



International
Labour
Organization

EMPLOYMENT
PROMOTION

IMPACT EVALUATION OF ACTIVE LABOUR MARKET PROGRAMS IN FYR MACEDONIA: KEY FINDINGS

Nikica Mojsoska-Blazevski
Marjan Petreski

Impact evaluation of active labour market programs in FYR Macedonia: key findings

**Nikica Mojsoska-Blazevski
Marjan Petreski**

Copyright © International Labour Organization 2015

First published 2015

Publications of the International Labour Office enjoy copyright under Protocol 2 of the Universal Copyright Convention. Nevertheless, short excerpts from them may be reproduced without authorization, on condition that the source is indicated. For rights of reproduction or translation, application should be made to ILO Publications (Rights and Licensing), International Labour Office, CH-1211 Geneva 22, Switzerland, or by email: rights@ilo.org. The International Labour Office welcomes such applications.

Libraries, institutions and other users registered with a reproduction rights organization may make copies in accordance with the licences issued to them for this purpose. Visit www.ifrro.org to find the reproduction rights organization in your country.

ILO Cataloguing in Publication Data

Mojsoska-Blazevski, Nikica; Petreski, Marjan

Impact evaluation of active labour market programs in FYR Macedonia : key findings / Nikica Mojsoska-Blazevski, Marjan Petreski ; International Labour Organization, ILO DWT and Country Office for Central and Eastern Europe. - Budapest: ILO, 2015

ISBN: 9789221298175; 9789221298182 (web pdf)

International Labour Organization; ILO DWT and Country Office for Central and Eastern Europe

labour market / employment / impact evaluation / methodology / data collecting / Macedonia, former Yugoslav Republic

13.01.2

The designations employed in ILO publications, which are in conformity with United Nations practice, and the presentation of material therein do not imply the expression of any opinion whatsoever on the part of the International Labour Office concerning the legal status of any country, area or territory or of its authorities, or concerning the delimitation of its frontiers.

The responsibility for opinions expressed in signed articles, studies and other contributions rests solely with their authors, and publication does not constitute an endorsement by the International Labour Office of the opinions expressed in them.

Reference to names of firms and commercial products and processes does not imply their endorsement by the International Labour Office, and any failure to mention a particular firm, commercial product or process is not a sign of disapproval.

ILO publications and digital products can be obtained through major booksellers and digital distribution platforms, or ordered directly from ilo@turpin-distribution.com. For more information, visit our website: www.ilo.org/publns or contact ilopubs@ilo.org.

Contents

- 1. Introduction..... 5**
 - 1.1 Labour market developments 6
- 2. The impact evaluation methodology and survey design..... 9**
 - 2.1 Sample 9
 - 2.2 Survey questionnaire and outcome variables..... 12
 - 2.3 Impact evaluation methodology..... 13
- 3. Descriptive analysis of the sample and the outcome indicators 15**
 - 3.1 Characteristics of the respondents 15
 - 3.2 Main outcome indicators..... 18
 - 3.3 Additional indicators 20
- 4. Impact evaluation: results and discussion 22**
 - 4.1 Self-employment program..... 23
 - 4.1.1 Summary of findings and recommendations 30
 - 4.2 Internship program 32
 - 4.2.1 Impact of 2010 Internship..... 32
 - 4.2.2 Impact of 2012 Internship..... 36
 - 4.2.3 Has Internship 2012 led to further gains over Internship 2010? 39
 - 4.2.4 Summary of findings and recommendations 43
 - 4.3 Training for known employer..... 44
 - 4.3.1 Impact of 2010 TKE 44
 - 4.3.2 Impact of 2012 TKE 48
 - 4.3.3 Has TKE 2012 led to further gains over TKE 2010? 51
 - 4.3.4 Summary of findings and recommendations 53
 - 4.4 Wage subsidy 54
 - 4.4.1 Summary of findings and recommendations 56
 - 4.5 Training for deficient occupations..... 58
 - 4.5.1 Summary of findings and recommendations 61
 - 4.6 Training in advanced IT skills..... 61
 - 4.6.1 Summary of findings and recommendations 65
- 5. Conclusion 66**
- References 72**
- Annex 1 73**

List of tables and figures

Table 1-2 Employment rates by gender, age and education, 2009-2014.....	8
Table 1-3 Unemployment rates by gender, age and education, 2009-2014	8
Table 2-1 Programs, size of the groups and control group selection.....	10
Table 2-2 Sample size and response rates.....	11
Table 2-3 Main reasons for non-responding	12
Table 3-1 Demographic characteristics of treatment and control group.....	16
Table 3-2 Labour market characteristics before program's entry/cut-off point, %	17
Table 3-3 Labour market status	18
Table 3-4 Self-assessment of the changes in employment prospects and the financial situation, %	19
Table 3-5 Employment characteristics of the control group and program participants, (today)	20
Table 3-6 Wage-employment characteristics of the control group and program participants.....	21
Table 3-7 Earnings, control group and program participants, %.....	21
Table 4-1 Means' comparison – Self-employment program	24
Table 4-2 Effects - Self-employment program	25
Table 4-3 Employment transition, by type of employment	26
Table 4-4 Means' comparison – Self-employment program: exercise with a 'general' control group	27
Table 4-5 Matching – Self-employment program: matching with a 'general' control group	28
Table 4-6 Effects – Self-employment program when compared to a 'general' control group.....	29
Table 4-7 Employment transition, by type of contract	30
Table 4-8 Means' comparison – Internship 2010 program.....	33
Table 4-9 Matching – Internship 2010 program	34
Table 4-10 Effects – Internship 2010 program	35
Table 4-11 Means' comparison – Internship 2012 program.....	36
Table 4-12 Matching – Internship 2012 program	37
Table 4-13 Effects – Internship 2012 program	38
Table 4-14 Means' comparison – Internship 2012 versus Internship 2010 programs	40
Table 4-15 Matching – Internship 2012 with Internship 2010 program	41
Table 4-16 Effects – Internship 2012 over Internship 2010.....	42
Table 4-17 Means' comparison – TKE 2010 program	45
Table 4-18 Matching – TKE 2010 program.....	46
Table 4-19 Results – TKE 2010 program	47
Table 4-20 Means' comparison – TKE 2012 program	48
Table 4-21 Matching – TKE 2012 program.....	49
Table 4-22 Effects – TKE 2012 program.....	50
Table 4-23 Means' comparison – TKE 2010 versus TKE 2012 programs	52
Table 4-24 Effects – TKE 2012 over TKE 2010 program	53
Table 4-25 Means' comparison – Wage subsidy programs.....	54
Table 4-26 Matching – Wage subsidy program	55
Table 4-27 Effects – Wage subsidy program.....	57
Table 4-28 Means' comparison – Training for deficient occupations program.....	58
Table 4-29 Matching – Training for deficient occupations program	59
Table 4-30 Effects – Training for deficient occupations program	60
Table 4-31 Means' comparison – Training for advanced IT skills program	62
Table 4-32 Matching – Training for advanced IT skills program.....	63
Table 4-33 Effects – Training in advanced IT skills program	64
Table 5-1 Employment outcomes of each program – summary table.....	67
Table A1.1: Explanatory variables included in the preferred specification of the regression model	73

1 Introduction

This report presents the findings of the impact evaluation carried out on a selection of active labour market programs implemented by the Employment Service Agency (ESA) of Macedonia during the period 2008-2012. We evaluate the following six programs: program for self-employment (implemented in 2008), internship (implemented in 2010 and 2012), training for known employer (2010 and 2012), wage subsidy program (2010), training for advanced IT skills (2010) and training for deficient occupations (2010). The choice of the years was based on the intention to assess not just the short-term (immediate) effectiveness of the program, but also the longer-term effects. Moreover, for two of the programs, we evaluate two years of program implementation, so as to see if some changes made within the program brought better effects.

The key research question is whether participation in the active labour market programs increased the probability of participants to find and retain gainful employment. However, in addition to this main outcome, we include other outcome variables such as: inactivity, unemployment, wages, changes in the prospects for employment after the program, changes in financial status, subjective labour market status, etc.

The process of planning, design and implementation of active labour market programs (ALMPs) in Macedonia has been streamlined from 2007 onwards. The planning and implementation of the programs is organized through annual Operational Plans for ALMPs (OPs) (the exception was the two-year OP for 2012-2013). OPs are following the objectives, principles and target groups set in the strategic documents, the National Employment Strategy and the National Action Plan for Employment. The OP sets the scope and financial allocation of ALMPs (the employment services are not directly treated as active programs). The Plan is managed by the national director, that is the Minister of Labour and Social Policy, and coordinated by the national coordinator – the director of the ESA. The ESA is the implementing body for the OPs.

The spending on ALMPs in Macedonia is comparable to that in the countries in the Western Balkan region, and to some of the EU Member States (for instance, Slovakia and Estonia), although being at a generally low level (below 0.1% of GDP)¹. If spending is compared to the overall unemployment (active jobseekers), in 2013 Macedonia spent about EUR 50 per jobseeker, whereas lowest spenders among the EU countries in 2012 were Romania (EUR 53 per jobseeker), Croatia (EUR 210) and Bulgaria (EUR 214).

Throughout years, as the country (policymakers) increased its experience with the design and implementation of ALMPs, there has been a growing need (both nationally and from the international community) for monitoring and evaluation of the active programs. The assessment of the effectiveness of active measures is regularly conducted through employment outcomes of different programs/interventions. However, such assessment fails to take into account several important elements: the cost-effectiveness of the programs; potential dead-weight loss; substitution effect; comparison of the employment outcomes with unemployed individuals who have not participated in the program; etc.² Only after consideration of these elements, an assessment will show whether the programs are bringing “value for money”, i.e. whether taxpayers’ money are used effectively. There were so far few attempts to provide a more rigorous evaluation of the ALMPs in Macedonia. In this regard, in 2012, the ILO conducted a performance measurement of three active programs: self-employment program, training for known employer and wage subsidies (see Corbanese, 2012). In 2012, an evaluation has been conducted of the self-employment program (2008-2011) implemented by the UNDP (see O’Higgins and Kirevska, 2012). Although those two studies were very useful for the policymakers in terms of evidence-based policymaking, they still fail to take full account of the previously mentioned elements. The current, impact evaluation study, is the first of such type implemented in Macedonia, and by domestic experts.

1 In 2009, as part of the anti-recession package of the Government, the spending on ALPMs was increased to 0.2% of GDP, mainly through increased scope and funding of public works.

2 The substitution effect indicates whether improved employment prospects for the participants in an ALMP comes at the expense of worsened employment prospects for the non-participants. Dead-weight loss represents a situation when a participant obtains employment through participation in ALMP that s/he would have still got without the program.

1.1 Labour market developments

The environment in which active programs operate is very important for their overall effectiveness. In particular, many studies show that ALMPs are less effective (if at all) in a situation of low labour demand (Betcherman et al., 2004). On the other hand, some active programs (such as subsidised employment and public works) can contribute towards bringing greater demand for workers, hence increasing overall employment.

The labour market in Macedonia is characterized with relatively high unemployment which is on a declining path for almost a decade, but at a slow pace. In the second quarter of 2014 (latest available data), unemployment rate was 28.2% (see Table 1-1). At the same time, there has been a steady path of increase in the employment rate, from 37.8% in Q1-2009 to 41.1% in Q2-2014.

Table 1-1 Main labour market indicators, 2009-2014

Time	Activity rate	Employment rate	Unemployment rate
Q1, 2009	56.2	37.8	32.7
Q2	57.0	38.8	31.9
Q3	57.3	39.2	31.7
Q4	56.1	37.9	32.4
Q1, 2010	56.3	37.5	33.5
Q2	56.0	38.1	32.1
Q3	57.6	39.3	31.7
Q4	57.8	39.9	30.9
Q1, 2011	57.1	39.3	31.2
Q2	56.6	38.8	31.3
Q3	56.9	39.1	31.2
Q4	56.5	38.5	31.8
Q1, 2012	56.4	38.6	31.6
Q2	56.5	38.8	31.2
Q3	56.3	39.1	30.6
Q4	56.7	39.3	30.6
Q1, 2013	57.0	40.0	29.9
Q2	57.1	40.7	28.8
Q3	57.3	40.8	28.7
Q4	57.3	40.9	28.6
Q1, 2014	57.3	41.0	28.4
Q2	57.3	41.1	28.2
Source: State Statistical Office. Data for population aged 15-79.			

Table 1-2 shows the employment rate by different characteristics of workers for the population aged 15-64. Apparently, males have much higher employment rate than females: the employment rate of males is by 49% higher than that of females. Young workers have very low employment rate of 14.4%, whereas the employment rate of prime-age workers (aged 25-49) is 60%. Workers with tertiary education have more than double the employment rate of primary-educated workers.

Females in Macedonia have higher unemployment rates compared to males, although the difference is relatively small (29% for females and 27.8% for males in Q2-2014) (see Table 1-3). Moreover, as Petreski et al. (2014) argue, females on average receive lower wages than males with same characteristics by 12.5%, although this gender wage gap disappears at tertiary education. One in two young persons (aged 15-24) searching for a job cannot find one. In addition, unemployment declines with education such that the unemployment rate of workers with completed primary education or less in Q2-2014 was 34%, as compared to 23.7% for individuals with completed tertiary education.

Table 1-2 Employment rates by gender, age and education, 2009-2014

Categories	2009				2010				2011				2012				2013				2014	
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
All	42.5	43.8	44.1	42.7	42.2	42.9	44.2	44.8	44.3	43.7	44.2	43.5	43.6	43.7	44.1	44.3	45.2	46.1	46.3	46.4	46.7	46.8
Gender																						
Males	52.6	53.1	53.6	51.8	51.5	52.7	53.0	53.9	52.2	52.2	52.2	52.5	51.3	52.2	52.7	53.3	52.6	54.1	55.3	55.9	56.0	55.8
Females	32.1	34.3	34.4	33.4	32.7	32.7	35.3	35.5	36.1	35.0	36.0	34.3	35.8	35.1	35.2	35.1	37.6	37.8	37.0	36.8	37.2	37.5
Age																						
15-24	16	16.1	16.2	14.6	14.6	14.6	16.5	16.1	14.9	14.3	14.6	13.6	14	14.9	16.9	16.1	15	15.7	16.6	17.5	15.4	14.4
25-49	54.3	56.5	56.4	54.5	54.8	55.5	55.6	57.2	57	56.2	57.1	55.8	56.6	55.5	55.2	55.8	57.6	58.6	57.6	58.1	59.6	59
50-64	41.3	41.9	43.4	42.9	40.2	41.3	44.8	44	43.1	43	42.7	43.4	41.4	43.2	43.3	43.5	43.9	44.5	46.3	45.1	44.9	46.6
Education																						
Primary and less	27.1	27.8	28.2	25.5	25.2	25.6	27.5	28.4	27.9	27.0	27.6	26.9	26.4	26.6	24.9	25.1	27.3	28.6	28.6	29.0	29.4	28.8
Secondary	49.6	51.3	49.8	49.0	49.1	50.5	49.9	50.2	49.8	51.1	48.8	48.1	48.7	49.6	50.6	51.6	51.3	52.5	53.0	52.7	52.5	53.3
Tertiary	70.5	68.8	70.9	72.3	71.2	70.3	71.1	70.2	68.3	67.0	70.1	68.9	68.8	68.3	68.6	66.6	67.9	68.2	67.3	66.6	69.1	69.2

Source: State Statistical Office.

Table 1-3 Unemployment rates by gender, age and education, 2009-2014

Categories	2009				2010				2011				2012				2013				2014	
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
All	32.9	32.0	31.9	32.6	33.6	32.3	31.9	31.2	31.4	31.6	31.3	31.9	31.7	31.4	30.8	30.8	30.1	28.9	28.9	28.8	28.5	28.3
Gender																						
Males	31.9	31.5	31.7	32.9	33.7	31.6	32.3	30.9	32.1	31.9	32.3	31.5	32.8	32.1	31.3	30.3	30.9	29.5	27.9	28.2	27.8	27.8
Females	34.7	32.7	32.2	32.2	33.6	33.4	31.2	31.8	30.4	31.1	29.9	32.5	30.0	30.3	30.1	31.7	28.8	27.9	30.4	29.7	29.5	29.1
Age																						
15-24	54.4	54.9	54.6	56.6	57.1	55.0	51.9	50.5	52.5	54.6	54.2	59.4	55.9	54.9	52.1	53.0	53.5	51.7	52.2	50.3	54.6	55.2
25-49	30.8	29.3	30.2	30.8	31.3	30.0	31.2	30.3	29.6	29.8	29.3	29.5	29.8	30.3	29.5	29.9	28.3	27.2	27.7	27.4	26.5	27.2
50-64	27.1	27.0	24.1	25.4	28.1	27.5	23.9	24.7	26.9	26.1	26.6	25.2	26.5	23.9	24.1	23.2	24.8	23.6	21.7	22.7	22.2	20.4
Education																						
Primary or less	39.3	39.3	37.9	40.5	43.5	41.2	38.1	35.9	36	39.4	39.6	37.8	37.7	36.8	39.6	40.4	36.1	34.8	35.4	32.7	32.7	33.9
Secondary	32.7	30.9	32.5	33.1	33	31.2	32.4	31.8	32.3	30.2	31.1	32.8	33	32.2	30.6	29.7	30.2	28.1	28.1	28.7	29	27.6
Tertiary	22	23	20.8	19.3	20.1	22	21.8	23.3	23.4	24.2	21.8	22.8	21.2	23.2	22	23.2	22.8	23.7	23.1	24.4	22	23.7

Source: State Statistical Office.

2 The impact evaluation methodology and survey design

The main objective of this report is to evaluate the effectiveness and efficiency of the active labour market policies in Macedonia. For this purpose, we measure the differences in labour market outcomes (employment at the time of survey and employment at any time between the end of the program and the survey date) and subjective wellbeing outcomes (subjective evaluation of the change in the financial situation and chances to find a job before and after program participation) between those who participated to the programs (*treatment group*) and those who did not (*control group*).

For a more precise estimate of program effects, it is necessary to “compare the comparable” (Heckman et al. 1999). This means that program participants need to be compared only to those non-participants who could have participated in the program (i.e. had an equal chance to be selected for participation as those who were actually treated). For four of the six ALMPs – self-employment, internship (2010 and 2012), wage subsidy and training for advanced IT skills – the control group has been chosen out of those who applied on the ESA open call, but who were ultimately non-selected. This is plausible, given that they not only satisfy the conditions of the call, but also they may be similar to the selected participants based on some unobservable characteristics (such as, motivation to participate in a program, or persuasion that program participation is beneficial). According to ESA, after all applicants passed through all eligibility filters/criteria for participation in a particular program, the selection into actual program participation is done randomly³. The exception is the self-employment program, whereby participants were selected based on their score obtained for the entrepreneurship test. However, even in this latter case, we dropped selected participants with a test score above 75, as well those non-selected with a score below 45 (60 being the threshold to proceed to the next level). The control groups for the two remaining programs – training for known employer and for deficient occupations were selected randomly from a sample obtained from ESA from persons who satisfy the conditions of for participation of the program “cleaned up” of those who during the interview claimed they knew of the particular ESA program, but did not apply because believed it would not change their employment prospects (i.e. their choice not to non-self-select has been correlated with the outcomes we evaluate here). Out of such defined potential control groups, the particular control groups are subsequently selected by means of a matching approach (see Section 2.3). The following section describes the main methodological problem addressed in constructing the treatment and the control group.

2.1 Sample

In order to be able to identify the causal impact of the ALMPs (namely, the intervention) we created credible control groups through matching non-participants with the participants on relevant observable characteristics. By doing so, the comparison of the employment, wage prospects and subjective well-being of the treatment and the control group enables us to isolate the impact of the program on those prospects. The assumption behind this is that the national labour market environment has the same effect on the outcomes for the participants in the programs and for the non-participants.

Table 2-1 presents information on the number of participants in each program (column 3) and the corresponding control sample frame (column 4). To ensure as much as possible precision in the evaluation, we survey the whole population of the program participants. Given that the evaluation includes a range of programs, years and variations in the characteristics of the treatment groups, the selection of the control groups has been specifically designed for each program. In general, there were two patterns of selecting a control group: i) from persons that applied for participation in the program, but were not selected, and ii) a so-called general control groups which were selected from the ESA registry based on the eligibility criteria for the specific program. The former was specifically chosen as to increase the probability that the individuals in the treatment and control groups are similar not just based on observable characteristics (which will be

3 One example being the case when an employer asks for an interns of certain profile. In such case, ESA sends to the employer all the CVs which match the request. The employer chooses an applicant from the CVs. Even if the employer makes his decision by observable characteristics, we address the selection on observables in this analysis, and hence this does not impose a bias. If, however, the employer makes the decision based on unobservables, this would impose a bias, which we do not address. However, even if the employer imposes unobservable factors into the decision, we believe this is not a general case.

ensured through the matching approach), but also on unobservable characteristics such as motivation for work, motivation to participate in a program, persuasion that program participation is beneficial, etc.

Table 2-1 shows the size of the treatment and control groups as well as details on the selection of control groups

Table 2-1 Programs, size of the groups and control group selection

Program	Year	Treatment group size	Control group size	Selection method for control group
Self-employment	2008	421	513	Participants that applied to the program but were not selected for participation and whose score on the entrepreneurship test ranged between 45 and 60 (60 being the threshold for selection). Similarly, all participants (treated) who had scored above 75 are not considered.
Internship	2010 2012	199 99	210 204	Persons who applied (hence satisfied the eligibility criteria), but were not selected. Here, the selection into treatment was fully random, as was driven by the demand for interns, which cannot be assumed to have been sector-specific or profile-specific.
Wage subsidy	2010	1,028	2,474	Persons who applied (hence satisfied the eligibility criteria), but were not selected. Here, the selection into treatment was fully random by the ESA.
Training in advanced IT skills	2010	81	112	Persons who applied (hence satisfied the eligibility criteria), but were not selected. Here, the selection into treatment was fully random by the ESA
Training for known employer	2010 2012	229 341	8,621 79,828	The selection into control groups was based on few criteria: i) persons who were unemployed at a cut-off date; ii) from employment offices where the program was running, iii) have education level based on the eligibility criteria, and iv) have educational (i.e. occupational) profiles similar to the program participants.
Training for deficient occupations	2010	54	18,530	The selection into control groups was based on few criteria: i) persons who were unemployed at a cut-off date for at least 6 months; and at least 3 months (for youth below the age of 27) ii) persons from Skopje where the program was running iii) educational level based on eligibility criteria (secondary general and vocational) iv) their occupation belongs to the four occupations identified as deficient by the ESA Action Plan (i.e. matches with the occupations of those treated).
* The cut-off date has been set at June, 2010 for all programs, except for self-employment (May 2008) and for Training for known employer (June, 2012). These cut-off dates were set around the time the selection of the participants was conducted.				

In total, we have data for 2,485 program participants from ESA for the six programs we evaluate. The sample frame for the control groups is larger, consisting of 111,566 individuals, from all control groups. The expected response rate differs across the programs and is mainly related to the size of the treatment group and the method used for selection of the control group.

Table 2-2 shows the planned samples size and achieved response rates. Apparently, the actual response rates are lower than the planned ones. The response rates are relatively low for the following programs: training for known employer 2010, where the response rate for the treatment group is 44% of the targeted sample, and for the treatment group for training in advanced IT skills, where the response rate is 51% of the targeted sample. On the other hand, for two programs – training for known employer (2010 and 2012) and for deficient occupations - the control groups were chosen from a “general” control group, being very large, so a low response rate was actually targeted and planned.

Table 2-2 Sample size and response rates

Type of program	Database from ESA		Targeted sample size		Actual sample size		Response rates relative to overall database		Response rates relative to targeted samples	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
SE	421	513	295	359	192	240	46%	47%	65%	67%
IN10	199	210	149	158	112	113	56%	54%	75%	72%
IN12	132	1,278	99	256	78	172	59%	13%	79%	67%
WS	1,028	2,474	308	619	284	603	28%	24%	92%	97%
IT	81	112	69	95	35	59	43%	53%	51%	62%
TKE10	229	8,621	172	259	75	263	33%	3%	44%	102%
TKE12	341	79,828	171	239	113	235	33%	0%	66%	98%
TDO	54	18,530	51	130	28	130	52%	1%	55%	100%
Total	2,485	111,566	1,314	2,114	917	1,815	37%		70%	86%

Source: Own calculations based on survey data.

The most frequent reason for the low response rates in the particular cases is the inability to reach a person, due to lack of correct contact information. Few attempts were made in most of the cases of non-response, and besides the mobile phones, the data collection agency has made efforts to find the missing persons through their home addresses. But, even that strategy did not bring large results. The rate of refusal to participate in the survey is actually between 0.2% and 22% (see Table 2-3). Highest percentages of the rejections to participate in the survey for the treatment group are present in the case of training in deficient occupations and internship program 2012. On average, the rejection happened in about 13% of the cases, suggesting that still the possible attrition bias is constrained in our case. There is a quite large share of persons who could not be reached by provided mobile phones, neither through their home addresses. What is surprising is that these shares are larger for the treatment than for the control group, given that one expects that the ESA keeps more reliable information for the program participants than for the other registered unemployed persons.

Table 2-3 Main reasons for non-responding

Type of program	Sample (respondents)		Persons who rejected		Persons who were unreachable		Rejection rate		Rate of unreachable	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
SE	192	240	3	16	226	256	0.7%	3.1%	53.7%	49.9%
IN10	112	113	7	28	78	68	3.5%	13.3%	39.2%	32.4%
IN12	78	172	26	18	28	66	19.7%	1.4%	21.2%	5.2%
WS	284	603	4	4	20	12	0.4%	0.2%	1.9%	0.5%
IT	35	59	14	19	22	30	17.3%	17.0%	27.2%	26.8%
TKE10	75	263	21	0	132	0	9.2%	n/a	57.6%	n/a
TKE12	113	235	32	0	196	0	9.4%	n/a	57.5%	n/a
TDO	28	130	12	0	14	0	22.2%	n/a	25.9%	n/a
Total	917	1815	119	85	716	432	13.0%	7.2%	78.1%	36.4%

Source: Own calculations based on survey data.

2.2 Survey questionnaire and outcome variables

Data were collected through questionnaires designed by the researchers, based on an ILO template and the questionnaires used in a similar study implemented by the FREN, in Serbia. Broadly, there are two questionnaires, one for the control groups and one for the treatment groups. However, the questionnaire for the control groups slightly differed depending on whether we had the control groups out of those who applied for the program, but were not selected, or a 'general' control group. The questionnaire for both types of control groups included questions on the employment history from the cut-off point (time of the selection to the program) until the time of the survey. For the general control group, the questionnaire also included control questions which tested the motivation to seek for work, availability for work, willingness to participate in the program and socio-demographic characteristics. Here, the strategy has been that we dropped the individuals (i.e. stopped the interview) if an individual was informed about the program but did not participate because s/he did not believe that the program will increase hers/his chances to find a job (i.e. the reason for non-participation was correlated with the outcomes we observe). The questionnaire for the treatment group is quite similar to that of the control groups. In addition to the questions for the control group, program participants answered questions related to their subjective assessment of the programs' usefulness for their future employment.

Draft questionnaires were sent for comments to the stakeholders, pre-tested on the pilot sample and adjusted in line with the comments received. Final questionnaire(s) were administered to the selected sample of the treatment and the control groups.

The outcome variables examined in this report are based on the survey questions. The study is based on the following outcome variables:

1. *Employment (i.e. current employment)*: The employment is defined on the basis of the ILO definition, that is all individuals who, in the reference week, performed some work for at least one hour for a remuneration (in cash or in-kind) and employed individuals who in the reference week were absent from work. To this definition, we added farmers and contributing family members.
2. *Non-employment-at-any-time*: Individuals who were unemployed (according to the above definition) at any time after the program's end (including those currently unemployed).

3. *Employed at program's end but currently unemployed.* Individuals who were employed at program's end (and certain period thereafter), but are currently unemployed.
4. *Inactivity.* Individuals who have not searched for a job in the last four weeks.
5. *Hourly wage:* Wage per hour earned on the current job.
6. *Changes in the prospects of employment after program participation /cut-off point.* This indicator is based on the subjective assessment of the respondent. Respondents rated the level of change on a three-point scale, from 1 ("Prospects are better") to 3 ("Prospects are worse").
7. *Changes in financial status after program participation/cut-off point* – based on the subjective assessment of the respondent. Respondents rated the change on a three-point scale, ranging from 1 ("Financial situation is better") to 3 ("Financial situation is worse").
8. *Subjective labour market status:* This is a self-assessment of the interviewed persons on their current labour market status. The status includes: employed, seasonal/temporary work, self-employed, trying to start-up a business, searching for a job, non-employed but not searching for a job, and involved in education/training.

Aside these main indicators, the survey also provided information on other labour market characteristics of program participants and the control group:

1. *Employment status:* wage-employment, self-employment and contributing family members;
2. *Informal employment:* individuals working in private unregistered business, or working in a registered business without an employment contract, including contributing family members;
3. *Ownership of enterprises* where employed individuals worked (private and public);
4. *Type of contract:* permanent or temporary;
5. *Wage levels.*

The main indicators are examined in both the descriptive (Section 3) and the econometric analysis (Section 4), while the additional indicators are used in Section 3 only.

2.3 Impact evaluation methodology

Any impact evaluation research has to deal with the problem of the counterfactual. This arises because it is impossible to directly observe a single individual in two different statuses (participation and non-participation in a program). Therefore, the main task of an impact evaluation study is to find a valid estimate of the counterfactual.

There are two methods to estimate the counterfactual: randomized experiments and non-experimental (also called quasi-experimental) methods. In principle, randomized experiments provide the most robust method to construct the counterfactual. In randomized experiments, individuals eligible for participation are randomly assigned to the treatment and control group. Since these two groups do not differ from each other (on average) both in observable and unobservable characteristics (i.e. the control group can be considered as "identical" to the treatment group), the average difference in outcomes between the two groups provides a simple answer to the counterfactual question. Often, however, randomized experiments are politically or socially unfeasible and they are not entirely free of estimation difficulties (Heckman et al., 1999).

The ESA-supported ALMPs were not designed as randomized experiments, which substantially lowered the chances to obtain *ex post* a control group with the same average characteristics as the treatment group. Still, the choice of a control group from those who applied (and hence were eligible for the program), but did not participate could mimic a natural experiment and the possibility of finding the treatment and the control group with essentially the same average characteristics was not excluded *a priori*.

However, a more realistic assumption would be that – if additional characteristics did play a role in determining the chances to participate in the ALMPs – one could not consider the treatment and the control group as “identical”. In this case, a simple comparison of mean outcomes (such as employment rates) between the two groups would be insufficient. The evaluation may also give biased results in case of substantial differences between the number of planned and accomplished interviews (i.e. low response rates) in both groups, since the selection of the control group was based on planned, rather than on accomplished interviews.

To assess whether program participation could be regarded as quasi-random, the characteristics of participants and non-participants were compared. Initially, statistical tests of the hypothesis of random assignment to participation were performed (i.e. random differences between the treatment and control group). In particular, we tested statistically whether the means of important socio-demographic characteristics and labour market outcomes were significantly different between treatment and control group. If the hypothesis of random assignment is rejected, it may be actually misleading to compute program's net effects as the difference in the average outcomes between participants and non-participants.

A common technique to solve the evaluation problem when participants and non-participants are not randomly assigned to a labour market program is the matching approach. This approach mimics a randomized experiment *ex post* by constructing a control group that resembles the treatment group as closely as possible. After matching, the probability to be selected for participation in the program of the control group' individuals, on the basis of their observable characteristics, is comparable to the probability of the individuals from the treatment group.

In the dataset there are many variables that presumably influence both the selection into the program and labour market outcomes. Hence, it appears reasonable to assume that selection into the program and labour market outcomes are independent conditional on these observables (the so-called conditional independence assumption). Under this assumption we apply one-to-one *nearest neighbour matching with replacement* and *the nearest neighbour matching with caliper*. Both approaches consist of two steps: (i) an estimation of the individual probabilities to participate in the program, depending on a set of observable characteristics; (ii) matching of participants and non-participants on the basis of these estimated probabilities. One-to-one matching means that each member of the treatment group is matched with a single member from the control group. Nearest neighbour matching means that the pairs are matched according to the minimum distance of the predicted probabilities of program participation, matching with replacement means that the data on individuals in the control group may be used more than once, provided that they are the nearest neighbour of an individual in the treatment group, and finally, matching with caliper means that control's propensity score belongs to a pre-defined radius.

3 Descriptive analysis of the sample and the outcome indicators

This section presents a descriptive analysis and comparison between the program participants (treatment group) and the control group. This analysis involves a bias since it compares two different groups of individuals, who have different demographic characteristics and distinctive labour market outcomes. In other words, at this stage, the comparison does not control for the participants' characteristics. However it is still giving an estimate of the raw impact of the programs. It assesses the change in the labour market status of the individuals that has happened after the program participation/cut-off date. We provide comparison between the treated and control group, as well as between the different types of programs.

3.1 Characteristics of the respondents

As previously explained, the analysis in this section presents the raw effects of the programs. Given that the treatment and control groups are likely to differ in their observable and unobservable characteristics, a comparison of their employment outcomes can be biased. In other words, program participants may have better employment outcomes not because of the effectiveness of the programs but because of, say, their better characteristics (higher education levels, prior work experience, shorter unemployment spells prior to program participation, etc.). Hence we proceed by analysing the differences in the main characteristics of the treatment and control groups, for each program.

Table 3-1 presents comparative information on several demographic characteristics of the treatment and control group, for each program. As expected, program participants (and the control group) are younger for the internship program (average age of the program participants is 31 years, and 26 and 30 years for the control groups, in 2010 and 2012, respectively). The same holds for the advanced IT training, where the average age of the treated is 36 and 33 for the non-participants. The differences in average age between the treatment and control groups are small. Males dominate in the treatment and control group in almost all programs except in the internship programs in both years, and training for known employer in both years. In an extreme, all participants in the training for deficient occupations were males, which is mainly due to the nature of the occupations in which training was held (male-dominated occupations).

Participants are predominantly with Macedonian ethnic origin in all programs, which matches the population structure. Ethnic Albanians are represented in higher shares in training in deficient occupations (11%), self-employment program (9%), and internship program in 2010 (8%).

Urban population dominates in all programs, with shares from 60% (in participants in the self-employment program and the control group for the training for known employer 2012), up to 95% in the internship program 2010. Rural population, on the other hand, has relatively high share in the self-employment program (36% and 33% among program participants and non-participants, respectively), and in the training for known employer 2012 (31% for both groups).

There are differences in the education structure both between the programs and within the participants and non-participants (see Table 3-1). The education structure of the treatment group for some of the programs is determined by the eligibility rules. The secondary education dominates within the following programs: self-employment, wage subsidies and training in deficient occupations (among treatment and control group), and training for known employer 2010 and 2012 (for the treatment group). On the other hand, most of the participants and non-participants in the internship program and training in advanced IT skills were holding tertiary education degree (either baccalaureate or master/PhD). Larger shares of primary educated individuals within the treatment groups are observed in the following programs: training for known employer (14%) and 12% in self-employment program, wage subsidies and training for known employer 2010 and 2012.

Table 3-1 Demographic characteristics of treatment and control group

Indicator	SE		IN10		IN12		WS		IT		TKE10		TKE12		TDO	
	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG
Age (mean)	45	44	31	30	31	26	46	41	36	33	36	44	37	46	39	44
Gender (in %)																
Males	66.1	69.6	35.7	43.4	47.4	35.1	57.0	55.7	74.2	49.2	42.7	52.9	32.7	59.6	100	59.2
Females	33.9	30.4	64.3	56.6	52.6	64.9	43.0	44.3	25.8	50.8	57.3	47.1	67.3	40.4	0.0	40.8
Ethnicity (in %)																
Macedonian	87.5	86.3	89.3	86.7	87.2	96.5	93.3	90.0	88.6	94.9	94.7	83.7	95.6	88.1	78.6	82.3
Albanian	8.9	9.6	8.0	9.7	8.9	2.3	2.8	3.3	5.7	3.4	4.0	6.1	3.5	1.7	10.7	14.6
Other	3.6	4.1	2.7	3.6	3.9	1.2	3.9	6.7	5.7	1.7	1.3	10.2	0.9	10.2	10.7	3.1
Urban/rural (in %)																
Urban	59.9	63.8	94.6	88.5	80.8	70.3	82.3	70.3	88.6	86.4	74.7	80.6	69.0	59.6	60.7	77.7
Suburban	3.6	2.9	2.7	4.4	5.1	5.2	2.8	6.5	0.0	6.8	0.0	7.2	0.0	8.9	17.9	10.0
Rural	36.5	33.3	2.7	7.1	14.1	24.5	14.9	23.2	11.4	6.8	25.3	12.2	31.0	31.5	21.4	12.3
Marital status (in %)																
Married	81.3	80.8	32.1	32.7	24.3	23.3	75.3	72.5	45.7	33.9	66.7	70.3	72.6	77.9	39.3	76.9
Highest education today (%)																
No education	0.5	0.4	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	1.6	0.0	2.2	0.0	0.0
Up to 4 years primary	0.0	1.7	0.0	0.0	0.0	0.0	1.4	0.3	0.0	0.0	0.0	3.1	0.9	7.3	0.0	0.0
5-7 years primary	0.5	1.3	0.0	0.0	0.0	0.0	0.7	0.5	0.0	0.0	1.3	0.8	0.0	0.4	0.0	0.8
Primary	12.0	17.9	0.0	0.0	0.0	0.0	11.6	15.8	0.0	0.0	12.0	39.3	12.4	65.5	14.3	4.6
Secondary, 3-year	12.0	13.8	0.9	1.8	3.8	1.8	10.9	14.4	5.7	0.0	8.0	8.2	16.8	8.6	25.0	21.5
Secondary, 4-year	41.7	40.0	24.1	32.7	17.9	16.4	37.7	34.3	20.0	18.6	52.0	37.4	37.2	15.1	35.7	56.9
High school (gymnasium)	14.6	11.3	4.5	4.4	6.4	2.9	12.3	11.9	5.7	1.7	4.0	5.1	7.1	0.4	10.7	13.8
Baccalaureate	17.7	13.3	59.8	57.5	62.8	67.8	23.6	21.6	60.0	74.6	18.7	4.3	24.8	0.4	14.3	2.3
Master's or PhD	1.0	0.4	10.7	3.5	9.0	11.1	1.8	1.0	8.6	5.1	4.0	0.4	0.9	0.0	0.0	0.0

Source: Own calculations based on survey data.

Note: If not otherwise stated, shares (%) are reported.

Note: SE – self-employment program; IN10 – internship program in 2010; WE – wage subsidy; IT – training in advanced IT skills; TKE10 – training for known employer; TDO – training in demand deficient occupations; TG – treatment group; CG – control group.

Within the programs, larger differences between treatment and control groups can be observed within the: i) training in advanced IT skills (the share of tertiary educated individuals within the treatment group is 69% compared to 80% in the control group); ii) training for known employer in both years (much larger share of primary educated individuals in the control group relative to the treatment group, and considerably less tertiary educated individuals); and iii) training in deficient occupations (larger share of secondary educated individuals within the control group, than the treatment group).

Data on labour market histories prior to entering the programs/cut-off date (see Table 3-2) show differences between programs (in line with the eligibility rules for each program) and between treatment and control groups. Across all programs, dominant share of participants and non-participants had prior unemployment spell of less than one month, which is contrary to our expectations given that in all programs there is a minimum required duration of unemployment as part of the eligibility conditions (except for the

internship program and training for known employer). We observe especially high percentage of participants (and non-participants) with less than 1 month unemployment spell in the internship program in 2012. This might imply some form of misuse or drain of the public funds as probably young persons register with ESA only to become eligible for the program. The same, although at a lower magnitude, can be observed for the training for known employer, where close to 70% of the participants in 2010 and 56% in 2012 had previous unemployment spells of less than one month. In addition, although in training for deficient occupations program there is a requirement that participants are unemployed at least 3 months (for young persons) and 6 months, in general, in 2010, 75% of the participants had shorter unemployment spells.

Table 3-2 Labour market characteristics before program's entry/cut-off point, %

	SE		IN10		IN12		WS		IT		TKE10		TKE12		TDO	
	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG
Unemployment spell (in % of the group)																
< than month	49.5	65.0	87.5	85.0	94.9	93.0	51.4	55.7	57.1	71.2	68.0	59.7	55.8	52.3	75.0	44.6
1-6 months	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6-12 months	4.2	1.7	6.3	5.3	1.3	0.6	4.6	7.1	14.3	6.8	10.7	4.9	4.4	0.0	3.6	10.0
12-24 months	3.6	2.1	1.8	3.5	0.0	0.6	3.9	4.3	8.6	5.1	0.0	4.2	9.7	7.2	7.1	8.5
24+ months	41.7	30.8	4.5	5.3	3.8	5.2	39.8	31.7	17.1	13.6	21.3	30.8	30.1	40.4	14.3	36.9
Prior work experience (in % of the group)																
Yes	69.3	62.9	22.3	23.9	10.2	9.9	72.5	65.1	57.1	45.8	46.7	58.9	54.9	57.9	50.0	73.8
No	30.7	37.1	77.7	76.1	89.8	90.1	27.5	34.9	42.9	54.2	53.3	41.1	45.1	42.1	50.0	26.2

Source: Own calculations based on survey data.

Programs with higher share of long term unemployed (above 1 year) are: self-employment program (45% of the participants were unemployed more than 1 year), wage subsidy (44%), training for known employer 2012 (40%), and training in advanced IT skills (26%). Self-employment programs elsewhere are commonly offered for unemployed with shorter spells, and this has been changed in the program design after 2008 (the year of program that we evaluate). The high share of participants with long duration of unemployment among participants in the training in advanced IT skills deserves some attention of the policymakers. Table 3-2 also shows that higher share of the program participants had some work experience before the start of the program/cut-off date. However, there are large variations between the programs, such that the share of participants in the internship program who had work experience prior to the program is about 28% (25% for the control group). This is related to the design of the program where only young individuals with no prior work experience are eligible for the program. The difference in the share of individuals with previous work experience between the treatment and control group is particularly large within the training in deficient occupations.

In summary, the analysis above shows that the program participants have similar demographic and labour market characteristics with the control group. Regarding the education structure, the participants in two programs (training in advanced IT skills and training for known employer) are slightly more educated compared to the control group. Program participants have better characteristics in terms of the prior work experience, for all programs. Two of the indicators we analyse are not in line with our expectations: i) large share of program participants stated to have very short unemployment spells prior to program participation; ii) there is high share of long-term unemployed among participants in training in advanced IT skills.

3.2 Main outcome indicators

Program participants (treatment group) have better employment outcomes relative to the control group in all programs, except in the subsidized employment (Table 3-3). In terms of the current employment (at the interview date), the most effective programs are the internship programs with 63% and 67% of the 2010 and 2012 program participants (respectively) being currently employed. The least effective program is the training in deficient occupations where only 21.4% of the participants are currently employed (based on the ILO definition of employment). This suggests that the program is probably not effective in enhancing the skills of the participants in the demanded occupations/professions, and/or that the choice of the sectors in excess demand of specific workers is incorrect. We will come back to this issue in the impact evaluation (Section 4). The difference between the treatment and control groups is largest for the training for known employer 2012, where participants have 185% higher employment today compared to the control group (similar results are present for the 2010 program). Though, as elaborated in Section 2, we cannot say based on this difference that the program is (very) effective, as we need to compare “same” individuals in order to do so. We provide such analysis in Section 4. The smallest differences between the treatment and control groups are found for the self-employment (only 5 p.p.) and training in advanced IT skills (7.5 p.p.). On the other hand, the current employment is lower for the individuals that participated in the wage subsidy program compared to non-participants. This may suggest an existence of a stigma effect, such that employers are not willing to employ workers that were previously part of the wage subsidy program based on a belief that those workers are of low quality. We come back to this issue later on.

Table 3-3 Labour market status

Type of program	Current employment		Not employed at any time after/cut-off		Employed at program end, currently not employed	
	TG	CG	TG	CG	TG	CG
SE	49.5	44.6	38.0	36.7	12.0	17.9
IN10	63.4	40.7	25.9	43.4	5.4	15.9
IN12	66.7	38.6	24.4	43.6	5.1	17.4
WS	36.0	40.1	53.9	41.5	9.2	17.1
IT	60.0	52.5	14.3	25.4	22.9	15.3
TKE10	53.3	23.2	37.3	61.2	8.0	15.6
TKE12	52.2	18.3	34.5	74.0	14.2	7.7
TDO	21.4	12.3	64.3	70.8	10.7	15.4

Source: Own calculations based on survey data.

Large share of participants in the training program for deficient occupations (64%) and in wage subsidies (54%) have not been employed any time after program end (Table 3-2). In most programs, participants have lower chances to be unemployed at any time after program end/cut-off, with exception of self-employment and wage subsidies. Apparently, these two programs are yielding very low results. Participants in the training for advanced IT skills are the least likely among all treatment groups to have not been unemployed at any time after program end. In general, lower share of program participants were employed at the program end/cut-off date, but not currently employed, compared to the control groups. The exception is the training for known employer in 2012.

Besides the generally better employment opportunities, program participants show slightly more positive attitude towards changes in well-being, as assessed i) by the change in the financial situation from the program beginning/cut-off date until the interview and ii) change in the employment prospects (see Table 3-4).

The share of program participants who reported better or same financial situation today compared to the beginning of the program is higher for all programs compared to the control group (except for the training in deficient occupations). However, the most frequent response for all programs and for both treatment and control groups is that there is no change in the financial situation. In 5 programs (self-employment, internship in both years, training in advanced IT skills, training for known employer 2012), program participants show better subjective well-being compared to the control group. The training in advanced IT skills and internship program 2012 are programs with highest shares of participants feeling that today their financial situation is better than before the program participation (27% and 21% of the respondents, respectively).

Table 3-4 Self-assessment of the changes in employment prospects and the financial situation, %

	SE		IN10		IN12		WS		IT		TKE10		TKE12		TDO	
	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG
Change in financial situation (in % of respondents)																
Better	15.7	12.6	19.1	12.5	20.5	15.8	11.6	14.3	27.3	8.5	1.3	12.1	16.1	5.1	3.6	10.1
Same	49.2	49.0	65.5	67.0	68.0	67.8	52.7	45.4	60.6	71.2	84.0	54.9	44.6	54.3	57.1	56.6
Worse	35.1	38.5	15.5	20.5	11.5	16.4	35.7	40.4	12.1	20.3	14.7	31.1	38.4	40.6	39.3	33.3
Change in employment prospects (in % of respondents)																
Better	5.4	7.1	17.1	5.7	8.7	15.5	2.8	9.1	6.7	7.1	14.3	10.0	6.1	3.7	0.0	2.7
Same	60.2	45.7	70.7	81.4	87.0	73.8	40.1	36.8	40.0	57.1	80.0	41.5	73.5	33.5	70.0	38.9
Worse	34.4	47.2	12.2	12.9	4.4	10.7	57.1	54.0	53.3	35.7	5.7	48.5	20.4	62.8	30.0	58.4
Subjective labour market status (in % of respondents)																
Employed	45.3	37.1	63.4	40.7	66.7	39.0	42.3	38.0	60.0	49.2	49.3	20.2	52.2	9.8	21.4	13.9
Not employed	54.7	62.9	36.6	59.3	33.3	61.1	57.8	62.0	40.0	50.9	50.7	79.9	47.8	90.2	78.6	86.2

Source: Own calculations based on survey data.

There is no clear pattern or distinction between treatment and control groups in terms of the subjectively assessed change in their employment prospects before and after program participation/cut-off. Highest shares of respondents stating improvement in their employment prospects after program end is observed for internship programs and training for known employer 2010. One can also note that there is no clear link between the assessment of the change in employment prospects and actual employment outcomes reported in Table 3-3.

Whereas the above two subjective measures of well-being and employment chances reveal rather mixed and small effects of the programs, the subjective (self-reported) labour market status shows much higher employment rate among program participants than the control group. In particular, as Table 3-4 shows, the share of participants from all programs who self-reported to be employed at the time of interview is higher than the share of non-participants who reported to be employed. The share of self-reported employment is highest among participants in the internship program in 2010 and 2012 (63% and 67%, respectively), and the training for known employer in 2012 (52%). The difference between treatment and control groups in self-reported employment is largest for the training for known employer in 2012. The subjective employment status of the participants is very close to their actual employment status (based on the ILO definition) reported in Table 3-3. There are slight differences for the self-employment program (the employment rate is 49% whereas subjective employment 45%), training for known employer (53% vs. 49%) and wage subsidies (36% vs. 42%).

In summary, analyses in this section show that in terms of the current employment (at the interview date), the most effective programs are the internship programs in 2010 and 2012, whereas the least effective program is the training in deficient occupations. The smallest differences in current employment between the treatment and control groups are found for the self-employment and training in advanced IT skills. For the wage subsidy program, the current employment is lower for the program participants. In most programs,

participants have lower chances to be unemployed at any time after program end/cut-off, with exception of self-employment and wage subsidies. This information signals that we can expect to find two of the programs – self-employment and wage subsidies – as least effective among all programs or even ineffective. Besides the generally better employment opportunities, program participants show slightly more positive attitude towards changes in well-being, as assessed by the change in the financial situation from the program beginning/cut-off date until the interview and change in the employment prospects. Though, in general, most of the respondents report same well-being today as at the program start/cut-off date.

3.3 Additional indicators

Program participants seem to hold more quality jobs today than the non-participants. The share of employed workers among program participants is in general higher relative to the control group, whereas the share of unpaid family workers lower (although it is quite low for the whole sample) (see Table 3-5). For the self-employment program, the share of self-employed among the program participants is higher than among non-participants, which is in line with the program objective. There is a large difference in the employment status of the participants and control group for the training for known employer in 2012 (and in 2010). While 97% of the program participants in 2012 have a status of employed, this is the case for only 36% of the control group (the majority being self-employed).

Table 3-5 Employment characteristics of the control group and program participants, (today)

	SE		IN		IN		WS		IT		TKE10		TKE12		TDO	
	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG
Employment status (% of respondents)																
Employed	35.8	50.9	100	97.8	96.2	92.4	88.3	83.5	100	96.7	92.5	72.6	96.7	35.7	71.4	88.2
Self-employed	62.1	39.6	0	2.2	3.8	7.6	10.7	12.0	0	3.3	7.5	21.0	1.7	54.8	28.6	11.8
Unpaid family worker	2.1	9.4	0	0	0	0	1.0	4.5	0	0	0	6.5	1.7	9.5	0	0
Informal/formal (% of respondents)																
Formal	82.4	70.4	98.6	100	96.1	96.7	91.2	87.1	100	96.6	97.3	84.4	98.3	100	80.0	86.7
Informal	17.6	29.6	1.4	0	3.9	3.3	8.8	12.9	0	3.4	2.7	15.6	1.7	0	20.0	13.3
Contract type (% of respondents)																
Open-end	67.9	81.6	80.0	77.8	81.6	64.4	80.7	65.9	66.7	75.0	86.1	71.1	70.2	53.3	25.0	92.3
Close-end	21.4	13.2	20.0	22.2	18.4	33.9	18.1	33.0	28.6	25.0	13.9	26.3	29.8	40.0	75.0	7.7
Seasonal	10.7	5.3	0	0	0	1.7	1.2	1.1	4.8	0	0	2.6	0	6.7	0	0
Source: Own calculations based on survey data.																

In addition, data for the training in deficient occupations shows that as much as 29% of the program participants are self-employed today (relative to 12% among the control group) suggesting that the program might have supported unemployed persons to start own businesses which is not an explicit objective of the program.

Only small share of program participants who are wage-employees work in the public sector (see Table 3-6). These shares range from literally zero for training in deficient occupations to 18% for internship 2010 and 29% for training in advanced IT skills. Within the control group, largest share of public sector employees is registered among participants in the training in deficient occupations (31%). Moreover, program participants are less likely to work in the public sector relative to the non-participants. The only exception is the training in advanced IT skills where opposite is the case.

Table 3-6 Wage-employment characteristics of the control group and program participants

	SE		IN10		IN12		WS		IT		TKE10		TKE12		TDO	
	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG
Ownership status (%)																
Public	8.4	11.3	18.3	19.6	7.5	22.7	4.9	12.0	28.6	23.3	5.0	15.5	5.0	4.8	0	31.3
Private-registered	84.2	70.8	81.7	80.4	90.6	74.2	90.3	80.2	71.4	76.7	92.5	67.2	95.0	47.6	100	68.8
Private-unregistered	6.3	7.5	0	0	1.9	3.0	2.9	3.7	0	0	2.5	8.6	0	38.1	0	0
Other	1.1	10.4	0	0	0.0	0.0	1.9	4.1	0	0	0	8.6	0	9.5	0	0

Source: Own calculations based on survey data.

For all programs, participants are less likely to work informally. Moreover, in some programs all of the gained wage employment of participants is formal (internship program 2010, training in advanced IT skills, training for known employer 2012 and training in deficient occupations). On contrary, a large share of control group for training for known employer 2012 (38%) work informally.

Table 3-7 shows that there are large shares of respondents (both participants and non-participants) who report 0 wages. The shares range between 13% (for the control groups for wage subsidies and training for known employer in 2010) to 62% for the participants in the training in advanced IT skills.

There is a clear wage disadvantage of the program participants in 3 programs (as assessed by the share of workers receiving wages above 10,000 MKD): training in advanced IT skills, training for known employer 2010 and training in deficient occupations. The disadvantage is very large among participants in the IT program where 29% of the treated individuals report wages above 10,000 MKD compared to 48% of the non-participants. As we explain in Section 4, we believe that the participants in this program exhibit high probability of emigration where those staying in the country may be with worse labour market characteristics, work more in public administration (Table 3-6) and hence earn lower wages.

Table 3-7 Earnings, control group and program participants, %

	SE		IN		IN12		WS		IT		KE10		KE12		TDO	
	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG	TG	CG
0	29.4	24.1	45.1	40.0	25.5	32.8	27.5	13.4	61.9	44.8	24.3	13.3	13.8	20.0	40.0	20.0
0-5,999	5.9	3.7	0	0	2.0	0	3.3	3.0	0	0	0	8.9	3.4	6.7	0	0
6,000-9,999	8.8	24.1	0	4.4	21.6	19.7	15.4	37.6	9.5	6.9	35.1	26.7	36.2	53.3	20.0	20.0
10,000-11,999	23.5	16.7	9.9	6.7	23.5	13.1	18.7	21.8	9.5	6.9	10.8	20.0	19.0	0.0	40.0	26.7
12,000-14,999	20.6	16.7	12.7	17.8	17.6	9.8	16.5	11.4	4.8	13.8	8.1	13.3	10.3	13.3	0	6.7
15,000-19,999	11.8	13.0	19.7	20.0	7.8	19.7	12.1	10.9	9.5	17.2	16.2	15.6	15.5	6.7	0	26.7
>20,000	0	1.9	12.7	11.1	2.0	4.9	6.6	2.0	4.8	10.3	5.4	2.2	1.7	0	0	0.0

Source: Own calculations based on survey data.

Program participants seem to hold more quality jobs today than the non-participants. The share of employed workers among program participants is in general higher relative to the control group, whereas the share of unpaid family workers lower (although it is quite low for the whole sample). Moreover, program participants are less likely to work in the public sector relative to the non-participants. For all programs,

participants are less likely to work informally. There is however no clear advantage of program participants in terms of earnings.

This Section provided analysis of the raw effects of the programs. While it does not unveil real effects of the program, it certainly provides some institution for the effectiveness. We turn to the impact evaluation of the programs in the next Section.

4 Impact evaluation: results and discussion

The objective of this report is to evaluate the effectiveness and efficiency of the active labour market programs implemented by ESA against a counterfactual reality where these programs did not exist. For this purpose, we compare labour market (employment, unemployment, inactivity and average net wage) and subjective wellbeing outcomes (self-assessment of past and current financial situation and evaluation of the chances to find a job). For a valid measurement of the program effects, we compare program participants – the *treatment group* – to the non-participants (*control group*), i.e. those who could have participated in the program, i.e. those who had an equal chance to be selected for participation in the program as the actually treated.

The evaluation will be conducted program by program. We conduct the analysis and present the results and findings in the following five steps.

First, to assess whether program participation can be regarded as quasi-random, we perform statistical tests of the hypothesis of random assignment to participation. Specifically, we test whether the means of important socio-demographic characteristics and labour market outcomes are significantly different between the two groups. In case they are, we observe the difference in the outcome variable between the treated and the non-treated, i.e. we jump to step five. On the contrary, if the hypothesis of random assignment is rejected, it would be misleading to measure net effects as the difference in average outcomes between the two groups, in which case we continue the analysis with the second step.

Second, in order to mimic a randomized experiment *ex post*, we construct a control group that resembles the treatment group by applying *one-to-one nearest neighbour matching with replacement*, in the cases where the hypothesis of random assignment has been rejected. This method comprises two steps: (i) an estimation of the individual probabilities to participate to the program, depending on a set of observable characteristics; and (ii) the matching of participants and non-participants on the basis of these estimated probabilities. The first step is conducted through using standard *probit* regression on the treated and the non-treated, whereby the estimated coefficients will provide insights in the factors influencing selection into treatment, but may also capture factors of attrition from the survey, i.e. factors explaining differential non-response rates in the treatment and in the control group.

Third, we apply the one-to-one nearest neighbour matching with replacement by using the estimated parameters from the probit regression of the previous step to predict the probability to participate in a treatment – the so-called *propensity score* – for each individual in the treatment and comparison groups. The propensity scores are used to match participants with comparable non-participants. For each treated individual, we look for the one individual among non-participants who is the closest neighbour in terms of the predicted probability of being treated. In other words, for each pair comprising a participant and a non-participant, the absolute difference in terms of the estimated propensity to participate in a certain treatment is minimized. To ensure that the matched pairs have reasonably similar probabilities to be treated, we exclude participants for whom the predicted probability to be in the program is larger than for any individual in the comparison group. In this way we achieve *common support*. Alternative matching procedures are used as robustness checks.

Fourth, we conduct evaluation of the matching quality. A way to do so is to compare the standardized bias before matching (SB^b) to the standardized bias after matching (SB^a). The standardized biases are defined as:

$$SB^b = \frac{(\bar{X}_1 - \bar{X}_0)}{\sqrt{0.5(V_1(X) - V_0(X))}}; SB^a = \frac{(\bar{X}_{1M} - \bar{X}_{0M})}{\sqrt{0.5(V_{1M}(X) - V_{0M}(X))}}$$

Where $X_1(V_1)$ is the mean (variance) in the treated group before matching and $X_0(V_0)$ is the analogue for the comparison group. $X_{1M}(V_{1M})$ and $X_{0M}(V_{0M})$ are the corresponding values after matching (Rosenbaum and Rubin, 1985). We also re-estimate the propensity score on the matched sample to compute the pseudo- R^2 before and after matching (like in Sianesi, 2004).

Fifth, we study the causal impact of the social programs on labour market outcomes, and the subjective well-being variables. The outcome variables are based on the labour market status after the participation in the program/cut-off date or at the time of the interview, namely: (i) unemployed at all times after the program, (ii) employment in a regular job, including self-employment, at the time of the interview, (iii) employed at program's end but unemployed at the time of the interview, and (iv) inactivity. In addition, we estimate the effects of the program (v) on the level of individual wage earned per hour from the main job.

4.1 Self-employment program

We start the evaluation of each program by looking at the observable differences between the treatment and the control groups. For the self-employment program, Table 4-1 shows that few of the demographic characteristics of the two groups are different: the treatment group has slightly higher education both before the treatment/cut-off date and today and has on average more members in the household. The Hotelling test at the bottom of the table actually suggests that there is no statistical difference in the vector of the means of the demographic variables of the control versus the treated individuals within the self-employment program. Surprisingly, the tests of the difference between the two groups further suggest that the treated have only lower employment rate at the end of the program than compared to the control group, as well higher perception of employment today, but the difference is significant only at the 10%.

Table 4-1 Means' comparison – Self-employment program

Socio-demographic characteristics	mean control	mean treated	p-value	sig
Age (years)	43.78	44.68	0.42	
Education today	5.77	6.13	0.01	***
Educational level before treatment/cut-off	5.72	6.08	0.00	***
Gender (1=male)	0.70	0.66	0.45	
Ethnicity (1=Macedonian)	0.86	0.88	0.70	
Ethnicity (1=Albanian)	0.10	0.09	0.79	
Marital status (1=married)	0.81	0.81	0.91	
Geography (1=urban)	0.64	0.60	0.41	
Geography (1=suburban)	0.03	0.04	0.68	
Geography (1=rural)	0.33	0.36	0.39	
Children (1=person has at least one child)	0.83	0.84	0.88	
Number of children	1.71	1.77	0.55	
Partner's employment status (1=if partner is employed)	0.44	0.39	0.33	
Number of household members	4.00	4.31	0.05	*
Number of household members below age 15	0.78	0.88	0.32	
Number of employed household members (age 15-64)	1.08	1.09	0.91	
Number of unemployed household members (age 15-64)	1.82	1.99	0.19	
Number of retired household members	0.29	0.34	0.37	
House ownership (1=owned)	0.88	0.90	0.39	
House size (sq. meters)	82.74	86.07	0.31	
Work experience (1=has work experience)	0.63	0.69	0.17	
Work experience (number of months)	77.63	88.71	0.26	
Contract (1=has written contract)	0.33	0.39	0.19	
Salary before treatment/cut-off (MKD)	6,272.92	6,070.83	0.81	
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	2.26	2.19	0.32	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.40	2.29	0.17	
Subjective employment status (1=employed)	0.37	0.45	0.09	*
Employment status (1=employed)	0.45	0.49	0.31	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.37	0.38	0.77	
Employment status (1=inactive)	0.21	0.16	0.13	
Employed at program's end, but unemployed now	0.18	0.12	0.08	*
Wage per hour (MKD)	21.08	22.50	0.69	
Hotelling T-squared test H0: The vector of means of all variables are equal across groups				
p-value	0.1164			
Source: Authors' calculations based on survey.				
Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.				

To pursue further the early signs that the treated may actually have worse positioning on the labour market than compared to the control group, Table 4-2 presents the result of the regression of the outcome variables on the treatment only, and on the treatment and a set of explanatory variables (age, education, gender, ethnicity, marital status, geography, the number of children and household members and if the person has a work experience).

Table 4-2 Effects - Self-employment program

Outcome variable	Only treatment included as explanatory			Treatment and a bunch of other explanatory variables		
	coefficient	p-value	sig	coefficient	p-value	sig
Financial condition after treatment/cut-off (1=better; 3=worse)	(0.11)	0.32		(0.11)	0.34	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	(0.22)	0.17		(0.31)	0.07	*
Subjective employment status (1=employed)	0.21	0.08	*	0.29	0.03	**
Employment status (1=employed)	0.12	0.31		0.20	0.13	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.04	0.77		(0.05)	0.69	
Employment status (1=inactive)	(0.21)	0.14		(0.30)	0.05	**
Employed at program's end, but unemployed now	(0.26)	0.09	*	(0.29)	0.07	*
Wage per hour (MKD)	1.41	0.69		2.90	0.41	

Source: Authors' calculations based on survey.
Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

The indicative results we obtained in the previous table are also found here: participants think that their chances for finding a job are better after they have participated in the program, as well they think their employability is significantly larger – actually by large 29 percentage points (p.p.) – compared to non-participants. The latter is confirmed by the finding that the inactivity rate of participants today is lower by large 30 p.p. than that of non-participants. Still, the employment at program's end is significantly lower for participants than for non-participants.

These findings portray important effects of the self-employment program. They actually suggest that it leads to building confidence among participants for their employability after the program participation. However, contrary to their perception, the actual figures suggest that they are not having an advantage relative to non-participants when the actual employment status and financial indicators are observed. In addition, caution in the interpretation of the effect is needed, since it is significant only at the 10%.

Overall, results suggest that the self-employment program has not exerted any visible difference for labour-market outcomes for those who were trained and subsequently obtained a grant to start a business, but likely changed the perception of these persons that with the program they became more employable, although they actually did not. This is in line with the observed labour-market transitions, by type of employment in Table 4-3. Namely, the shares of those losing a job (transition from employment status before the program/cut-off date to unemployment after the program/cut-off date) are similar (slightly above 50%); in addition, though, the share of jobs created (transitions from unemployment before the program/cut-off date to employment after the program/cut-off date) is slightly larger for the treatment group (54%) than for the control group (44%). In that line, 59% of the newly created jobs within the treatment group were actually self-created, while the control group' job creation leaned toward wage employment (51% versus 49% for self-employment type of jobs).

Table 4-3 Employment transition, by type of employment

			Employment status now				
			Unemployed	Employed			TOTAL
				Wage employed	Self employed	Unpaid family worker	
Employment status before participation / cut-off date	T	Unemployed	27	13	19	0	59
		Employed	70	21	40	2	133
	C	Unemployed	50	20	19	0	89
		Employed	83	34	23	10	150
		TOTAL	230	88	101	12	431

Source: Authors' calculations based on survey.

What are the likely explanations/reasons for this result? First, recall that we use as a control group the persons who scored on the entrepreneurship test below the threshold, but not lower than 45 point and hence they did not proceed with training and funding of their business ideas. From this viewpoint, those persons are likely to have similar motivation, (initial) courage and capability to establish and run a business. In other words, the selection on unobservables is minimized in this way, i.e. the two groups we proved are identical on observables (Table 4-1), are also likely identical (or very similar) on unobservables. Results actually suggest that people who applied but were rejected have the same employment prospects as the ones who got the grant and started the business: they were able to start the business even without the grant and/or were able to find wage employment after rejection, despite the former further built their self-confidence over the program. This may suggest that the program entails large dead-weight loss.

To test this explanation, we conduct an exercise whereby we swap the control group with a random control group whereby we cannot assume it is identical on unobservables. Namely, we have on disposal a randomly selected sample of persons which we later use as a control group for the training for known employer program in 2010. The disadvantage of using this group as a control group for self-employment program is that they refer to different years, it is not selected from all occupations and not from all ESA centres. However, we conduct the evaluation for comparison purposes.

Table 4-4 indeed suggests that the two groups are genuinely different on observables (the discussion on the particular differences being unimportant for the time being).

Table 4-4 Means' comparison – Self-employment program: exercise with a 'general' control group

Socio-demographic characteristics	mean control	mean treated	p-value	sig
Age (years)	43.99	44.68	0.55	
Education today	5.05	6.13	0.00	***
Educational level before treatment/cut-off	5.05	6.08	0.00	***
Gender (1=male)	0.53	0.66	0.00	***
Ethnicity (1=Macedonian)	0.84	0.88	0.25	
Ethnicity (1=Albanian)	0.06	0.09	0.27	
Marital status (1=married)	0.70	0.81	0.01	***
Geography (1=urban)	0.81	0.60	0.00	***
Geography (1=suburban)	0.07	0.04	0.09	*
Geography (1=rural)	0.12	0.36	0.00	***
Children (1=person has at least one child)	0.72	0.84	0.00	***
Number of children	1.45	1.77	0.00	***
Partner's employment status (1=if partner is employed)	0.35	0.39	0.38	
Number of household members	3.85	4.31	0.00	***
Number of household members below age 15	0.71	0.88	0.10	*
Number of employed household members (age 15-64)	0.91	1.09	0.21	
Number of unemployed household members (age 15-64)	1.84	1.99	0.24	
Number of retired household members	0.36	0.34	0.77	
House ownership (1=owned)	0.87	0.90	0.26	
House size (sq. meters)	74.58	86.07	0.00	***
Work experience (1=has work experience)	0.59	0.69	0.02	**
Work experience (number of months)	91.00	88.71	0.83	
Contract (1=has written contract)	0.37	0.39	0.70	
Salary before treatment/cut-off (MKD)	4,487.83	6,070.83	0.02	**
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	2.23	2.19	0.58	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.39	2.29	0.21	
Subjective employment status (1=employed)	0.20	0.45	0.00	***
Employment status (1=employed)	0.23	0.49	0.00	***
Unemployed at any time after program's end/cut-off (1=unemployed)	0.61	0.38	0.00	***
Employment status (1=inactive)	0.30	0.16	0.00	***
Employed at program's end, but unemployed now	0.16	0.12	0.27	
Wage per hour (MKD)	16.45	22.50	0.19	
Hotelling T-squared test H0: The vector of means of all variables are equal across groups				
p-value	0.0000			
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.				

Given this, we pursue a matching to pair the treatment and control individuals with same or similar observable characteristics. Table 4-5 exhibits the *probit* estimation results (estimated coefficients and marginal effects), underlying the propensity scores for the treatment.

Table 4-5 Matching – Self-employment program: matching with a ‘general’ control group

Variable	Marginal effects	p-value	sig
Age	0.04	0.01	***
Age squared	(0.00)	0.02	**
Education today	0.17	-	***
Gender (1=male)	0.20	-	***
Ethnicity (1=Macedonian)	0.09	0.44	
Ethnicity (1=Albanian)	(0.08)	0.58	
Marital status (1=married)	0.08	0.28	
Geography (1=urban)	(0.39)	-	***
Geography (1=suburban)	(0.32)	0.00	***
Number of children	0.03	0.39	
Number of household members	0.03	0.05	*
Work experience (1=has work experience)	0.05	0.41	
# Observations	448		
Log-pseudolikelihood	-231.14635		
Pseudo R ²	0.2445		

Source: Authors’ calculations based on survey.
Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

If we assume that the control group is sufficiently well-designed counterpart of our treatment group, then the results in Table 4-5 suggest that older persons (up to certain age), more educated, males and rural inhabitants, as well those living in households with more members have larger probability to be selected for treatment (obtaining a grant to start off a business).

Based on the propensity scores obtained from the probit regression, in what follows we perform a matching, which comes in two options: nearest-neighbour matching with replacements, based on the absolute value of the difference between the propensity score of the selected treatment and that of the control under consideration, so that an individual from the control group could be a match of more than one individual in the treatment group; and caliper matching, whereby we impose additional restriction that the control’s propensity score belongs to a pre-defined radius (so called caliper). Table 4-6 offers two pieces of information: the results for the different outcome variables (both before and after matching); and the reduction of the mean and median bias after matching, so as to judge matching quality. Note that after matching, the observable differences between the treatment and control group should disappear. While we strongly ensure that this condition is satisfied after matching is performed, these results are available on request due to space.

Results robustly suggest that when samples are paired only on observables, the impact evaluation of the self-employment program would produce arguably plausible (and sizeable) results. Our simulation suggests that the program more than doubles the chance that a treated individual would find a job than compared to a non-treated individual, hence also reducing the probability that he/she stays unemployed after the program finishes. However, the uncontrolled unobservables which apparently have a role in this framework, render the estimates in Table 4-6 biased and are presented for illustrative purposes. Overall, Table 4-6 lends indirect support to our claim that persons self-select into application for the self-employment program also based on their unobservables: motivation, courage and inclination to start a business, ideas for business, entrepreneurial spirit and so on, which leads them to have certain employment prospects irrespective of whether they get the grant (proceed to the next stage – training – of the selection process) or not (compete but get rejected for a grant as they fail on the entrepreneurial test); and irrespective of whether their self-confidence is built up over the process or not.

Table 4-6 Effects – Self-employment program when compared to a ‘general’ control group

Outcome variable	Unmatched				Nearest neighbour matching with replacement				Matching with caliper (0.01)						
	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig
Financial condition after treatment/cut-off (1=better; 3=worse)	2.19	2.22	(0.03)	(0.39)		2.19	2.20	(0.01)	(0.04)		2.17	2.12	0.05	0.50	
Bias reduction						Mean	-40%	Median	-37%		Mean	-66%	Median	-61%	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.29	2.39	(0.10)	(1.21)		2.29	2.41	(0.12)	(0.75)		2.25	2.25	-	-	
Bias reduction						Mean	-22%	Median	-15%		Mean	-58%	Median	-37%	
Subjective employment status (1=employed)	0.45	0.20	0.25	5.97	***	0.45	0.15	0.30	3.98	***	0.49	0.18	0.31	4.57	***
Bias reduction						Mean	-30%	Median	-43%		Mean	-55%	Median	-66%	
Employment status (1=employed)	0.49	0.23	0.26	5.94	***	0.49	0.18	0.32	4.03	***	0.53	0.22	0.32	4.54	***
Bias reduction						Mean	-30%	Median	-43%		Mean	-55%	Median	-66%	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.38	0.62	(0.24)	(5.10)	***	0.38	0.71	(0.33)	(3.64)	***	0.34	0.64	(0.30)	(3.84)	***
Bias reduction						Mean	-28%	Median	-43%		Mean	-56%	Median	-66%	
Employment status (1=inactive)	0.16	0.29	(0.14)	(3.42)	***	0.16	0.32	(0.17)	(2.01)	**	0.14	0.29	(0.14)	(2.10)	**
Bias reduction						Mean	-32%	Median	-43%		Mean	-61%	Median	-66%	
Employed at program's end, but unemployed now	0.12	0.15	(0.03)	(0.87)		0.12	0.11	0.01	0.08		0.12	0.14	(0.03)	(0.46)	
Bias reduction						Mean	-38%	Median	-35%		Mean	-66%	Median	-62%	
Wage per hour (MKD)	22.50	16.90	5.59	1.13		22.50	9.55	12.94	1.91	*	24.45	11.72	12.73	2.15	**
Bias reduction						Mean	-32%	Median	-34%		Mean	-62%	Median	-61%	

Source: Authors' calculations based on survey.

Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

A second possible explanation for the non-effectiveness of the self-employment program is the work of the informal economy and employment. Table 4-7, however, suggests that this is an unlikely explanation: only 20% of the informal contracts transitioned to formal ones within the treatment group, and 23% within the control group, which is the same rate and likely signifies the work of forces affecting both groups (like the more stringent regulations, being one example), and not the work of the ESA self-employment program. Still, Table 4-7 needs to be approached with caution, because about 1/3 of the respondents did not reveal the type of contract.

Table 4-7 Employment transition, by type of contract

			Employment status now			
			Unemployed	Employed		
				Formal contract	Informal contract	TOTAL
Employment status before participation / cut-off date	T	Unemployed	27	12	1	40
		Formal contract	41	8	1	50
		Informal contract	22	6	2	30
	C	Unemployed	50	14	6	70
		Formal contract	43	11	4	58
		Informal contract	32	11	5	48
	TOTAL	215	62	19	296	

Source: Authors' calculations based on survey

Overall, we conclude that the ESA self-employment program did not exert influence on the actual labour market outcomes of the participants, but likely affected their self-perceived employment prospects through building their self-confidence. The main reason for this finding is related to the characteristics of the individuals from both groups (treatment and control group) which self-select themselves into application to the program (and not to the treatment). Namely, it is likely that persons who are motivated, determined, encouraged and with business ideas apply for the call, which then is the main reason why their chances for finding a job, employability and overall financial condition are good, irrespective of whether they ultimately get selected for training and funding within the ESA program.

4.1.1 Summary of findings and recommendations

The evaluation shows that the program for self-employment (2008) positively affected subjective measures of well-being of the program participants, but not their actual employment. In particular, higher share of participants believes that their employment chances after program end improved compared to the non-participants. In addition, significantly higher share of program participants (subjectively) state to be employed today. However, these results are not confirmed with the findings for current employment status, which do not show any significant difference between participants and non-participants. Participants are still significantly less likely to be inactive today and to have been employed after program's end but unemployed now, compared to the non-participants. These results suggest that the program participation likely changed the perception of the participants that they became more employable through the program, although they actually did not.

The likely explanation for the same employment outcomes of the treatment and control group is that persons who apply for the program likely self-select themselves out of the other unemployed based on unobservable characteristics (for instance, motivation, courage, entrepreneurial spirit, etc.). These persons, irrespective of whether they will participate in the program or not, have high intention and probability to start a business. Hence, both at the program end and currently, there are no statistically significant differences in the outcome indicators between program participants and non-participants (those who applied to the

program but failed the entrepreneurship test and were not selected to be trained and funded). This implies that the program has a dead-weight loss, meaning that the businesses would have been established even if there were no grants (public money) involved into them, or that these persons would have found wage employment anyway. In addition, it may also cautiously suggest that the training component of the program (training in entrepreneurship and preparation of business plan) was not effective in further improving the employment prospects and success outlook of those selected to get funding.

However, the so-called learning curve/effect suggests that over time administration of the program probably improved (the design itself, implementation by the ESA, etc.), as well its effectiveness. Following changes were also made to the program through years:

- the training component increased from 2 days to 30 hours active training;
- based on experts' advices and evaluation studies of similar programs elsewhere, the targeting of the program has been re-directed towards unemployed with shorter spells of unemployment (initially, main target group were those registered as unemployed for more than 2 years and through years declined to 6 months);
- initially, there was a rule that participants cannot re-enter the ESA registry in 3-year time. In 2010, an additional rule was added that those that started the program (training in preparation of business plan) and exited the program afterwards, cannot register as unemployed for at least 1 year. The program was made even stricter in 2011, when an obligation was introduced that if a person shuts down a business that was established within the program in less than 1 year from the registration, s/he had to pay back 80% of the grant (in a year).

There are important lessons and recommendations that can be drawn if the findings from this evaluation are combined with the results of the performance evaluation of self-employment program that was carried out by ILO (Corbanese, 2012). These are:

1. The program could contain a component related to business management and technical skills;
2. There could be control of the training quality;
3. The size of the program (i.e. number of program participants) could be reduced whereas the amount of the grant increased;
4. The training could be designed so that it does not necessarily leads to a grant even for the successful applicants. Instead, a possibility for assistance (e.g. business mentoring) for access to commercial credits could be offered to the selected business plans, especially those where commercialization is deemed to open more possibilities for business growth and development.
5. If recommendation 4 is accepted, a form of Guarantee fund could be established by the government, whereby businesses who were selected through the business plan competition but ultimately were directed to the commercial market for funding will obtain a guarantee against failure;
6. The amount of the grant given to applicants can vary according to the type of business and required technology for establishing and running the business,
7. Related to the previous recommendation, the government might set priority sectors/industries for which grants will be given;
8. The self-selection into application may be replaced with a random selection, whereby ESA invites randomly selected eligible persons to attend a training, with a possibility that they get the grant if the business plan they will prepare satisfy the present criteria for selection. Presently, this seems a bold recommendation;
9. If recommendation 8 is accepted, the training should be also tailored toward including motivation components, likely by presenting successful businesses which were erected from scratch;
10. If recommendation 8 is accepted, training participation should be probably motivated with a small cash grant, given that majority of the randomly selected participants into the training may lack initial motivation and courage to attend a training leading to establishing a business.

Further improvement of the program (and potentially, its effectiveness) is very important given that this program is the most expensive one in terms of cost per participant.

4.2 Internship program

We evaluate the internship program in two years 2010 and 2012. As argued in the previous sections, the design of the program slightly changed in-between. In particular, in the internship program 2012 an obligation was introduced for the host companies to employ at least 50% of the interns after the termination of the program. Hence, besides evaluating the impact of the program for participants against the non-participants, we opt to evaluate which of the two designs delivered better results. Therefore, we first analyse each of the two programs separately, while in the third sub-section we analyse them together.

4.2.1 Impact of 2010 Internship

Table 4-8 provides early evidence that the treatment and the control samples are systematically different in a couple of aspects; in addition, the Hotelling test rejects the null of joint equality between the means of the two groups at the 10%. In particular, interns are on average 2 years older than the non-treated; more frequently from urban areas; they live in smaller households, in particular where the number of unemployed members in the household is almost twice lower that of the control group; and larger percent of treated own a house. The two samples are different in terms of the outcome variables, treated being better positioned on the labour market than non-treated.

Table 4-8 Means' comparison – Internship 2010 program

Socio-demographic characteristics	mean control	mean treated	p-value	sig
Age (years)	29.91	32.11	0.04	**
Education today	7.28	7.65	0.15	
Educational level before treatment/cut-off	7.23	7.48	0.50	
Gender (1=male)	0.43	0.35	0.46	
Ethnicity (1=Macedonian)	0.87	0.87	0.90	
Ethnicity (1=Albanian)	0.10	0.09	0.92	
Marital status (1=married)	0.33	0.33	0.66	
Geography (1=urban)	0.88	0.97	0.10	*
Geography (1=suburban)	0.04	0.03	0.96	
Geography (1=rural)	0.06	-	0.02	**
Children (1=person has at least one child)	0.25	0.24	0.59	
Number of children	0.39	0.31	0.31	
Partner's employment status (1=if partner is employed)	0.25	0.29	0.98	
Number of household members	4.12	3.60	0.01	***
Number of household members below age 15	0.50	0.31	0.15	
Number of employed household members (age 15-64)	1.52	1.87	0.17	
Number of unemployed household members (age 15-64)	1.74	1.05	0.00	***
Number of retired household members	0.33	0.36	0.95	
House ownership (1=owned)	0.87	0.96	0.02	**
House size (sq. meters)	76.23	73.12	0.32	
Work experience (1=has work experience)	0.24	0.28	0.71	
Work experience (number of months)	96.00	10.29	0.96	
Contract (1=has written contract)	0.13	0.17	0.56	
Salary before treatment/cut-off (MKD)	2,176.99	1,986.67	0.57	
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	2.08	1.96	0.14	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.07	1.95	0.23	
Subjective employment status (1=employed)	0.41	0.63	0.00	***
Employment status (1=employed)	0.41	0.63	0.00	***
Unemployed at any time after program's end/cut-off (1=unemployed)	0.43	0.26	0.01	***
Employment status (1=inactive)	0.12	0.10	0.54	
Employed at program's end, but unemployed now	0.16	0.05	0.01	***
Wage per hour (MKD)	27.00	34.81	0.27	
Hotelling T-squared test H0: The vector of means of all variables are equal across groups				
p-value	0.0572			
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.				

Given the systematic difference between the two samples, Table 4-9 presents the results of a probit regression, to check which characteristics of a person lead to higher or lower probability that he/she is selected in the program. As suggested in Table 4-9, results suggest that older persons have higher probability to be selected as interns, up to a certain age. An additional household member reduces the chance that one is selected as an intern.

Table 4-9 Matching – Internship 2010 program

Variable	Marginal effects	p-value	sig
Age	0.10	0.00	***
Age squared	(0.00)	0.00	***
Education today	0.04	0.33	
Gender (1=male)	(0.04)	0.61	
Ethnicity (1=Macedonian)	(0.04)	0.87	
Ethnicity (1=Albanian)	0.07	0.78	
Marital status (1=married)	(0.06)	0.56	
Geography (1=urban)	0.24	0.13	
Geography (1=suburban)	0.16	0.50	
Number of children	(0.04)	0.60	
Number of household members	(0.06)	0.05	**
Work experience (1=has work experience)	(0.06)	0.47	
# Observations	224		
Log-pseudolikelihood	-142.15842		
Pseudo R ²	0.0844		

Source: Authors' calculations based on survey.
Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

The propensity score obtained from the previous regression is used to match the treatment and control samples, so as to obtain the unbiased impact of the internship programs on the battery of outcomes. These, along the evaluation of the matching quality are presented in the following Table 4-10.

Results robustly suggest that the internship 2010 program played sizeable role for the employment prospects of the involved individuals. In particular, the program led to current employment gains of about 25 p.p., both in terms of the actual and perceived employment status. Similarly, the program reduced the incidence of unemployment over the entire post-program period by 21 p.p., compared to non-participants, who actually face almost twice larger probability to stay unemployed over a prolonged period of time. The fact that the program does not exert influence on the employment rate at program's end suggests that the benefits of the program are reaped over longer time horizon only, which in our framework is about four years. Also, findings suggest that the program does not make difference in financial terms, as neither the financial condition nor the wage per hour was found systematically different between the treatment and control group. The program likely works through a couple of channels: the networking effects; the prevention of certain skills to erode; practicing and acquiring new skills, in particular, soft skills; and building a sense of organizational culture; all hence increasing the chances for faster and likely more suitable employment later. One should note here also that compared to all other ESA programs, the targeted individuals within the internship program are wider spread across all income groups of the households in Macedonia.

Table 4-10 Effects – Internship 2010 program

Outcome variable	Unmatched				One-to-one matching with replacement				Matching with caliper (0.01)						
	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig
Financial condition after treatment/cut-off (1=better; 3=worse)	1.96	2.08	(0.12)	(1.50)		1.95	2.11	(0.16)	(1.56)		1.94	2.11	(0.12)	(1.50)	
Bias reduction						Mean	-41%	Median	-41%		Mean	-36%	Median	-47%	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	1.95	2.07	(0.12)	(1.29)		1.86	2.03	(0.17)	(1.20)		1.81	2.00	(0.19)	(1.35)	
Bias reduction						Mean	36%	Median	45%		Mean	43%	Median	35%	
Subjective employment status (1=employed)	0.63	0.41	0.22	3.42	***	0.64	0.41	0.23	2.34	**	0.64	0.41	0.23	2.38	**
Bias reduction						Mean	-39%	Median	-40%		Mean	-35%	Median	-49%	
Employment status (1=employed)	0.64	0.41	0.23	3.57	***	0.65	0.40	0.25	2.57	**	0.65	0.40	0.25	2.61	***
Bias reduction						Mean	-38%	Median	-40%		Mean	-34%	Median	-49%	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.25	0.43	(0.18)	(2.90)	***	0.25	0.46	(0.21)	(2.14)	**	0.25	0.47	(0.21)	(2.16)	**
Bias reduction						Mean	-37%	Median	-40%		Mean	-34%	Median	-49%	
Employment status (1=inactive)	0.10	0.12	(0.02)	(0.59)		0.10	0.09	0.01	0.16		0.10	0.08	0.02	0.32	
Bias reduction						Mean	-50%	Median	-19%		Mean	-45%	Median	-27%	
Employed at program's end, but unemployed now	0.05	0.16	(0.11)	(2.57)	**	0.06	0.14	(0.09)	(1.21)		0.05	0.14	(0.09)	(1.22)	
Bias reduction						Mean	-45%	Median	-40%		Mean	-41%	Median	-49%	
Wage per hour (MKD)	35.12	27.00	8.13	1.15		36.15	29.37	6.79	0.59		36.86	28.83	8.02	0.70	
Bias reduction						Mean	-47%	Median	-40%		Mean	-42%	Median	-49%	

Source: Authors' calculations based on survey.

Note: *, **, and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

4.2.2 Impact of 2012 Internship

We iterate the same procedure as in 4.2.1 and provide the estimates for the 2012 internship program. Table 4-11 provides evidence that the treatment and the control samples are systematically different in a couple of aspects; in addition, the Hotelling test rejects the null of joint equality between the means of the two groups at the 1% (i.e. more strongly than with the 2010 internship program). Interns are on average 5 years older than the non-treated; with lower education before treatment, more frequently males or Albanians, from urban areas, they live in larger households (which is the opposite of the 2010 groups), in particular where the number of employed and retired members in the household is higher that of the control group; and live in larger houses. The two samples are different in terms of the outcome variables, treated being better positioned on the labour market than non-treated, on average.

Table 4-11 Means' comparison – Internship 2012 program

Socio-demographic characteristics	mean control	mean treated	p-value	sig
Age (years)	25.91	30.55	0.01	***
Education today	7.70	7.55	0.27	
Educational level before treatment/cut-off	7.64	7.26	0.01	***
Gender (1=male)	0.35	0.47	0.07	*
Ethnicity (1=Macedonian)	0.97	0.87	0.02	**
Ethnicity (1=Albanian)	0.02	0.09	0.06	*
Marital status (1=married)	0.23	0.24	0.85	
Geography (1=urban)	0.70	0.81	0.07	*
Geography (1=suburban)	0.05	0.05	0.97	
Geography (1=rural)	0.24	0.14	0.06	*
Children (1=person has at least one child)	0.13	0.14	0.78	
Number of children	0.17	0.19	0.80	
Partner's employment status (1=if partner is employed)	0.17	0.10	0.14	
Number of household members	3.96	4.27	0.05	**
Number of household members below age 15	0.26	0.21	0.53	
Number of employed household members (age 15-64)	1.82	2.09	0.07	*
Number of unemployed household members (age 15-64)	1.61	1.60	0.96	
Number of retired household members	0.21	0.35	0.08	*
House ownership (1=owned)	0.90	0.94	0.27	
House size (sq. meters)	78.21	86.19	0.04	**
Work experience (1=has work experience)	0.10	0.10	0.93	
Work experience (number of months)	2.94	3.64	0.78	
Contract (1=has written contract)	0.06	0.08	0.59	
Salary before treatment/cut-off (MKD)	813.95	788.46	0.95	
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	2.01	1.91	0.22	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	1.95	1.96	0.96	
Subjective employment status (1=employed)	0.39	0.67	0.00	***

Employment status (1=employed)	0.39	0.67	0.00	***
Unemployed at any time after program's end/cut-off (1=unemployed)	0.44	0.24	0.00	***
Employment status (1=inactive)	0.05	0.09	0.31	
Employed at program's end, but unemployed now	0.17	0.05	0.00	***
Wage per hour (MKD)	21.91	34.95	0.02	**
Hotelling T-squared test H0: The vector of means of all variables are equal across groups				
p-value				0.0000
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.				

Given the systematic difference between the two samples, Table 4-12 presents the results of a probit regression, to check which characteristics of a person lead to higher or lower probability that he/she is selected in the program. We only find that urban inhabitants have higher probability to be selected for participation in the program.

Table 4-12 Matching – Internship 2012 program

Variable	Marginal effects	p-value	sig
Age	0.04	0.28	
Age squared	(0.00)	0.51	
Education today	(0.05)	0.16	
Gender (1=male)	0.06	0.39	
Ethnicity (1=Macedonian)	(0.37)	0.21	
Ethnicity (1=Albanian)	0.07	0.83	
Marital status (1=married)	0.05	0.66	
Geography (1=urban)	0.21	0.01	***
Geography (1=suburban)	0.20	0.26	
Number of children	(0.07)	0.45	
Number of household members	0.04	0.15	
Work experience (1=has work experience)	(0.11)	0.27	
# Observations	249		
Log-pseudolikelihood	-136.24459		
Pseudo R ²	0.1199		
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.			

The propensity score obtained from the previous regression is used to match the treatment and control samples, so as to obtain the unbiased impact of the internship programs on the battery of outcomes. These, along the evaluation of the matching quality are presented in the following Table 4-13.

Table 4-13 Effects – Internship 2012 program

Outcome variable	Unmatched				One-to-one matching with replacement				Matching with caliper (0.01)						
	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig
Financial condition after treatment/cut-off (1=better; 3=worse)	1.91	2.01	(0.10)	(1.23)		1.94	2.14	(0.19)	(1.48)		1.91	2.12	(0.10)	(1.23)	
Bias reduction						Mean	-29%	Median	-61%		Mean	-31%	Median	-50%	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	1.96	1.95	0.01	0.04		1.95	1.90	0.05	0.29		1.95	1.90	0.05	0.28	
Bias reduction						Mean	-22%	Median	-27%		Mean	-5%	Median	23%	
Subjective employment status (1=employed)	0.67	0.39	0.27	4.15	***	0.65	0.35	0.31	2.89	***	0.66	0.34	0.32	3.01	***
Bias reduction						Mean	-29%	Median	-63%		Mean	-31%	Median	-53%	
Employment status (1=employed)	0.67	0.39	0.28	4.24	***	0.65	0.36	0.29	2.79	***	0.66	0.35	0.31	2.90	***
Bias reduction						Mean	-30%	Median	-63%		Mean	-32%	Median	-53%	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.24	0.44	(0.20)	(2.98)	***	0.25	0.42	(0.17)	(1.57)		0.26	0.43	(0.17)	(1.57)	
Bias reduction						Mean	-35%	Median	-63%		Mean	-38%	Median	-53%	
Employment status (1=inactive)	0.09	0.05	0.04	1.11		0.08	-	0.08	2.54	**	0.09	-	0.09	2.55	**
Bias reduction						Mean	-29%	Median	-61%		Mean	-31%	Median	-50%	
Employed at program's end, but unemployed now	0.05	0.18	(0.12)	(2.67)	***	0.06	0.22	(0.17)	(2.23)	**	0.05	0.22	(0.17)	(2.26)	**
Bias reduction						Mean	-28%	Median	-63%		Mean	-31%	Median	-53%	
Wage per hour (MKD)	34.95	22.04	12.91	2.24	**	36.08	16.46	19.62	2.70	***	35.70	14.30	21.40	2.86	***
Bias reduction						Mean	-28%	Median	-63%		Mean	-29%	Median	-53%	

Source: Authors' calculations based on survey.

Note: *, **, and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

Results robustly suggest that the internship 2012 program also played sizeable role for the employment prospects of the involved individuals. In particular, the program led to current employment gains of about 31 p.p., both in terms of the actual and perceived employment status. Interestingly though, results suggest that interns had quite lower employment rate at the end of the program than compared to the non-treated persons over the same period of time. Still, this may be in line with the finding of the 2010 program whereby unemployed at any time after program's end has been reduced for participants, in a sense that participants face larger employment opportunities in the long run only, i.e. they are not immediately hired by the host company at program end. That this may be the case is confirmed by the last finding whereby participants exhibit larger earnings today compared to non-participants, despite the latter were more frequently employed at program's-end time. Earnings are found more than twice larger, which is a sizeable impact in terms of the financial status that could be ascribed to the program.

4.2.3 Has Internship 2012 led to further gains over Internship 2010?

The quantitative assessment of the internship program impact in both 2010 and 2012 suggests that both programs bring gains for interns, predominantly over the longer time horizon. In terms of employment, both subjective and actual, both designs led to very favourable results, increasing participants employability by 25 to 31 p.p. (in 2010 and 2012, respectively) compared to non-participants. The better employment probability of participants is slightly higher in the 2012 program. This is probably related to the introduced obligation to host companies to hire at least 50% of the interns. Both designs did not produce short-term impacts; quite the contrary, the 2012 evaluation suggests that more frequently interns are not hired at the program end but reap the benefits of it only in the longer term. The 2012 program also found significant financial gains for participants.

Although we give a (slight) preference to internship program in 2012 in the above discussion, we still perform a quantitative evaluation to check which of the two exerted stronger impact. To do so, we perform the following exercise: we compare the treated in 2012 with the treated in 2010 and derive the difference in a similar manner as we do when comparing treated versus control groups. Designed in this way, the treatment dummy captures two influences: i) potential differential results between Internship 2010 and Internship 2012; and ii) differential macroeconomic and labour market conditions.

Table 4-14 presents the tests of the means differences and provides evidence that the two groups are still different in terms of the average observable demographics. This may suggest that the targeting of the two programs actually changed between the two years, or simply the program became more popular, so that the amount and variety of (selected) applicants increased. However, the insignificance of the differences in the outcome variables may actually suggest that, even if the targeting may have changed, the outcomes may have not.

Table 4-14 Means' comparison – Internship 2012 versus Internship 2010 programs

Socio-demographic characteristics	mean IN10	mean IN12	p-value	sig
Age (years)	31.21	30.55	0.72	
Education today	7.55	7.55	0.99	
Educational level before treatment/cut-off	7.41	7.26	0.30	
Gender (1=male)	0.36	0.47	0.11	
Ethnicity (1=Macedonian)	0.89	0.87	0.66	
Ethnicity (1=Albanian)	0.08	0.09	0.82	
Marital status (1=married)	0.32	0.24	0.24	
Geography (1=urban)	0.95	0.81	0.01	***
Geography (1=suburban)	0.03	0.05	0.41	
Geography (1=rural)	0.03	0.14	0.01	***
Children (1=person has at least one child)	0.24	0.14	0.08	*
Number of children	0.32	0.19	0.13	
Partner's employment status (1=if partner is employed)	0.27	0.10	0.00	***
Number of household members	3.68	4.27	0.00	***
Number of household members below age 15	0.29	0.21	0.31	
Number of employed household members (age 15-64)	1.78	2.09	0.05	*
Number of unemployed household members (age 15-64)	1.22	1.60	0.03	**
Number of retired household members	0.34	0.35	0.94	
House ownership (1=owned)	0.96	0.94	0.57	
House size (sq. meters)	76.19	86.19	0.03	**
Work experience (1=has work experience)	0.22	0.10	0.02	**
Work experience (number of months)	8.82	3.64	0.14	
Contract (1=has written contract)	0.14	0.08	0.14	
Salary before treatment/cut-off (MKD)	1,419.64	788.46	0.22	
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	1.96	1.91	0.53	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	1.95	1.96	0.96	
Subjective employment status (1=employed)	0.63	0.67	0.64	
Employment status (1=employed)	0.63	0.67	0.64	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.26	0.24	0.81	
Employment status (1=inactive)	0.10	0.09	0.84	
Employed at program's end, but unemployed now	0.05	0.05	0.94	
Wage per hour (MKD)	34.81	34.95	0.98	
Hotelling T-squared test H0: The vector of means of all variables are equal across groups				
p-value	0.0114			
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.				

Still, to verify this quantitatively, we match the two groups and then provide the average treatment effect. Indeed, Table 4-15 provides some evidence of changed targeting of the program. Younger individuals were more frequently a target of the 2012 program than compared to the 2010 one. Then, individuals with more education, of other ethnicity than Macedonian and Albanian, and from households with more household members had higher chance to be selected in 2012 than compared to 2010. If ESA did not change the targeting intentionally, this may indeed suggest that the announcement of the 2012 program had larger reach and more “different” applicants (both interns and companies) applied and became part of the program.

Table 4-15 Matching – Internship 2012 with Internship 2010 program

Variable	Marginal effects	p-value	sig
Age	(0.25)	-	***
Age squared	0.00	-	***
Education today	0.12	0.02	**
Gender (1=male)	0.09	0.31	
Ethnicity (1=Macedonian)	(0.61)	0.01	**
Ethnicity (1=Albanian)	(0.44)	0.01	***
Marital status (1=married)	(0.01)	0.91	
Geography (1=urban)	(0.29)	0.12	
Geography (1=suburban)	(0.24)	0.29	
Number of children	0.14	0.25	
Number of household members	0.13	0.01	***
Work experience (1=has work experience)	(0.08)	0.54	
# Observations	189		
Log-pseudolikelihood	-91.80105		
Pseudo R²	0.2834		

Source: Authors' calculations based on survey.

Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

Results after the matching suggest that the impact of the 2012 internship program has been hardly stronger than the one in 2010, in quantitative terms. When the matching with caliper is pursued, results suggest that the 2012 program is better in terms of employment, both actual and subjective, but the finding is only significant at the 10%. In addition, the caliper matching does not lead to any important efficiency gains over the usual nearest-neighbour matching: actually, the mean bias is less reduced with caliper, while the median bias is slightly more reduced with caliper, hence not giving firm grounds to claim that the finding of employment gains of 2012 program over the one in 2010 is actually statistically significant.

Table 4-16 Effects – Internship 2012 over Internship 2010

Outcome variable	Unmatched				One-to-one matching with replacement				Matching with caliper(0.01)						
	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig
Financial condition after treatment/cut-off (1=better; 3=worse)	1.91	1.96	(0.05)	(0.62)		1.91	1.87	0.04	0.25		1.88	2.02	(0.14)	(0.95)	
<i>Bias reduction</i>						Mean	-57%	Median	-65%		Mean	-38%	Median	-52%	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	1.96	1.95	0.01	0.58							1.93	2.07	(0.13)	(0.80)	
<i>Bias reduction</i>						Mean		Median			Mean	-45%	Median	-45%	
Subjective employment status (1=employed)	0.67	0.63	0.04	0.51		0.67	0.47	0.19	1.47		0.67	0.43	0.24	1.95	*
<i>Bias reduction</i>						Mean	-43%	Median	-48%		Mean	-25%	Median	-49%	
Employment status (1=employed)	0.67	0.64	0.03	0.38		0.67	0.47	0.19	1.47		0.67	0.43	0.24	1.95	*
<i>Bias reduction</i>						Mean	-43%	Median	-48%		Mean	-25%	Median	-49%	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.24	0.25	(0.01)	(0.13)		0.24	0.36	(0.12)	(0.92)		0.24	0.38	(0.14)	(1.18)	
<i>Bias reduction</i>						Mean	-47%	Median	-48%		Mean	-31%	Median	-49%	
Employment status (1=inactive)	0.09	0.10	(0.01)	(0.21)		0.09	0.13	(0.04)	(0.39)		0.10	0.17	(0.07)	(0.74)	
<i>Bias reduction</i>						Mean	-52%	Median	-48%		Mean	-34%	Median	-49%	
Employed at program's end, but unemployed now	0.05	0.05	(0.00)	(0.08)		0.05	0.12	(0.06)	(0.88)		0.05	0.12	(0.07)	(1.00)	
<i>Bias reduction</i>						Mean	-46%	Median	-48%		Mean	-31%	Median	0%	
Wage per hour (MKD)	34.95	35.12	(0.17)	(0.03)		34.95	29.38	5.57	0.47		35.47	18.29	17.19	1.52	
<i>Bias reduction</i>						Mean	-52%	Median	-48%		Mean	-28%	Median	-49%	

Source: Authors' calculations based on survey.

Note: *, **, and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

In qualitative fashion, still 2012 internship produces slightly more gains for interns than compared to the 2010 design and hence should be maintained. Note that the results in Table 4-16 disprove the claim that macro and labour-market conditions in 2012 significantly differed from those in 2010, since had they differed it would have been in positive direction only (i.e. 2012 cannot be claimed to have been worse economic year than 2010). Hence, it is unlikely that worse economic conditions compensate the potentially positive effect of 2012 program over the one in 2010. The qualitatively better results for the 2012 program may be actually revealing the better administration of the program and confirming our initial notion that over time the pool of applicants (and selected applicants) and host companies may have become more diverse in both observable and unobservable characteristics. Though, we cannot claim that the better outcomes of the 2012 program are related to the obligation introduced in 2012 that employers hire at least 50% of the interns at the program end.

4.2.4 Summary of findings and recommendations

The evaluation of the internship programs (for both years) shows that the program is effective in terms of employment outcomes of the program participants. Particularly, the program has no significant impact on the participants after the program end (meaning after the end of the internship), but has large positive effects on the subsequent employment. Similarly, the subjective employment is higher for the treatment group. There is difference between the two groups in employment after program end/cut-off date only in 2012. The comparison between the program in 2010 and 2012 suggests that effects are similar, but of a higher magnitude in 2012. The 2012 program also found significant financial gains for participants, which is not the case with the 2010 program.

These results show that employers do not immediately hire the interns, but once there is a need for a worker, their former interns are first to hire.⁴ In addition, the interns have higher probability of employment relative to young persons who were not taking part in this program as employers are favouring workers with previous work experience. In general, better employment prospects come from different channels such as networking while on internship, new skills gained, initial on-the-job training received, but also maybe the signalling effect (interns signalling to employers that they are of better quality, more motivated, etc. compared to their peers). Our results do not show the presence of the so-called internship “trap” whereby young persons are trapped into internships, moving from one to next, without being hired, and only substituting the regular staff.

Based on the evaluation, we propose following changes to the program:

1. As the evaluation shows that both programs (2010 and 2012) are effective, the Ministry may lift the obligation for employers to hire 50% of the interns. This can increase the intake by the employers, and still in the future contribute to greater employment prospects of the interns;
2. We suggest that the government aligns further this program with similar programs (and terminology) of the EU countries. Such programs are generally implemented across the EU countries as traineeship programs rather than internship programs. The evaluation of the EU traineeship programs shows that they are effective only when they are well-structured (EC, 2013);
3. In line with the experience of EU countries, the length of the internship could be made more flexible and of longer duration, ranging from 3 to 9 months, rather than the current 3 months, hence giving more time that both sides (the intern and the host company) reap the benefits of the internship;
4. In order to increase the learning experience and gained knowledge and skills by interns, the internship program could involve an internship agreement stipulating results to be achieved and knowledge to be gained by the intern. This agreement could be signed by the host company, the intern and the ESA. The agreement (with specifically outlined learning content of the internship) will increase the future value of the internship for the interns, as they are otherwise not getting any formal qualification at the end of the program. Moreover, the agreement can be used in future for better monitoring of the program;

4 We cannot assess with the data if interns are hired by the same company where they were on internship.

5. Before starting the internship, each person could receive an initial training organized by the ESA in job-related skills (job readiness training).

4.3 Training for known employer

We evaluate the training for known employer (TKE) program in two years 2010 and 2012. As argued in the previous sections, the design of the program slightly changed in-between. In particular, the obligation of employers to employ at least 70% of the participants in 2010 was reduced to 50% of the participants in 2012. Hence, besides evaluating the impact of the program for participants against the non-participants, we opt to evaluate which of the two designs delivered better results. Therefore, we first analyse each of the two programs separately, while in the third sub-section we analyse them together.

4.3.1 Impact of 2010 TKE

Table 4-17 gives the comparisons between the unmatched data. The evidence suggests that the two samples: the treated and the control one are systematically different, both observed by individual variables and jointly through the Hotelling test. The treated individuals are on average younger and more educated than the control individuals; the share of Macedonians is larger in the former than in the latter group; treated have on average fewer children and more employed members in the household. Participants have less work experience than non-participants. The two groups are also different in terms of the outcomes: the treated are better positioned on the labour market than non-treated, both in terms of observed labour-market outcomes and self-assessed wellbeing.

Table 4-17 Means' comparison – TKE 2010 program

Socio-demographic characteristics	mean control	mean treated	p-value	sig
Age (years)	43.99	36.21	0.00	***
Education today	5.05	6.17	0.00	***
Educational level before treatment/cut-off	5.05	6.03	0.00	***
Gender (1=male)	0.53	0.43	0.12	
Ethnicity (1=Macedonian)	0.84	0.95	0.00	***
Ethnicity (1=Albanian)	0.06	0.04	0.44	
Marital status (1=married)	0.70	0.67	0.55	
Geography (1=urban)	0.81	0.75	0.29	
Geography (1=suburban)	0.07	-	0.00	***
Geography (1=rural)	0.12	0.25	0.02	**
Children (1=person has at least one child)	0.72	0.64	0.19	
Number of children	1.45	1.19	0.06	*
Partner's employment status (1=if partner is employed)	0.35	0.37	0.71	
Number of household members	3.85	4.19	0.05	*
Number of household members below age 15	0.71	0.68	0.83	
Number of employed household members (age 15-64)	0.91	1.56	0.00	***
Number of unemployed household members (age 15-64)	1.84	1.67	0.30	
Number of retired household members	0.36	0.29	0.35	
House ownership (1=owned)	0.87	0.96	0.00	***
House size (sq. meters)	74.58	79.81	0.23	
Work experience (1=has work experience)	0.59	0.47	0.06	*
Work experience (number of months)	91.00	55.95	0.01	***
Contract (1=has written contract)	0.37	0.36	0.84	
Salary before treatment/cut-off (MKD)	4,487.83	4,433.33	0.94	
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	2.23	2.13	0.12	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.39	1.91	0.00	***
Subjective employment status (1=employed)	0.20	0.49	0.00	***
Employment status (1=employed)	0.23	0.53	0.00	***
Unemployed at any time after program's end/cut-off (1=unemployed)	0.61	0.37	0.00	***
Employment status (1=inactive)	0.30	0.11	0.00	***
Employed at program's end, but unemployed now	0.16	0.08	0.05	**
Wage per hour (MKD)	16.45	29.97	0.04	**
Hotelling T-squared test H0: The vector of means of all variables are equal across groups				
p-value	0.0000			
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.				

Table 4-18 presents the results of the probit estimation for estimating the probability that one gets selected for the program. Surprisingly, only the urban variable is significant, suggesting that persons living in urban areas have slightly lower probability to be selected for the program than rural inhabitants.

Table 4-18 Matching – TKE 2010 program

Variable	Marginal effects	p-value	sig
Age	(0.01)	0.58	
Age squared	0.00	0.90	
Education today	0.08	-	***
Gender (1=male)	(0.08)	0.08	*
Ethnicity (1=Macedonian)	0.15	0.17	
Ethnicity (1=Albanian)	(0.05)	0.76	
Marital status (1=married)	(0.04)	0.58	
Geography (1=urban)	(0.19)	0.01	***
Number of children	(0.03)	0.33	
Number of household members	0.04	0.02	**
Work experience (1=has work experience)	0.00	0.97	
# Observations	313		
Log-pseudolikelihood	-139.29907		
Pseudo R ²	0.1918		

Source: Authors' calculations based on survey.
 Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

The propensity score is then used to conduct the matching, which is presented in Table 4-19. Results robustly suggest that the program exerted positive influence for both participants' self-assessed wellbeing and the observed labour-market results. In particular, the program is found to have increased the chances of finding a job by nearly 50 p.p., as well the subjective and observed employment probabilities by large 35 and 42 p.p., respectively. It is also found to have led to sizeable monetary gains, leading to more than threefold larger earnings per hour than compared to the non-participants.

Table 4-19 Results – TKE 2010 program

Outcome variable	Unmatched				One-to-one matching with replacement				Matching with caliper (0.01)						
	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig
Financial condition after treatment/cut-off (1=better, 3=worse)	2.13	2.21	(0.07)	(0.92)		2.12	2.15	(0.03)	(0.31)		2.13	2.15	(0.02)	(0.17)	
<i>Bias reduction</i>						Mean	-70%	Median	-73%		Mean	-72%	Median	-69%	
Chances to find a job after treatment/cut-off (1=better, 3=worse)	1.91	2.39	(0.48)	(4.12)	***	1.90	2.33	(0.43)	(3.06)	***	1.88	2.35	(0.46)	(2.96)	***
<i>Bias reduction</i>						Mean	-17%	Median	79%		Mean	-28%	Median	101%	
Subjective employment status (1=employed)	0.49	0.20	0.29	5.14	***	0.51	0.20	0.31	3.46	***	0.53	0.18	0.35	3.91	***
<i>Bias reduction</i>						Mean	-59%	Median	-67%		Mean	-57%	Median	-68%	
Employment status (1=employed)	0.53	0.23	0.30	5.16	***	0.54	0.14	0.40	4.60	***	0.57	0.15	0.42	4.77	***
<i>Bias reduction</i>						Mean	-54%	Median	-67%		Mean	-55%	Median	-68%	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.37	0.62	(0.25)	(3.86)	***	0.35	0.62	(0.26)	(2.62)	***	0.32	0.62	(0.30)	(3.06)	***
<i>Bias reduction</i>						Mean	-61%	Median	-67%		Mean	-60%	Median	-68%	
Employment status (1=inactive)	0.11	0.29	(0.18)	(3.19)	***	0.11	0.23	(0.12)	(1.53)		0.10	0.22	(0.12)	(1.48)	
<i>Bias reduction</i>						Mean	-66%	Median	-67%		Mean	-67%	Median	-68%	
Employed at program's end, but unemployed now	0.08	0.15	(0.07)	(1.50)		0.09	0.20	(0.11)	(1.43)		0.10	0.22	(0.12)	(1.55)	
<i>Bias reduction</i>						Mean	-63%	Median	-63%		Mean	-64%	Median	-65%	
Wage per hour (MKD)	29.97	17.25	12.71	1.62		27.01	7.42	19.59	2.11	**	28.92	8.04	20.88	2.26	**
<i>Bias reduction</i>						Mean	-62%	Median	-63%		Mean	-63%	Median	-65%	

Source: Authors' calculations based on survey.

Note: *, **, and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

4.3.2 Impact of 2012 TKE

The same steps as in 4.3.1 are conducted for the TKE program in 2012. Table 4-20 suggests that selection on observables is present in this case also. Similarly as in 2010, the participants in 2012 were on average younger and more educated than non-participants, but distinctly than in 2010, selected were more frequently women. In 2012, participants were also different from non-participants in terms of whether they have a child or not, the number of children and partner's labour market status. Moreover, treated had shorter working history and more frequently a written contract than control group. Also, their positioning on the labour market has been quite better than non-participants, similarly as in 2010. Hence, selection on observables has a strong role to play.

Table 4-20 Means' comparison – TKE 2012 program

Socio-demographic characteristics	mean control	mean treated	p-value	sig
Age (years)	46.13	36.70	0.00	***
Education today	4.20	6.14	0.00	***
Educational level before treatment/cut-off	4.20	6.02	0.00	***
Gender (1=male)	0.60	0.33	0.00	***
Ethnicity (1=Macedonian)	0.88	0.96	0.01	***
Ethnicity (1=Albanian)	0.02	0.04	0.34	
Marital status (1=married)	0.78	0.73	0.29	
Geography (1=urban)	0.60	0.69	0.08	*
Geography (1=suburban)	0.09	-	0.00	***
Geography (1=rural)	0.31	0.31	0.92	
Children (1=person has at least one child)	0.84	0.60	0.00	***
Number of children	1.75	1.12	0.00	***
Partner's employment status (1=if partner is employed)	0.26	0.44	0.00	***
Number of household members	3.82	3.90	0.64	
Number of household members below age 15	0.65	0.59	0.61	
Number of employed household members (age 15-64)	0.55	1.48	0.00	***
Number of unemployed household members (age 15-64)	2.44	1.58	0.00	***
Number of retired household members	0.25	0.25	0.99	
House ownership (1=owned)	0.96	0.84	0.00	***
House size (sq. meters)	77.62	79.28	0.62	
Work experience (1=has work experience)	0.58	0.55	0.60	
Work experience (number of months)	96.03	56.19	0.00	***
Contract (1=has written contract)	0.38	0.50	0.04	**
Salary before treatment/cut-off (MKD)	4,396.17	5,369.90	0.15	
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	2.35	2.24	0.15	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.59	2.14	0.00	***
Subjective employment status (1=employed)	0.10	0.52	0.00	***
Employment status (1=employed)	0.18	0.52	0.00	***

Unemployed at any time after program's end/cut-off (1=unemployed)	0.74	0.35	0.00	***
Employment status (1=inactive)	0.35	0.12	0.00	***
Employed at program's end, but unemployed now	0.08	0.14	0.08	*
Wage per hour (MKD)	4.56	29.13	0.00	***
Hotelling T-squared test H0: The vector of means of all variables are equal across groups				
p-value	0.0000			
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.				

That this is the case, Table 4-21 presents the probit results, whereby more variables than in the case of 2010 were found to have played a role for the selection into treatment: more educated individuals and women had larger role to be selected to TKE 2012, as well Macedonians compared to other ethnicities; as well those with fewer children, more household members and with working history.

Table 4-21 Matching – TKE 2012 program

Variable	Marginal effects	p-value	sig
Age	(0.01)	0.62	
Age squared	(0.00)	0.98	
Education today	0.21	0.00	***
Gender (1=male)	(0.25)	0.00	***
Ethnicity (1=Macedonian)	0.25	0.07	*
Ethnicity (1=Albanian)	0.39	0.20	
Marital status (1=married)	0.09	0.37	
Geography (1=urban)	(0.07)	0.30	
Number of children	(0.12)	0.01	**
Number of household members	0.05	0.04	**
Work experience (1=has work experience)	0.14	0.05	**
# Observations	324		
Log-pseudolikelihood	-115.41533		
Pseudo R²	0.4492		
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.			

The propensity scores from the previous step were used to match the two samples and obtain comparable ones on observables, so as to appropriately measure the program's effect. Results are presented in the following Table 4-22:

Table 4-22 Effects – TKE 2012 program

Outcome variable	Unmatched				One-to-one matching with replacement				Matching with caliper(0.01)						
	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig
Financial condition after treatment/cut-off (1=better; 3=worse)	2.24	2.33	(0.09)	(1.26)		2.28	2.32	(0.04)	(0.31)		2.26	2.28	(0.02)	(0.14)	
<i>Bias reduction</i>						Mean	-70%	Median	-34%		Mean	-73%	Median	-39%	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.14	2.58	(0.44)	(4.87)	***	2.16	2.52	(0.35)	(2.11)	**	2.10	2.55	(0.45)	(2.47)	**
<i>Bias reduction</i>						Mean	-39%	Median	0%		Mean	-39%	Median	-20%	
Subjective employment status (1=employed)	0.52	0.09	0.43	9.67	***	0.54	0.06	0.48	5.50	***	0.53	0.06	0.47	5.41	***
<i>Bias reduction</i>						Mean	-55%	Median	-47%		Mean	-58%	Median	-59%	
Employment status (1=employed)	0.52	0.19	0.33	6.58	***	0.55	0.17	0.38	3.50	***	0.53	0.19	0.34	3.30	***
<i>Bias reduction</i>						Mean	-59%	Median	-47%		Mean	-63%	Median	-59%	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.35	0.74	(0.39)	(7.46)	***	0.37	0.59	(0.23)	(1.86)	*	0.40	0.66	(0.26)	(2.37)	**
<i>Bias reduction</i>						Mean	-66%	Median	-47%		Mean	-67%	Median	-59%	
Employment status (1=inactive)	0.12	0.32	(0.19)	(3.91)	***	0.10	0.38	(0.28)	(2.38)	**	0.11	0.40	(0.28)	(2.72)	***
<i>Bias reduction</i>						Mean	-59%	Median	-47%		Mean	-62%	Median	-59%	
Employed at program's end, but unemployed now	0.14	0.07	0.08	2.24	**	0.10	0.24	(0.14)	(1.75)	*	0.08	0.15	(0.08)	(1.07)	
<i>Bias reduction</i>						Mean	-62%	Median	-39%		Mean	-70%	Median	-53%	
Wage per hour (MKD)	29.13	3.53	25.59	9.01	***	29.72	2.08	27.64	5.80	***	25.57	1.93	23.65	4.88	***
<i>Bias reduction</i>						Mean	-57%	Median	-47%		Mean	-62%	Median	-59%	

Source: Authors' calculations based on survey.

Note: *, **, and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

Results suggest that TKE 2012 led to gains for the treated participants. Chances for finding a job increased by about 45 p.p. compared to non-participants, as well the subjective employment by large 47 p.p. The actual employment changed by large 34 p.p. The incidence of unemployment and inactivity over the entire post-program period declined by 26 and 28 p.p., respectively, while the average wage rate has been significantly larger for treated than compared to non-treated.

4.3.3 Has TKE 2012 led to further gains over TKE 2010?

Both TKE 2010 and 2012 were found to have led to sizeable gains for the persons selected in the program. The program showed similar results in the two observed years: participants declared that both chances for finding a job and their individual perception of employment opportunities significantly improved, and the actual labour market outcomes confirm this: current employment increased sizeably, as well the wage rate. Effects, however, remain mixed. TKE 2010 is found stronger in the gains in terms of actual employment, while TKE 2012 in terms of subjective employment and wages. Overall, both evaluations suggested the programs have rather longer-term impact, as the employment at program's end was found indifferent than the one of non-participants.

We further perform a quantitative evaluation to check which of the two exerted stronger impact. To do so, we perform the following exercise: we compare the treated in 2012 with the treated in 2010 and derive the difference in a similar manner as we do when comparing treated versus control groups. Designed in this way, the treatment dummy captures two influences: i) potential differential results between TKE 2010 and TKE 2012; and ii) differential macroeconomic and labour market conditions.

Table 4-23 presents the tests of the means differences and rather provides strong evidence that the two groups are quite similar in terms of the average observable demographics. Slight difference arises in terms of the chances to find a job, but these will be evaluated in the next table.

Table 4-23 Means' comparison – TKE 2010 versus TKE 2012 programs

Socio-demographic characteristics	mean KE10	mean KE12	p-value	sig
Age (years)	36.21	36.70	0.76	
Education today	6.17	6.14	0.88	
Educational level before treatment/cut-off	6.03	6.02	0.96	
Gender (1=male)	0.43	0.33	0.17	
Ethnicity (1=Macedonian)	0.95	0.96	0.78	
Ethnicity (1=Albanian)	0.04	0.04	0.87	
Marital status (1=married)	0.67	0.73	0.39	
Geography (1=urban)	0.75	0.69	0.40	
Geography (1=suburban)	-	-	.	
Geography (1=rural)	0.25	0.31	0.40	
Children (1=person has at least one child)	0.64	0.60	0.60	
Number of children	1.19	1.12	0.64	
Partner's employment status (1=if partner is employed)	0.37	0.44	0.35	
Number of household members	4.19	3.90	0.15	
Number of household members below age 15	0.68	0.59	0.52	
Number of employed household members (age 15-64)	1.56	1.48	0.57	
Number of unemployed household members (age 15-64)	1.67	1.58	0.66	
Number of retired household members	0.29	0.25	0.57	
House ownership (1=owned)	0.96	0.84	0.00	***
House size (sq. meters)	79.81	79.28	0.91	
Work experience (1=has work experience)	0.47	0.55	0.27	
Work experience (number of months)	55.95	56.19	0.99	
Contract (1=has written contract)	0.36	0.50	0.07	*
Salary before treatment/cut-off (MKD)	4,433.33	5,369.90	0.29	
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	3.09	2.99	0.34	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.13	2.24	0.19	
Subjective employment status (1=employed)	1.91	2.14	0.03	**
Employment status (1=employed)	0.49	0.52	0.70	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.53	0.52	0.88	
Employment status (1=inactive)	0.37	0.35	0.70	
Employed at program's end, but unemployed now	0.11	0.12	0.72	
Wage per hour (MKD)	0.08	0.14	0.18	
Hotelling T-squared test H0: The vector of means of all variables are equal across groups				
p-value	0.6734			
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.				

As selection on observables is not active in this particular case, we only present only Table 4-24 whereby the outcome variables are regressed on a dummy taking a value of 1 for TKE 2012 and 0 for TKE 2010. Results strongly confirm our qualitative assessment that there is no statistically significant difference in actual/observed outcomes between TKE 2010 and TKE 2012. However, the evidence suggests that the TKE 2012 exerts more positive influence on participants in terms of their perception for employability. Namely, participants in 2012 have higher inclination to consider themselves employed than compared with participants in 2010 by 61 to 75%. Hence, the present design of TKE could be maintained without prejudice that it performs worse than the earlier design. The insignificant results for the observed outcomes also confirm that the macroeconomic conditions in 2012 were not better than in 2010, what we also documented when comparing the Internship programs above.

Table 4-24 Effects – TKE 2012 over TKE 2010 program

Outcome variable	Only treatment included as explanatory			Treatment and a bunch of other explanatory variables		
	coefficient	p-value	sig	coefficient	p-value	sig
Financial condition after treatment/cut-off (1=better; 3=worse)	0.19	0.26		0.19	0.28	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	0.61	0.03	**	0.75	0.03	**
Subjective employment status (1=employed)	0.07	0.70		0.07	0.73	
Employment status (1=employed)	(0.03)	0.88		(0.02)	0.93	
Unemployed at any time after program's end/cut-off (1=unemployed)	(0.08)	0.69		(0.08)	0.69	
Employment status (1=inactive)	0.09	0.72		(0.05)	0.87	
Employed at program's end, but unemployed now	0.33	0.20		0.30	0.28	
Wage per hour (MKD)	(0.84)	0.89		0.74	0.90	

Source: Authors' calculations based on survey.
Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

4.3.4 Summary of findings and recommendations

Program evaluation has shown that training for known employer brings large gains for the participants. It is important to note that these gains are not only related to the obligation of the employer to keep at least 70% of the participants (actually, those that successfully completed the training) on the job, but the program brings long-term positive effects for the program participants (i.e. after the expiry of the obligation). In addition, program participants believe that the program has improved their chances to find a job. The evaluation also finds statistically higher wages of program participants in both years.

The test we ran did not lend evidence that there are statistically significant differences between the programme effects in the two years, except for the perception that chances for getting a job was higher in 2012 than in 2010. Although, the latter may also reflect more optimistic outlook for the overall economy. Results suggest that the program can be kept in the current content. In addition, we observe the learning effect, such that the program implementation is improved with the time as more experience is gained by the ESA.

The results also show that the current employment rate of individuals participating in the 2010 program is 57%, meaning that there is diminishing effect on employment of the program with time. It also suggests that the program may be used by employers to cover their short term needs for labour (or that they are not able to plan their labour needs in a longer time span). These findings are in line with the results obtained by the performance measurement of the training program (Corbanese, 2012).

Potential recommendations for further improvement of the program are:

1. Priority in this program should be given to employers from sectors which are set by the government as priority ones for the future development of the country, or those that are most propulsive sectors and have high growth potential. This will increase employment rate after program end;
2. An external assessment system can be set up to assess the effectiveness of the training. The quality of the training is important for improving the longer-term employment prospects of the participants (after the expiry of the obligation to employers to keep trained workers on the job).

4.4 Wage subsidy

We start the evaluation of the wage subsidy program by comparing the means of observables across the treatment and control groups (see Table 4-25). Results suggest that selection is likely present, as judged by the individual differences and the Hotelling test at the bottom of the table. Treated individuals were found older, more frequently from urban areas, with slightly more children and with higher share of employed and retired persons in the household; with longer employment experience, more frequently with a written contract, but with lower salary on previous employment. The treated are also identified with worse positioning on the labour market than controls. Hence, selection on observables is likely in place.

Table 4-25 Means' comparison – Wage subsidy programs

Socio-demographic characteristics	mean control	mean treated	p-value	sig
Age (years)	41.96	46.44	0.00	***
Education today	6.08	6.23	0.16	
Educational level before treatment/cut-off	6.06	6.18	0.24	
Gender (1=male)	0.56	0.57	0.71	
Ethnicity (1=Macedonian)	0.90	0.93	0.09	*
Ethnicity (1=Albanian)	0.03	0.03	0.68	
Marital status (1=married)	0.72	0.75	0.36	
Geography (1=urban)	0.70	0.82	0.00	***
Geography (1=suburban)	0.06	0.03	0.01	***
Geography (1=rural)	0.23	0.15	0.00	***
Children (1=person has at least one child)	0.75	0.75	0.91	
Number of children	1.42	1.56	0.06	*
Partner's employment status (1=if partner is employed)	0.35	0.40	0.14	
Number of household members	3.96	4.01	0.68	
Number of household members below age 15	0.57	0.60	0.67	
Number of employed household members (age 15-64)	1.24	1.40	0.03	**
Number of unemployed household members (age 15-64)	1.90	1.69	0.02	**
Number of retired household members	0.23	0.32	0.04	**
House ownership (1=owned)	0.90	0.92	0.28	
House size (sq. meters)	82.08	87.32	0.09	*
Work experience (1=has work experience)	0.65	0.73	0.03	**
Work experience (number of months)	115.71	133.07	0.09	*
Contract (1=has written contract)	0.48	0.55	0.04	**
Salary before treatment/cut-off (MKD)	6,135.24	5,137.11	0.04	**

Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	2.26	2.24	0.69	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.45	2.54	0.08	*
Subjective employment status (1=employed)	0.38	0.42	0.23	
Employment status (1=employed)	0.40	0.36	0.24	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.41	0.54	0.00	***
Employment status (1=inactive)	0.25	0.25	0.84	
Employed at program's end, but unemployed now	0.17	0.09	0.00	***
Wage per hour (MKD)	22.48	21.93	0.89	
Hotelling T-squared test H0: The vector of means of all variables are equal across groups				
p-value	0.0000			
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.				

The probit estimates in Table 4-26, suggest that more educated and persons living in urban areas have larger probability to be selected for the program, as well those with previous working experience. On the contrary, the higher the number of months with previous working experience and the higher the previous salary, the lower the probability of being selected.

Table 4-26 Matching – Wage subsidy program

Variable	Marginal effects	p-value	sig
Age	-0.0042759	0.696	
Age squared	0.0001014	0.393	
Education today	0.0224984	0.078	*
Gender (1=male)	0.0405567	0.223	
Ethnicity (1=Macedonian)	0.0988146	0.165	
Ethnicity (1=Albanian)	0.1061411	0.382	
Marital status (1=married)	-0.0380216	0.405	
Geography (1=urban)	0.0877842	0.039	**
Geography (1=suburban)	-0.0689996	0.407	
Number of children	0.0263074	0.265	
Number of household members	0.0154538	0.148	
Work experience (1=has work experience)	0.1142615	0.041	**
Work experience (number of months)	-1.19	0.0001	***
Contract (1=written contract)	0.0170609	0.707	
Salary before treatment/cut-off (MKD)	-3.54	0.0000	***
# Observations	887		
Log-pseudolikelihood	-528.9812		
Pseudo R ²	0.0489		
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.			

We use the probit estimates to arrive at propensity scores to be used for matching. Results of the matching and its evaluation are presented in the next Table 4-27. After matching, results suggest that the wage subsidy program exerts only one significant effect: the one on employment at program's end. Surprisingly the effect is negative, i.e. after program finishes, participants face twice smaller perspectives for employment than non-participants.

4.4.1 Summary of findings and recommendations

The impact evaluation of the wage subsidy program shows that the program is not effective in improving the long-term chances of participants to find and sustain a job in an open labour market. There is no statistically significant effect of the program on none of the outcomes studied here, neither actual ones (current employment, inactivity, non-employment) nor subjective ones. The only difference is in the employment after program end, which is actually lower for the program participants. This suggests that employers immediately fire workers after the period in which they are obliged to keep the workers (12 months following the 6 months subsidy period); the program is used by employers mainly to overcome short-term needs for workers. In addition, there might be a so-called stigma effect (usually found in public-works programs), so that potential future employers do not hire workers than have previously been part of the wage subsidy program. This works opposite to the expectation that the work experience gained within the program will improve the likelihood of the participants to find a job (possibly also with another employer).

Some recommendations for improvement of the program are:

1. Introduce training before assigning unemployed to the program;
2. Enhance certification program for training offered by employers to increase the long-term value of the training and the experience gained while in program.

In the current setting, we propose that the program is not implemented anymore.

Table 4-27 Effects – Wage subsidy program

Outcome variable	Unmatched					One-to-one matching with replacement					Matching with caliper(0.01)				
	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig
Financial condition after treatment/cut-off (1=better; 3=worse)	2.24	2.26	(0.02)	(0.38)		2.24	2.23	0.01	0.11		2.24	2.23	0.01	0.11	
<i>Bias reduction</i>						Mean	-63%	Median	-56%		Mean	-63%	Median	-56%	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.54	2.45	0.09	1.63		2.54	2.49	0.05	0.60		2.54	2.49	0.05	0.60	
<i>Bias reduction</i>						Mean	-51%	Median	-52%		Mean	-51%	Median	-52%	
Subjective employment status (1=employed)	0.42	0.38	0.04	1.22		0.42	0.34	0.08	1.65		0.42	0.34	0.08	1.65	
<i>Bias reduction</i>						Mean	-52%	Median	-39%		Mean	-52%	Median	-39%	
Employment status (1=employed)	0.36	0.40	(0.04)	(1.16)		0.36	0.37	(0.01)	(0.21)		0.36	0.37	(0.01)	(0.21)	
<i>Bias reduction</i>						Mean	-60%	Median	-59%		Mean	-60%	Median	-59%	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.54	0.41	0.12	3.48	***	0.54	0.46	0.08	1.47		0.54	0.46	0.08	1.47	
<i>Bias reduction</i>						Mean	-57%	Median	-47%		Mean	-57%	Median	-47%	
Employment status (1=inactive)	0.25	0.25	0.01	0.21		0.25	0.29	(0.04)	(0.74)		0.25	0.29	(0.04)	(0.74)	
<i>Bias reduction</i>						Mean	-55%	Median	-39%		Mean	-55%	Median	-39%	
Employed at program's end, but unemployed now	0.09	0.17	(0.08)	(3.14)	***	0.09	0.18	(0.09)	(2.17)	**	0.09	0.18	(0.09)	(2.17)	**
<i>Bias reduction</i>						Mean	-51%	Median	-47%		Mean	-51%	Median	-47%	
Wage per hour (MKD)	21.93	22.48	(0.55)	(0.16)		21.93	18.52	3.42	0.70		21.93	18.52	3.42	0.70	
<i>Bias reduction</i>						Mean	-56%	Median	-44%		Mean	-56%	Median	-44%	

Source: Authors' calculations based on survey.

Note: *, **, and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

4.5 Training for deficient occupations

Table 4-28 presents the tests for the means differences between the treatment and control group characteristics for the training for deficient occupations program. Both individual and the aggregate test suggest that the two groups have systematically different characteristics, i.e. there is selection on observables. All treated persons were men and from Macedonian ethnic origin, less frequently married, younger, without or with fewer children and with a partner less frequently employed than compared to non-treated. Their working history has been scarcer than compared to the control group and they less frequently had written contract. Not many differences are observed in terms of the outcomes, except for the inactivity whereby participants were less inactive than non-participants.

Table 4-28 Means' comparison – Training for deficient occupations program

Socio-demographic characteristics	mean control	mean treated	p-value	sig
Age (years)	43.57	39.11	0.05	*
Education today	5.85	5.86	0.99	
Educational level before treatment/cut-off	5.84	5.79	0.82	
Gender (1=male)	0.59	1.00	0.00	***
Ethnicity (1=Macedonian)	0.82	0.79	0.67	
Ethnicity (1=Albanian)	0.15	0.11	0.56	
Marital status (1=married)	0.77	0.39	0.00	***
Geography (1=urban)	0.78	0.61	0.10	
Geography (1=suburban)	0.10	0.18	0.32	
Geography (1=rural)	0.12	0.14	0.71	
Children (1=person has at least one child)	0.81	0.46	0.00	***
Number of children	1.68	0.93	0.00	***
Partner's employment status (1=if partner is employed)	0.45	0.14	0.00	***
Number of household members	4.09	3.93	0.61	
Number of household members below age 15	0.82	0.82	0.99	
Number of employed household members (age 15-64)	1.04	1.00	0.84	
Number of unemployed household members (age 15-64)	1.68	1.57	0.62	
Number of retired household members	0.35	0.46	0.47	
House ownership (1=owned)	0.84	0.79	0.54	
House size (sq. meters)	83.66	76.82	0.24	
Work experience (1=has work experience)	0.74	0.50	0.03	**
Work experience (number of months)	112.15	58.86	0.01	***
Contract (1=has written contract)	0.49	0.25	0.01	**
Salary before treatment/cut-off (MKD)	8,406.15	5,000.00	0.08	*
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	2.23	2.36	0.30	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.56	2.30	0.14	
Subjective employment status (1=employed)	0.14	0.21	0.38	
Employment status (1=employed)	0.12	0.21	0.29	

Unemployed at any time after program's end/cut-off (1=unemployed)	0.71	0.64	0.52	
Employment status (1=inactive)	0.32	0.11	0.00	***
Employed at program's end, but unemployed now	0.15	0.11	0.49	
Wage per hour (MKD)	7.83	8.85	0.81	
Hotelling T-squared test H0: The vector of means of all variables are equal across groups				
p-value				0.0001
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.				

Table 4-29 suggests that non-Albanian, non-married, and persons without previous working experience had higher chance to be selected into the program. Based on these estimates, we conduct a matching and evaluate the program.

Table 4-29 Matching – Training for deficient occupations program

Variable	Marginal effects	p-value	sig
Age	0.03	0.30	
Age squared	(0.00)	0.25	
Education today	0.02	0.69	
Macedonian	(0.21)	0.25	
Albanian	(0.30)	0.03	**
Marital status (1=married)	(0.25)	0.05	**
Geography (1=urban)	(0.13)	0.30	
Geography (1=suburban)	(0.07)	0.66	
Number of children	(0.02)	0.77	
Number of household members	0.05	0.13	
Work experience (1=has work experience)	(0.26)	0.04	**
# Observations	105		
Log-pseudolikelihood	-49.891637		
Pseudo R ²	0.1806		
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.			

Table 4-30 presents the results. They suggest that the program exerted no effect on the treated, once they are paired with non-treated on observables. Note that some of these results may be driven by the small sample size, so that the efficiency of the estimates is considerably reduced, leading to large standard errors which render estimates insignificant. However, there are only two results which deserves attention: i) persons trained for deficient occupations thought that their chances for finding a job after the training were better than compared to non-participants; and ii) the subjective opinion of the financial situation after program end/cut-off was better for the non-participants than for participants.

Table 4-30 Effects – Training for deficient occupations program															
Outcome variable	Unmatched				One-to-one matching with replacement				Matching with caliper (0.01)						
	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig
Financial condition after treatment/cut-off (1=better; 3=worse)	2.36	2.26	0.10	0.78		2.39	1.96	0.43	1.92	*	2.32	2.00	0.32	1.41	
<i>Bias reduction</i>						Mean	-23%	Median	-24%		Mean	-41%	Median	-49%	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.30	2.59	(0.29)	(1.53)		2.38	2.88	(0.50)	(1.97)	**					
<i>Bias reduction</i>						Mean	-30%	Median	-34%						
Subjective employment status (1=employed)	0.21	0.10	0.11	1.47		0.17	0.17	-	-		0.21	0.21	-	-	
<i>Bias reduction</i>						Mean	-47%	Median	-45%		Mean	-58%	Median	-73%	
Employment status (1=employed)	0.21	0.10	0.11	1.47		0.17	0.17	-	-		0.21	0.21	-	-	
<i>Bias reduction</i>						Mean	-47%	Median	-45%		Mean	-58%	Median	-73%	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.64	0.73	(0.08)	(0.83)		0.65	0.78	(0.13)	(0.78)		0.63	0.74	(0.11)	(0.59)	
<i>Bias reduction</i>						Mean	-38%	Median	-25%		Mean	-51%	Median	-50%	
Employment status (1=inactive)	0.11	0.39	(0.28)	(2.84)		0.13	0.09	0.04	0.35		0.16	0.11	0.05	0.38	
<i>Bias reduction</i>						Mean	-50%	Median	-45%		Mean	-59%	Median	-61%	
Employed at program's end, but unemployed now	0.11	0.16	(0.05)	(0.63)		0.13	0.13	-	-		0.16	0.16	-	-	
<i>Bias reduction</i>						Mean	-45%	Median	-29%		Mean	-56%	Median	-65%	
Wage per hour (MKD)	8.85	6.55	2.29	0.46		7.21	16.55	(9.34)	(0.75)		8.73	20.04	(11.30)	(0.86)	
<i>Bias reduction</i>						Mean	-32%	Median	-24%		Mean	-41%	Median	-49%	

Source: Authors' calculations based on survey.

Note: *, **, and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

4.5.1 Summary of findings and recommendations

The impact evaluation of the training in deficient occupations showed that the program does not exert any significant effect on the participants, as compared to the non-participants. Though, the results have to be treated with caution due to the small sample size, so that the efficiency of the estimates is considerably reduced, leading to large standard errors which render estimates insignificant. The only positive effect of the program is the belief that the employment chances have improved (with no actual improvement).

In addition, data show relatively low outcomes for the program participants. The employment after program end was only 11%, whereas current employment is at 21%. This implies that the program is not effective in enhancing the skills of the participants in the demanded occupations/professions, and/or that the choice of the sectors in excess demand of specific workers is incorrect. If the former is true, then ESA could improve the monitoring of the training conducted in the training institutions, should require certification of the training, could ask the VET Centre to re-define the training curricula, etc. If latter holds, then ESA should improve the methodology behind its skills demand survey, or use other instruments to collect more reliable information on skills in insufficient supply. Another practical problem in implementation of the program is the geographical distance between areas with excess demand for workers, training providers and place of residence of potential trainees. To overcome this problem, ESA could select trainees only from the locations where there is an excess demand (as the mobility within the country is very low), or carefully chose occupations for which it will provide the training, based on availability of training providers and of unemployed persons in the near neighbourhood. The program may be supplemented with some financial remuneration for the daily mobility (to training).

In the current setting, we propose that the program is not implemented anymore.

4.6 Training in advanced IT skills

Table 4-31 gives the individual and joint tests for the statistical significance of the mean difference between the treatment and control group for the training for advanced IT skills program. They suggest that the treated persons are on average older, less educated, more frequently males, less frequently own a house and more frequently possess a written contract, than compared their non-treated counterpart. Despite few differences are significant, overall, the Hotelling tests suggests that the two samples are different on observables. Hence, we continue with matching.

Table 4-31 Means' comparison – Training for advanced IT skills program

Socio-demographic characteristics	mean control	mean treated	p-value	sig
Age (years)	32.68	36.37	0.05	**
Education today	7.66	7.46	0.35	
Educational level before treatment/cut-off	7.59	7.23	0.09	*
Gender (1=male)	0.49	0.74	0.01	**
Ethnicity (1=Macedonian)	0.95	0.89	0.31	
Ethnicity (1=Albanian)	0.03	0.06	0.62	
Marital status (1=married)	0.34	0.46	0.27	
Geography (1=urban)	0.86	0.89	0.76	
Geography (1=suburban)	0.07	-	0.04	**
Geography (1=rural)	0.05	0.11	0.31	
Children (1=person has at least one child)	0.37	0.49	0.29	
Number of children	0.53	0.69	0.35	
Partner's employment status (1=if partner is employed)	0.31	0.37	0.52	
Number of household members	3.28	3.69	0.17	
Number of household members below age 15	0.53	0.54	0.92	
Number of employed household members (age 15-64)	1.36	1.37	0.95	
Number of unemployed household members (age 15-64)	0.93	1.14	0.40	
Number of retired household members	0.44	0.43	0.94	
House ownership (1=owned)	0.88	0.66	0.02	**
House size (sq. meters)	73.43	74.14	0.91	
Work experience (1=has work experience)	0.46	0.57	0.29	
Work experience (number of months)	27.78	53.37	0.16	
Contract (1=has written contract)	0.25	0.49	0.03	**
Salary before treatment/cut-off (MKD)	3,211.86	3,842.86	0.65	
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	2.12	1.85	0.04	**
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.29	2.47	0.37	
Subjective employment status (1=employed)	0.49	0.60	0.31	
Employment status (1=employed)	0.53	0.60	0.49	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.25	0.14	0.18	
Employment status (1=inactive)	0.02	0.06	0.36	
Employed at program's end, but unemployed now	0.15	0.23	0.38	
Wage per hour (MKD)	30.91	23.59	0.51	
Hotelling T-squared test H0: The vector of means of all variables are equal across groups				
p-value	0.0374			
Source: Authors' calculations based on survey. Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively				

Table 4-32 gives the probit results which suggest that ethnic Albanians are less frequently selected into the program, opposed to individuals originating from larger households who have higher chances to be selected. The propensity score we obtain from here is next used for obtaining the effect of the program.

Table 4-32 Matching – Training for advanced IT skills program

Variable	Marginal effects	p-value	sig
Age	0.04	0.48	
Age squared	(0.00)	0.63	
Education today	(0.04)	0.64	
Macedonian	0.32	0.01	
Albanian	(0.72)	-	***
Marital status (1=married)	(0.52)	.	
Geography (1=urban)	0.07	0.70	
Geography (1=suburban)	(0.09)	0.70	
Number of children	(0.04)	0.75	
Number of household members	0.09	0.10	*
Work experience (1=has work experience)	0.17	0.18	
# Observations	41		
Log-pseudolikelihood	-19.585037		
Pseudo R ²	0.296		

Source: Authors' calculations based on survey.
Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

Table 4-33 presents the results. Both matched and non-matched cases robustly suggest that the program has only one significant and positive effect related to the subjective well-being: participants assessed their financial conditions to have improved after the program. However, no real gains occurred in terms of employment prospects and wage gains. There are two caveats in the interpretation of these results: i) we operate with a small sample of 35 treated individuals, out of the sample frame of 81 persons who were involved in this training in 2010. The small sample size negatively affects the efficiency of our estimates; and ii) while large share of those who were treated but did not respond the questionnaire were unreachable (hence, in general terms, not imposing any correlation between the error term and the outcome), it is still likely that such correlation may exist. Namely, since the program consisted of training for advanced IT skills (leading to internationally recognized certificate), which are in demanded abroad, it may be that major part of the unreachable treated persons in our sample migrated by having accepted a job abroad. We may assume that there is positive selection into migration, i.e. those which are more capable and with brighter employment opportunities will migrate first. Hence, we may actually work with the remaining part of the sample which has systematically lower-performance indicators than compared to those who emigrated. Hence, the largely no-result of this program should be also interpreted in this light. Though, we have no strong evidence to support this view.

Table 4-33 Effects – Training in advanced IT skills program

Outcome variable	Unmatched				One-to-one matching with replacement				Matching with caliper (0.01)						
	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig
Financial condition after treatment/cut-off (1=better; 3=worse)	1.85	2.13	(0.28)	(2.28)	**	1.85	2.22	(0.37)	(2.36)	**	1.73	2.18	(0.28)	(2.28)	**
<i>Bias reduction</i>						Mean	-45%	Median	-52%		Mean	-29%	Median	-34%	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.47	2.27	0.20	0.99		2.56	2.22	0.33	1.10		2.50	2.25	0.25	0.71	
<i>Bias reduction</i>						Mean	-13%	Median	22%		Mean	-5%	Median	13%	
Subjective employment status (1=employed)	0.60	0.47	0.13	1.18		0.57	0.43	0.14	0.86		0.59	0.45	0.14	0.80	
<i>Bias reduction</i>						Mean	-51%	Median	-46%		Mean	-44%	Median	-30%	
Employment status (1=employed)	0.60	0.53	0.07	0.66		0.57	0.50	0.07	0.43		0.59	0.50	0.09	0.53	
<i>Bias reduction</i>						Mean	-53%	Median	-40%		Mean	-45%	Median	-22%	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.14	0.26	(0.12)	(1.35)		0.14	0.29	(0.14)	(0.97)		0.18	0.32	(0.14)	(0.91)	
<i>Bias reduction</i>						Mean	-49%	Median	-48%		Mean	-43%	Median	-33%	
Employment status (1=inactive)	0.06	0.02	0.04	0.96		0.07	-	0.07	1.44		0.05	-	0.05	1.00	
<i>Bias reduction</i>						Mean	-47%	Median	-40%		Mean	-44%	Median	-22%	
Employed at program's end, but unemployed now	0.23	0.13	0.10	1.17		0.25	0.21	0.04	0.31		0.23	0.18	0.05	0.37	
<i>Bias reduction</i>						Mean	-56%	Median	-47%		Mean	-48%	Median	-35%	
Wage per hour (MKD)	23.59	34.41	(10.82)	(0.87)		22.38	12.78	9.60	0.76		23.32	16.27	7.05	0.52	
<i>Bias reduction</i>						Mean	-53%	Median	-40%		Mean	-47%	Median	-29%	

Source: Authors' calculations based on survey.

Note: *, **, and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.

4.6.1 Summary of findings and recommendations

Training in advanced IT skills does not bring higher effects for the participants compared to the non-participants in terms of their employment probability. The only significant and positive effect is related to the subjective well-being: participants assessed their financial conditions to have improved after the program. However, no real gains occurred in terms of employment prospects and wage gains. This implies that the program is very costly for the society, as there are no benefits attached to it both from societal and individual viewpoint. In addition, as data show that 60% of the participants are employed today, the program is also likely to have significant dead-weight loss.

There are two caveats in the interpretation of these results: i) we operate with a small sample of 35 treated individuals (out of 81 participants) which negatively affects the efficiency of our estimates; and ii) while large share of those who were treated but did not respond the questionnaire were unreachable (hence, in general terms, not imposing any econometric problems in the estimation), it is still likely that such correlation may exist. Namely, we assume that some of the program participants (i.e. those more capable and with better labour market characteristics) were likely to migrate which, if true, biases our results. Though, we have no strong evidence to support this view.

This suggests that either the program design should be changed or monitoring of former participants to the program improved. The ESA might closely follow the participants through years, potentially imposing a clause that they cannot leave the country for a certain period after the training took place, and hence have better information on the effectiveness of the program.

5 Conclusion

This report presents the findings of the impact evaluation carried out on a selection of active labour market programs implemented by the Employment Service Agency (ESA) of Macedonia during the period 2008-2012. We evaluate the following programs: program for self-employment (implemented in 2008), internship (implemented in 2010 and 2012), training for known employer (implemented in 2010 and 2012), wage subsidy program (2010), training for advanced IT skills (2010) and training for deficient occupations (2010). The choice of the years was based on the intention to assess not just the short-term (immediate) effectiveness of the program, but also the longer-term effects. Moreover, for two of the programs, we evaluate two years of program implementation, as to see if some changes made to the program brought better effects.

The key research question was whether participation in the active labour market programs increased the probability of participants to find and retain gainful employment. However, in addition to this main outcome, we include other outcome variables such as: inactivity, unemployment, wages, changes in the prospects for employment after the program, changes in financial status, subjective labour market status, etc.

To answer such questions (i.e. to assess the program outcomes) we employ a quasi-experimental impact evaluation method. This is the first time the impact evaluation is implemented in Macedonia on such a large scale, and by national experts. The method, in essence, measures and compares the differences in labour market outcomes between those who participated in the programs (*treatment group*) and those who did not (*control group*). In this type of analysis, the precision of the estimate of program effects critically depends on whether the control group is the “same” or largely comparable to the treatment group; in other words, whether we can “compare the comparable” (Heckman et al. 1999). This means that program participants need to be compared only to those non-participants who could have participated in the program (i.e. had an equal chance to be selected for participation as those who were actually treated). In this regard, we have carefully chosen the control groups for each program and then used econometric techniques to match participants with “same” non-participants.

Data for the evaluation were gathered through a one-to-one survey that was run during October 2014, covering the participants (treatment groups) and non-participants (control groups). The survey was implemented by the Institute for Political Research-Skopje (IPIS). In total, we have data for 2,485 program participants from ESA for the six programs we evaluate. The response rates were relatively low for the participants in the following programs: training for known employer 2010 (44%) and for training in advanced IT skills (51%).

The main findings from the evaluation show that results are mixed. Some programs bring comparatively better labour-market outcomes for the program participants relative to the non-participants. i.e. that programs are effective. This holds true both for the actual outcomes and subjective ones. However, there are also programs which prove not to be effective in improving the labour market outcomes of the participants.

Table 5-1 Employment outcomes of each program – summary table

Program	Actual outcome variables				Subjective well-being			
	Current employment	Unemployed at any time	Employment at program end/cut-off, but not today	Wages	Inactivity	Subjective employment	Better financial situation	Better chances to find a job
Internship 2010	√ 25 p.p. higher chances of participants to be currently employed	√ 21 p.p. lower probability of participants to be unemployed at any time after program end/cut-off date				√ 23 p.p. higher employment rate of participants based on self-assessment		
Internship 2012	√ 29 p.p. higher probability of participants to be employed currently		√ (negative) 17 p.p. lower probability of participants to be employed at program end/cut-off	√ More than double higher hourly wages of participants	√ Very small difference (0.08 p.p.)	√ 31 p.p. higher employment rate of participants based on self-assessment		
Training for known employer 2010	√ 42 p.p. higher probability of participants to be employed currently	√ 30 p.p. lower probability of participants to be unemployed at any time after program end/cut-off date		√ Very large gain in wages for participants (more than three times higher wages)		√ 35 p.p. higher employment of participants based on self-assessment		√ Self-reported better chances for employment after program end/cut-off
Training for known employer 2012	√ 34 p.p. higher probability of participants to be employed currently	√ 26 p.p. lower probability of participants to be unemployed at any time after program end/cut-off date		√ Very large gain in wages for participants	√ 28 p.p. lower chances of participants to be inactive today	√ 47 p.p. higher employment of participants based on self-assessment		√ Reported better chances for employment after program end

Continue from page 67 **Table 5-1 Employment outcomes of each program – summary table**

Program	Actual outcome variables				Subjective well-being			
	Current employment	Unemployed at any time	Employment at program end/cut-off, but not today	Wages	Inactivity	Subjective employment	Better financial situation	Better chances to find a job
Self-employment			√ 29 p.p. lower chances for participants to be employed at program end, but unemployed now		√ 30 p.p. lower chances to be inactive today	√ 29 p.p. higher for participants		√ Reported better chances for employment after program end
Training in advanced IT skills							√ Reported higher improvement in financial situation by participants	
Wage subsidies			√ (negative) Lower chances to be employed at program end/cut-off but unemployed today					
Training for deficient occupations							√ (negative) Reported less improvement in the financial situation after program end/cut-off	√ Reported better chances for employment after program end

Based on the employment outcomes of each program, we provide following categorization of programs in terms of their effectiveness:

- 1. Programs that are effective and should be further implemented in the current design (in grey in Table 5-1)** – This holds for the internship program and training for known employer. For these two programs we have evaluated two years of implementation, as there were slight changes to the program design. We found that there are no significant differences in the effectiveness between years for both programs implying that there is **no need for making the criteria more or less stringent**. These two programs are the only ones that bring real differences for the participants: i) higher employment today (longer-term effect); ii) that reduced the changes for non-employment at any time after program end, and iii) to bring financial gains to the participants (in terms of wages). We still give some recommendations for further improvement of the programs (see Sections 4.2.4 and 4.3.4, for more details)

Internship program:

- As the evaluation shows that both programs (2010 and 2012) are effective, the Ministry may lift the obligation for employers to hire 50% of the interns to make the program more attractive for employers;
- We suggest that the government aligns further this program with similar programs (and terminology) of the EU countries;
- In line with the experience of EU countries, the length of the internship could be made more flexible and of longer duration, ranging from 3 to 9 months;
- Enhance the learning experience and gained knowledge and skills by interns;
- Provide job readiness training before placing young people on internship.

Training for known employer:

- Priority in this program should be given to employers from sectors which are set by the government as priority ones for the future development of the country, or those that are most propulsive sectors and have high growth potential;
- An external assessment system can be set up to assess the effectiveness of the training.

- 2. Programs that bring some positive effects, but the design need to be improved (in light blue in Table 5-1)** – The self-employment program and training in advanced IT skills fall into this category. We found that the former is effective in improving employability at program end (but not afterwards), in reducing the inactivity of the participants, and bringing positive changes in the subjective well-being. But, we do not find evidence that the program brought effects in terms of employment after the program ended, which is the main outcome variable we are interested in.

As we explain in Section 4.1.1, the **self-employment program** has been changed since 2008 and made more stringent for the participants, which is likely to prolong for some time the exit of the established businesses from the market, but that will not make any real change in terms of the long-term effect of the program. We therefore recommend changing the eligibility rules and selection procedure, rather than the penalties for the non-compliers. In order to improve the effectiveness of the self-employment program, we propose following (see also 4.1.1 for more details):

- Applicants in the program should not be self-selected, but rather randomly selected out of the pool of eligible candidates. Then, they should enter the filters of the program (training, motivation, developing a business plan, funding).
- The program could contain a component related to business management and technical skills;
- The quality of the training component should be monitored;
- The size of the program (i.e. number of program participants) could be reduced whereas the amount of the grant increased;

- The training could be designed so that it does not necessarily leads to a grant but may help participants to gain access to commercial credits. This should be accompanied by establishment of a Guarantee fund;
- The amount of the grant given to applicants can vary according to the type of business and required technology for establishing and running the business,
- The government might set priority sectors/industries for which grants will be given;

The **training in advanced IT skills** is found only to improve subjective well-being, but not the actual labour market outcomes of the participants. However, as we explain in Section 4.6.1, there were two caveats in the interpretation of these results: i) we operate with a small sample of 35 treated individuals (out of 81 participants) which negatively affects the efficiency of our estimates; and ii) given the large share of non-reachable program participants at the interview date, we assume that some of the program participants (i.e. those more capable and with better labour market characteristics) were likely to migrate which, if true, biases our results. Though, we have no strong evidence to support this view. Because of these two caveats (especially the migration-related one) we recommend that the program continues, but with following changes:

- changes in eligibility and selection (towards unemployed with shorter spell of unemployment, with better initial IT skills and knowledge);
- imposing some requirements to the program participants, such as staying in the country for some minimum period of time, obligation for regular update of contact information 5 years after program end, etc.;
- program design should be changed or monitoring of former participants to the program improved. The ESA should closely follow the participants through years, and develop a longitudinal database.

Otherwise, the program is costly for the society.

3. Programs that call for major revisions or discontinuation (in blue in Table 5-1) – Two of the programs, wage subsidies and training for deficient occupations are not bringing positive effects to the participants (except in one measure of subjective well-being), and are even worsening the position of the participants. Moreover, the participants in the program have very low employment rates today.

For the **wage subsidy program**, we found that the employment after program end (the subsidy period plus the period with the obligation to keep the subsidised workers on the job) is very low, suggesting that employers are using this program to address only short term needs of the workers. We provide following suggestions for improving the effectiveness of the wage subsidies program:

- Introduce training to unemployed prior to program assignment;
- Enhance certification program for training offered by employers to increase the long-term value of the training and the experience gained while in program.

For the **training in deficient occupations**, the employment after program end was only 11%, whereas current employment is at 21%. This implies that the program is not effective in enhancing the skills of the participants in the demanded occupations/professions, and/or that the choice of the sectors in excess demand of specific workers is incorrect. If the program is to be continued, we recommend the following:

- ESA should improve the monitoring of the quality of training conducted in the training institutions and should require certification of the training;
- ESA needs to improve the methodology behind its skills demand survey, or use other instruments to collect more reliable information on skills in insufficient supply;
- ESA could select trainees only from the locations where there is an excess demand (as the mobility within the country is very low), or carefully chose occupations for which it will provide the training, based on availability of training providers and of unemployed persons in the near neighbourhood (and not solely on the skills in deficient supply).

Besides the specific findings and recommendations presented for each program, we also derive some more general conclusions and suggestions for improving the ALMPs in Macedonia:

- Although each program design sets eligibility criteria which are met by the program participants, self-reported information collected by the participants (as opposed to the administrative data) show that the **targeting of some programs is not in line with the intended one**. In particular, data collected for this study showed: i) large share of program participants stated to have very short unemployment spells prior to program participation (less than 1 month); ii) there is a high share of long-term unemployed among participants in training in advanced IT skills. These differences in planned and actual targeting may partly be explained by the engagement of unemployed in informal employment prior to program participation. Even in that case, it deserves some attention from the policymakers as it suggests that programs are involving “easy-to-place” individuals rather than those in real need of support from the ESA.
- There is a quite **large share of persons who could not be reached** by provided mobile phones, neither through their home addresses. It is surprising is that these shares are larger for the treatment than for the control group, given that one expects that the ESA keeps more reliable information for the program participants than for the other registered unemployed persons. In our case, this especially holds for the training in advanced IT skills where most participants were unreachable even after many attempts were made. If ESA intends to better monitor the effects of its policies (through time), it needs to improve its database, collect information on the program participants for longer period of time (and not just at the program end) and consequently implement longitudinal studies. This is also important given the costs relatively large associated with impact evaluation studies. Correct contact information about ALMPs participants is also essential for future impact evaluation studies.
- We also propose ESA to **implement in future a randomized experiment method for program evaluation**. It can choose a program in which individuals eligible for participation will be randomly assigned to the treatment and control group.

References

- Betcherman, G., Olivas, K. and Dar, A. (2004) *Impacts of Active Labor Market Programs: New Evidence from Evaluations with Particular Attention to Developing and Transition Countries*. World Bank Social Protection Discussion Papers Series, No. 0402.
- Blunder, R., Dearden, L. and Sianesi, B. (2005) Evaluating the Effects of Education: Models, Methods and Results from the National Child Development Survey, *Journal of the Royal Statistical Society, Series A*, 168, 473-512.
- Corbanese, V. (2012) *Performance monitoring of active labour market programs implemented in 2007-2010: Key findings*. Geneva: ILO. Mimeo.
- European Commission (2013) *Apprenticeship and Traineeship Schemes in EU27: Key Success Factors. A Guidebook for Policy Planners and Practitioners*. Brussels: European Commission.
- Heckman, J., LaLonde, R. and Smith J. (1999) The Economics and Econometrics of Active Labour Market Policy, in Ashenfelter, O. and Card, D. (eds.), *The Handbook of Labour Economics*, Volume III, Amsterdam: Elsevier Science.
- Imbens, G. and Wooldridge, J. (2009) Recent Developments in the Econometrics of Program Evaluation), *Journal of Economic Literature*, 47, 5-86.
- O'Higgins, N. and Kirevska, S. (2012) Independent Evaluation of UNDP program, "Self-employment II-V and Technical Assistance to Support the Government Operational Plan for Active Labour Market Measures (ALMMs)". Mimeo.
- Petreski, M., Mojsoska-Blazevski, N. and Petreski, B. (forthcoming) Gender wage gap when women are highly inactive: Evidence from repeated imputations with Macedonian data. *Journal of Labour Research*, 35(4), p.393-411.
- Rosenbaum P. R. and Rubin D.B. (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score, *The American Statistician*, 1985, Vol.39, No1).
- Rubin, D. (1974), Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies), *Journal of Educational Psychology*, 66, 688-701.
- Sianesi, B. (2004): An Evaluation of the Active Labour Market Programs in Sweden, *The Review of Economics and Statistics*, 86(1), 133-155.
- Smith, H. (1997) Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies, *Sociological Methodology*, 27, 325-353.
- Smith, J., and P. Todd (2005) Does Matching Overcome LaLonde's Critique of Non-experimental Estimators?, *Journal of Econometrics*, 125(1-2), 305 - 353.

Annex 1

Table A1.1: Explanatory variables included in the preferred specification of the regression model

Name of variable	Survey question	Description
Sex	What is your sex?	1: Female 2: Male
Age	What is your exact age?	Age (in years)
Age squared		Age (in years) squared
Married	What is your marital status?	1: If married 0: Otherwise
# Employment of a partner	What is the employment status of your partner?	1: If employed 0: Otherwise
# Members of household	Number of members of household?	Number: 1-18
# Members of household able unemployed	Number of members of household who are unemployed and able to work?	Number: 0-10
# Children in the family	Number of children in family?	Number: 0-9
# Retired household members	Number of household members over 64 years?	Number: 0-3
Size of the apartment	Size of the apartment? (in sq meters)?	Number: 10-500
House ownership status	What is your house ownership status?	1: Ownership, without credit/mortgage 2: Ownership, with credit/mortgage 3: Rental agreement 4: Non-paying rental agent
Education: less than primary school	What is your highest educational level?	1: If without education, up to 4 years of primary school, 5 to 7 years of primary school 0: Otherwise
Education: primary		1: If primary school 0: Otherwise
Education: vocational		1: If vocational (3 years), 0: Otherwise
Education: secondary		1: If secondary special school (4 years), 0: Otherwise
Place of living	What is your place of living?	1: Urban 0: Otherwise
# Months of work experience	How many months of work experience?	Number: 0-182
# Years of work experience on the previous job	How many years of work experience on the job which precedes the current one?	Number: 0-45
Economy sector of previous job	What was industry sector of previous job?	1: Agriculture 2: Manufacturing 3: Services
Salary on previous job	Your salary on previous job before 2011?	Number, salary in MKD

