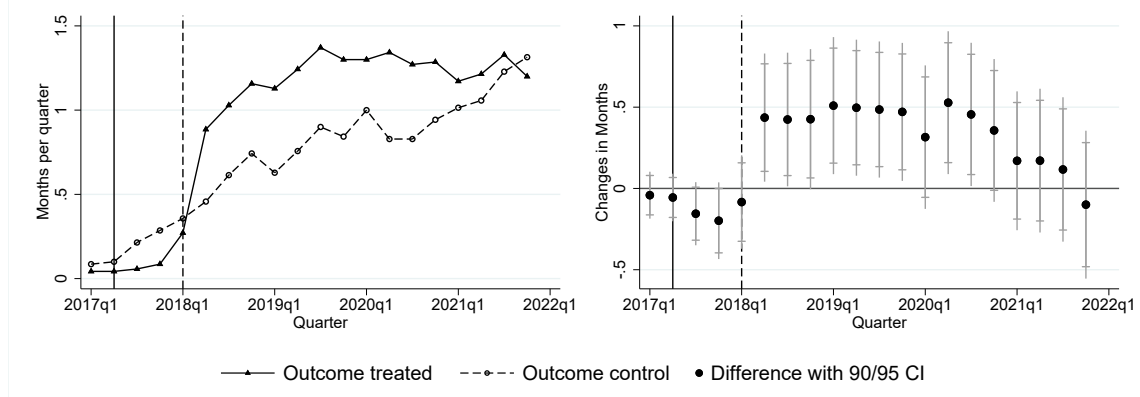
Similar results are also shown if we use earnings as an outcome directly, see Figure A7.

Finally, as part of our sensitivity analysis, we investigate whether our main results change when controlling for baseline characteristics, as seen in Figure A8. If anything, our estimated effects become somewhat larger, while the overall conclusions remain consistent.

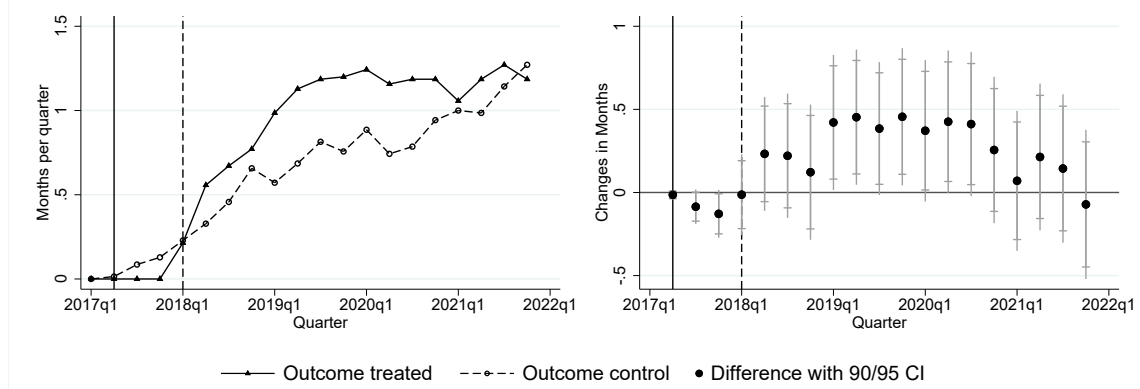
²⁹The income base amount, ”inkomstbasbelopp”, is an income-based index used to calculate the income pension in Sweden each year. To put the figures in perspective, it can be noted that in 2021 the average monthly salary for a person working in the restaurant sector was 24,800 SEK (see <https://www.scb.se/hitta-statistik/sverige-i-siffror/lonesk/Search/?lon=restaurant>), implying that three income base amounts in 2021 amount to almost 70 percent of the mean salary in the restaurant sector in 2021.



(a) Employment defined as earnings above 1 IBB during a month



(b) Employment defined as earnings above 2 IBB during a month



(c) Employment defined as earnings above 3 IBB during a month

Figure A6: Sensitivity analysis with different cut-offs for employment. Months per quarter.

Notes: ITT effects from OLS estimations controlling for gender.

Table A9: Income base amount different years, and the cut off we used for monthly incomes

Year	Yearly		Monthly	
	IBB	1 IBB	2 IBB	3 IBB
2017	61500	5125	10250	15375
2018	62500	5208.333	10416.67	15625
2019	64400	5366.667	10733.33	16100
2020	66800	5566.667	11133.33	16700
2021	68200	5683.333	11366.67	17050

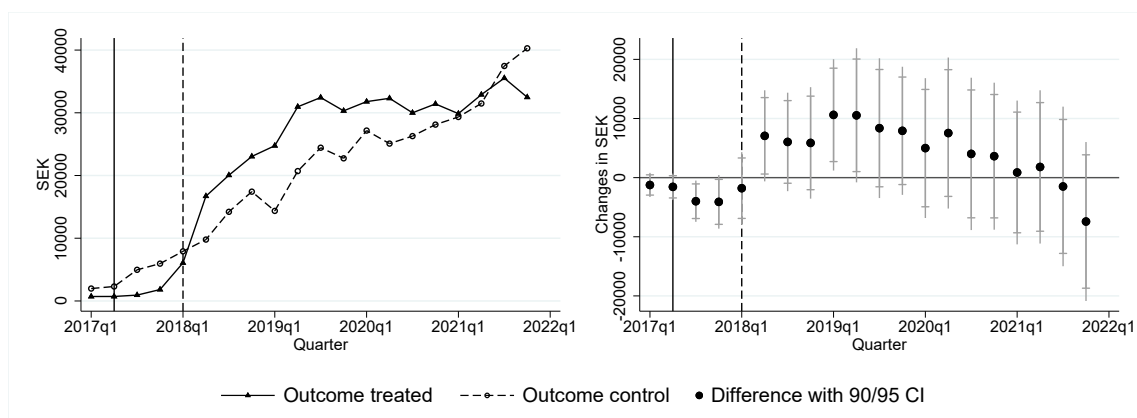


Figure A7: Sensitivity analysis earnings from employment, per quarter.

Notes: ITT effects from OLS estimations controlling for gender.

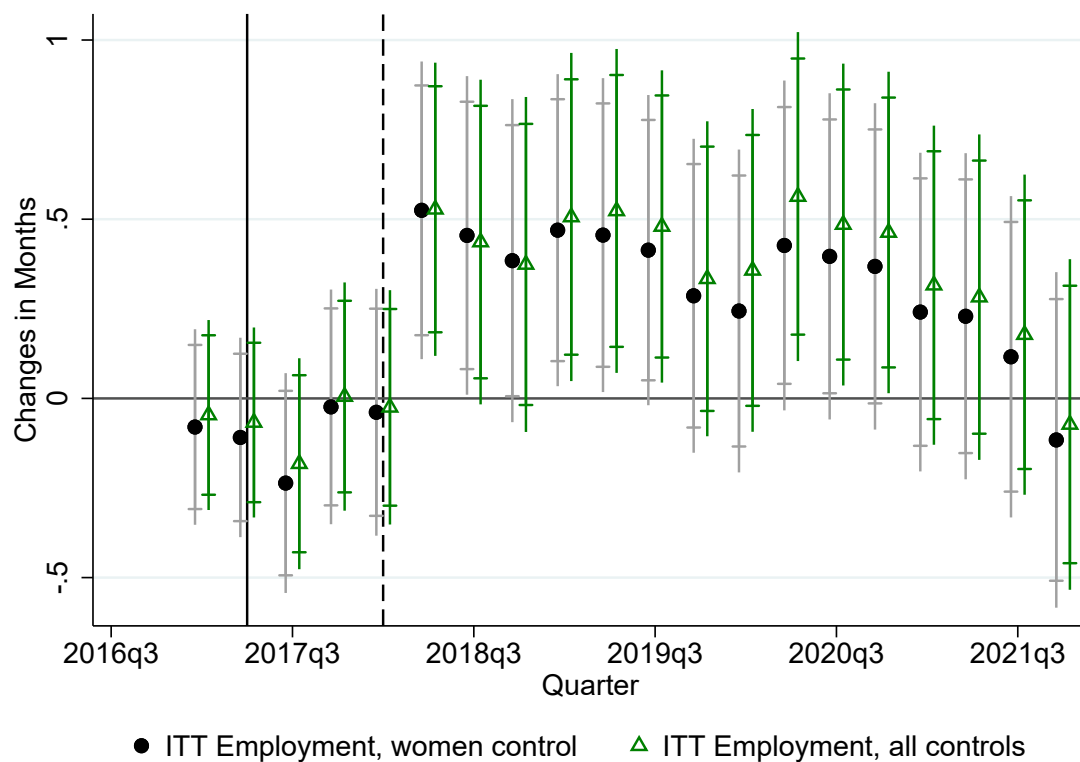


Figure A8: Sensitivity analysis also including additional control. Months per quarter.

Notes: ITT effects from OLS estimations controlling for gender (black dots) and additional pre-randomization variables, education, age, origin, time in Sweden and if residence permit was permanent (green triangles).

A.3 The questions in the survey

We base our questions on the multidimensional integration index (the IPL-index) developed by Harder et al. (2018). They develop two forms of the index, a short form (IPL-12) and a long-form (IPL-24). To increase the response rate in our group of low-educated refugees that are fairly new to Sweden, we choose to adopt the index with the fewest set of questions (which is the IPL-12 index). The IPL-index covers integration along six dimensions: psychological, linguistic, economic, political, social, and navigational. There are two questions per dimension in IPL-12. Our adaptation of the IPL-12 is as follows:

- *Psychological integration*

Here we use the same questions as in the IPL-12 index:

1. How connected do you feel with Sweden?³⁰
2. How often do you feel like an outsider in Sweden?³¹

- *Linguistic integration*

Here we use the same questions as in the IPL-12 index:

1. Please evaluate your own skills in Swedish. How well can you do the following when reading Swedish: I can read and understand the main points in simple newspaper articles on familiar subjects.³²
2. Please evaluate your own skills in Swedish. How well can you do the following when speaking Swedish: In a conversation, I can speak about familiar topics and express personal opinions.³³

- *Economic integration*

Since we have very good and detailed register information on how the individuals are doing on the labor market, we choose not to ask any questions on economic integration but rather make use of the register information. We use two sets of information from the registers. The first one is how many months the individual is employed or not and the other is whether the individual is on social assistance or not. We measure this by the time of the survey (in the fall, September–December, of 2021). Individuals are seen as more economically integrated the more months employed and the less months with social assistance they have. Distribution over month employed and with Social assistance are shown in Figure A9.

- *Political integration* Here we use the same questions as in the IPL-12 index:

1. How well do you understand the important political issues facing Sweden?³⁴

³⁰In Swedish: Hur stark tillhörighet känner du att du har till Sverige?

³¹In Swedish: Hur ofta känner du dig som en främling i Sverige?

³²In Swedish: Gör en egen bedömning av dina kunskaper i svenska språket. Vilken nivå har du när du läser svenska texter?

³³In Swedish: Gör en egen bedömning av dina kunskaper i svenska språket. Vilken nivå har du när du pratar svenska?

³⁴In Swedish: Hur bra förstår du de viktiga politiska frågor som rör Sverige?

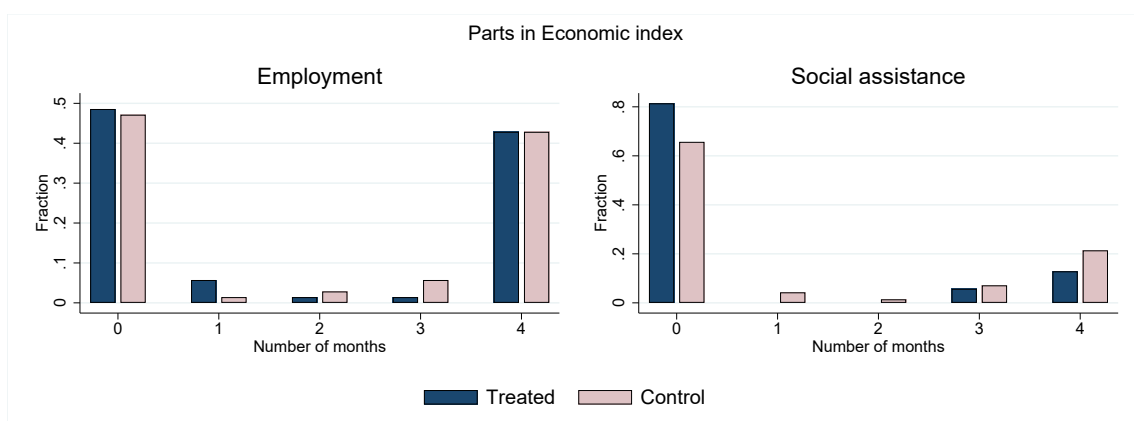


Figure A9: Distribution of months with employment and Social assistance included in the economic dimension of the integration index.

2. In the last 12 months, how often did you typically discuss major political issues facing Sweden with others?³⁵

- *Social integration:*

When it comes to Social integration one of the questions in IPL-12 concerns how many times the respondent have had dinner with natives during the last year. We asked the question during the Covid-19 pandemic and even if Sweden had less restrictions than many other countries, the population was prompted not to invite anyone outside their own family home. We therefore excluded that question and used other questions from the IPL-24 questionnaire. The questions we used to measure Social integration was thereby:

1. Please think about the Swedes in your address book or your phone contacts. With how many of them did you have a conversation - either by phone, messenger chat, or text exchange - in the last 4 weeks?³⁶
2. People in Sweden sometimes participate in different kinds of groups or associations in their leisure time. How often do you do that?³⁷
 - (a) If you think about the members of this group, how many of them are Swedes?³⁸

- *Navigational integration*

In the IPL-12, they ask if the respondent know how to make an appointment with a doctor and whether he or she knows how to find a vacant job to apply for. Since all our respondents took part in a labor market program organized by the PES in Sweden (making it likely that they know how to apply for a job), we choose to add a question on another navigational aspect

³⁵In Swedish: Hur ofta det senaste året har du diskuterat viktiga politiska frågor som rör Sverige med andra?

³⁶In Swedish: Tänk på de svenskar du har i din telefons kontakter. Hur många av dessa har du pratat, mejlat, chattat eller "SMS-at" med under de senaste 4 veckorna?

³⁷In Swedish: Personer i Sverige deltar ibland i olika typer av grupper eller föreningar på sin fritid. Hur ofta gör du det?

³⁸In Swedish: Hur många är svenskar i de grupper eller föreningar du deltar i?

that is important in the Swedish context: if they know how to find/apply for an apartment or house to live in.³⁹ The two questions we use to construct the index is hence given by:

1. In Sweden, do you know how to make an appointment with a doctor?⁴⁰
2. In Sweden, do you know how to find an apartment or a house to live in?⁴¹

A score of 1-5 is given on each question, where a higher score implies that the respondent is considered as more integrated. An integration index for each of the six dimension of integration, as well as for the overall IPL-12 Index, is given by rescaling the total sum to a score between 0 and 1.

³⁹Sweden has a regulated rental housing market, which means that it is in principle only possible to get rental housing by signing up to the municipal housing queue. Knowledge on how to do this is very important in order to get housing in Sweden.

⁴⁰In Swedish: I Sverige, vet du hur du skall gå till väga för att boka en tid hos en läkare?

⁴¹In Swedish: I Sverige, vet du hur du skall gå till väga för att söka efter en lägenhet eller ett hus att bo i?