Ministry of Housing,
Communities &
Local Government

Evaluation of the STEP Ukraine Programme

Process, Impact, and Economic Evaluations

Ministry of Housing, Communities and Local Government

This document/publication is also available on our website at www.gov.uk/dluhc

If you have any enquiries regarding this document/publication, complete the form at http://forms.communities.gov.uk/ or write to us at:

Department for Levelling Up, Housing and Communities
Fry Building
2 Marsham Street
London
SW1P 4DF
Telephone: 030 3444 0000

For all our latest news and updates follow us on Twitter: https://twitter.com/luhc

April 2026

Foreword

The UK's response to Russia's invasion of Ukraine has provided safety and stability to those forced to flee their homes. Supporting people to rebuild their lives in the UK is central to this effort. This includes enabling individuals to improve their English language skills and access sustainable employment, which are key foundations for successful integration.

The Ministry of Housing, Communities and Local Government launched the STEP Ukraine programme to meet this need, offering a combined package of English language provision and tailored employment support to Ukrainian visa holders across the UK. This report presents the findings of an independent evaluation of the programme, examining delivery, outcomes, and value for money.

The evaluation provides strong evidence that the programme has delivered positive impacts. Participants made measurable improvements in their English language proficiency and were more likely to move into employment compared to similar individuals who did not receive support. The findings also indicate that the programme represents good value for money, generating benefits for both participants and the wider economy.

Importantly, the report highlights lessons for future policy and delivery. It demonstrates the value of integrating language and employment support, the importance of flexibility in meeting participants' needs, and the role of tailored provision in supporting people with diverse experiences and circumstances.

I would like to thank World Jewish Relief and the British Council, who designed and delivered the programme; Alma Economics and RAND Europe for their rigorous evaluation; members of the steering group for their expert input and challenge; and the participants who shared their experiences. Their contributions have been essential in building this evidence base.

As the government continues to support those displaced by conflict, the insights from this evaluation will help inform the design of future programmes and ensure that support is both effective and evidence-based.

Stephen Aldridge
Director for Analysis and Data & Chief Economist
Ministry of Housing, Communities and Local Government

Contents

List of Tables

List of Figures

List of Boxes

Glossary of Key Definitions

Term	Definition
Ashenfelter's dip	A small deterioration in the employment rate of the treated group relative to the control just before the intervention.
Attrition bias	A type of bias that occurs when participants drop out of a study or programme in ways that are not random (e.g., less motivated learners leaving). This can distort results if the remaining sample is systematically different from those who left.
Barrier	A factor or condition that facilitates effective programme delivery or positive outcomes for participants.
Before-after (pre-post) analysis	A method for evaluating programme impact by comparing outcomes measured before participants begin an intervention with outcomes measured afterwards, without using a separate control group.
Beneficiary	An individual who receives direct support or services from a programme, in this case, participants in the STEP Ukraine programme.
Control group	Individuals who were eligible for the intervention being studied but did not participate – in this case, eligible Ukrainians who did not enrol in the STEP Ukraine programme.
Control variable ("control")	A variable to indicate a relevant factor, which may influence the relationship being estimated (e.g., age, gender, time spent in the UK), included in a statistical model to reduce bias and isolate the effect of the programme itself from these other factors.
Counterfactual	An estimate of what the outcomes of participants (e.g., English ability) would have looked like in the absence of the programme. This is crucial for estimating the causal impact of an intervention.
Difference-in-differences	A statistical method for estimating causal effects by comparing outcome changes for a treatment group against a comparison group. A pseudo-difference-in-differences approach was used for the language impact evaluation, since no true control group was available. This involved deriving a counterfactual estimate from the relationship between test scores at the point of entering the programme and time spent in the UK.
Descriptive statistics	Basic summaries of data, such as averages, percentages, and distributions, providing an overview without deeper statistical modelling.

Term	Definition
Dummy variable	A variable that takes the value of 0 or 1 to represent categorical information in regression models (e.g., gender: male = 1, female = 0).
Employment probability	The probability that someone is employed, as opposed to not being employed.
Employment sustainability	Being in sustained employment over specified durations (e.g., employed for at least 3, 6, or 12 consecutive months).
Enabler	A factor or condition that facilitates effective programme delivery or positive outcomes for participants.
Impact evaluation	A systematic study assessing the changes in outcomes (e.g., English proficiency, employment) that can be attributed to a programme, as distinct from changes due to other factors.
Intrinsic motivation	A person's internal drive to achieve goals (e.g., learning English), independent of programme participation.
Line of best fit	A straight line drawn on a scatter plot of data to represent the general trend of the data points.
Logic model	A visual or narrative tool used to describe what impact is expected from a programme or policy intervention. It outlines the sequence of intended outcomes, showing how inputs and activities are expected to lead to outputs and results.
Mixed-methods approach	A research design combining qualitative and quantitative data collection and analysis to provide a more comprehensive understanding of programme processes and outcomes.
Model for Adaptation Design and Impact (MADI)	A framework used to categorise and assess adaptations made during programme delivery. It evaluates the nature of adaptations (e.g., proactive versus reactive, systematic versus unsystematic) and their potential impact on outcomes.
Multivariate regression analysis	Regression analysis involving multiple explanatory variables (see entry for 'regression analysis' below), allowing researchers to estimate the effect of one factor while controlling for others.
Natural progression	The improvement in English skills expected simply from living in the UK, even without formal training.
Output	Direct products or services delivered by a programme (e.g., number of workshops held, participants trained).
Outcome variable	The main result being measured in an evaluation (e.g., English proficiency).

Term	Definition
Process evaluation	A study focused on understanding how a programme is designed and implemented, including delivery mechanisms, fidelity, adaptations, and contextual factors affecting outcomes.
Process mapping	A method used to visually and descriptively outline the steps, decision points, responsibilities, and flows within a programme. It helps identify how activities are implemented, where adaptations occur, and potential areas for improvement.
Propensity score matching	A statistical technique used to estimate the causal effect of a treatment by matching comparable members of a treated group to one or more members of an untreated (control) group, allowing for a more balanced comparison between the two groups.
Regression analysis	A statistical technique for estimating the relationship between an outcome ("dependent") variable (e.g., English score) and one or more determining or "explanatory" variables (e.g., months in the UK, gender). Common usage is to say the dependent variable is "regressed" on the explanatory variables. We then look at the estimated "coefficient" on each explanatory variable to determine its influence on the dependent variable.
Robustness check	An additional analysis conducted to verify that the results are stable and not overly sensitive to specific assumptions or methods.
Sequential explanatory design	A mixed-methods evaluation approach in which quantitative data are collected and analysed first, followed by qualitative data to help explain and interpret the quantitative findings.
Selection bias	A bias introduced when individuals self-select into a programme, meaning they may differ systematically from non-participants (e.g., motivated learners versus less-engaged individuals).
Standard deviation	A statistical measure of how spread-out data points are from the average. Changes expressed in standard deviations show the size of improvements relative to the variability in scores.
Standard error	A measure of how precise an estimate is; smaller standard errors mean greater statistical confidence in the results.
Statistical significance	Statistical significance refers to the extent to which the result is unlikely to be observed due to random chance if there were truly no real effect. The highest conventional level of statistical significance, where we can be most confident that the result is not due to chance, is 1%.
t-test	A t-test checks whether the difference between two averages is statistically significant, i.e., larger than is likely to have been observed due to random chance.

Term	Definition
Thematic analysis	A qualitative analysis method used to identify, analyse, and report patterns or themes within textual data. It helps summarise key insights, capture recurring issues, and interpret different perspectives.
Theory of Change	A framework that explains why and how a programme or policy is expected to achieve its intended outcomes. It provides the story behind the causal links in the logic model, outlining assumptions, contextual factors, and mechanisms of change that connect activities to results.
Treatment group	Participants who have received the intervention being studied – in this case, individuals enrolled in the STEP Ukraine programme.
Treatment effect	The impact of a programme (or "treatment") estimated specifically for those who actually take part in it. This is more specifically known as the average treatment effect on the treated (ATT).
Triangulation	The use of multiple data sources, methods, or perspectives to cross-check findings and strengthen the reliability and validity of evaluation conclusions.

List of Acronyms and Abbreviations

Acronym	Full term
AI	Artificial Intelligence
ATT	Average treatment effect on the treated
BC	British Council
BCR	Benefit-cost ratio
BITC	Business in the Community
CEFR	Common European Framework of Reference for Languages
CPI	Consumer Price Index
CPIH	Consumer Price Index including owner-occupiers' housing costs
CRM	Customer Relationship Management
CSDID	Callaway and Sant'Anna Difference-in-Differences framework
CV	Curriculum Vitae
DiD	Difference-in-Differences
DoB	Date of birth
DWP	Department for Work and Pensions
EAs	Employment Advisers
ESOL	English for Speakers of Other Languages
ETs	English tutors
HMRC	His Majesty's Revenue and Customs
IELTS	International English Language Testing System
KPI	Key Performance Indicator
MADI	Model for Adaptation Design and Impact
MHCLG	Ministry of Housing, Communities and Local Government
MWS	Migrant Worker Scan
NINOs	National Insurance Numbers
OBR	Office for Budget Responsibility

Acronym	Full term
ONS	Office for National Statistics
PAYE	Pay As You Earn
PSM	Propensity score matching
PTSD	Post-Traumatic Stress Disorder
QED	Quasi-experimental design
RIBA	Royal Institute of British Architects
SAMHSA	Substance Abuse and Mental Health Services Administration
SIM	Subscriber Identity Module
STEP	Skills, Training, Employment Programme
TERN	The Entrepreneurial Refugee Network
ToC	Theory of Change
UC	Universal Credit
UK	United Kingdom
UK ENIC	UK National Information Centre
VfM	Value for money
WJR	World Jewish Relief

Executive Summary

This executive summary sets out key findings, lessons, and recommendations from an evaluation of the STEP Ukraine Programme. These can be used to help inform policy and the future design of innovative integration support for refugees and forced migrants.

Programme overview

The STEP Ukraine programme was developed by the Ministry of Housing, Communities and Local Government (MHCLG) and delivered by World Jewish Relief (WJR) in partnership with the British Council. The pilot was developed in direct response to the displacement of Ukrainian nationals following the outbreak of the Russia-Ukraine war. The programme aimed to accelerate the integration of Ukrainian visa holders in the UK by enhancing English language proficiency and improving employment prospects. The core intervention consisted of 12 weeks of tailored employment support combined with ten weeks of intensive online English language training, all provided free of charge.

The programme initially ran from August 2023 to August 2024. The first stage of the programme engaged 17,282 participants, of whom 9,813 were onboarded. Following a contract extension, delivery was continued from August 2024 to March 2025. In total, the programme engaged 45,512 participants, with 14,850 signing up for the programme. This evaluation focuses on the initial programme from August 2023 to August 2024.

Evaluation objectives

Alma Economics and RAND Europe were commissioned by MHCLG to conduct an independent evaluation of STEP Ukraine to assess its processes, impact on language skills and employment outcomes, and Value for Money (VfM).

The evaluation aimed to:

1. Assess the design, delivery, and experience of the STEP Ukraine programme.
2. Quantitatively estimate the impact of the programme on English skills.
3. Quantitatively estimate the impact of the programme on employment outcomes.
4. Quantitatively estimate the programme's VfM.

To achieve these objectives, the evaluation employed a mixed-methods approach using both qualitative and quantitative methods.

Qualitative methods include:

- In-depth interviews and focus groups with Ukrainian guests, alongside interviews with programme managers, implementers, and stakeholders, to gather insights into the programme's implementation and participant experiences.

Quantitative methods include:

- Analysis of pre- and post-programme English proficiency test scores to measure the programme's impact on language skills.

- Analysis of participant employment outcomes compared to a counterfactual group of Ukrainian guests who did not participate in STEP Ukraine.

- Analysis of the VfM generated by the programme. This compares programme costs to the estimated value of the benefits that it generated for society and for the public finances.

Key findings

1. Process evaluation

The process evaluation revealed that the programme aimed to address both immediate integration needs and longer-term career aspirations through a number of routes. The key takeaways from the process evaluation are:

- The programme's dual-focus model was ambitious and well-conceived. In practice, however, the language and employment support strands often operated in silos, with a lack of communication between English tutors and Employment Advisers (EAs). In turn, this undermined the intended synergy of the model.

- Operational agility was demonstrated when the programme was launched, as within six weeks of the contract being awarded, the programme was operational. However, the speed of the launch may also have resulted in some of the practical difficulties subsequently experienced, including high caseloads (up to 35 clients per EA), rigid scheduling and timetabling of English, and employment support concurrently indicate that.

- Personalised employment support was seen as one of the strongest aspects of the STEP Ukraine programme. Participants consistently valued the culturally sensitive, one-to-one guidance from Ukrainian-speaking EAs, however, STEP Ukraine staff themselves felt that more training could be provided on mental health concerns.

- English language training appeared to boost the confidence of participants, however, there were some equity concerns related to the starting English level of participants. Findings indicate that learners who started with a lower level of English were particularly disadvantaged, and that the model likely worked best for those who already had a baseline level of English language on entry to the programme.

- Barriers to participation and completion of the STEP Ukraine programme were largely related to the flexibility of the programme. Barriers cited include concurrent timetabling of support, which hindered completion, digital access, which limited the ability to participate, and caring/work responsibilities, which prevented attending sessions during the daytime. This highlights a tension between the programme structure and participants' real-life constraints to participate.

2. English language impact evaluation

On average, participants in the STEP programme showed significant improvement in their English test scores between programme entry and exit, even when controlling for time spent in the UK:

- The increase in average English test scores between pre- and post-STEP Ukraine participation ranged from 20.8% (in the case of our more conservative measure) to 29.2% (in the case of our less conservative measure).

- Some of that increase in test scores would be expected because English tends to improve with time spent in the UK. However, STEP participants saw an additional increase in English test scores, over and above the improvement that would be expected as participants spend longer in the UK, of between 17.6% (using our more conservative measure) and 25.9% (using our less conservative measure). These estimated effects are statistically significant, meaning they are unlikely to have been observed by chance.

3. Employment impact evaluation

At the UK level, STEP Ukraine had a positive effect on the employment outcomes of participants:

- Participants were on average 6.6 percentage points more likely to be in employment over the 20 months following entry to the programme than if they had not participated. These impacts emerged shortly after participation and grew steadily up to approximately seven months after treatment.

- We also estimate the positive effects of STEP Ukraine on the probability of sustained employment. It increased the likelihood of being in employment for at least three consecutive months by 7.7 percentage points, and at least six consecutive months by 6.5 percentage points.

- We estimated that STEP Ukraine had a positive effect on the monthly earnings of participants who were already employed prior to the programme, raising their monthly earnings by around 10.7% after 20 months.

We also examined impacts at the level of the devolved administrations in Scotland, Wales, and Northern Ireland.

- The estimated impacts of STEP on employment are broadly the same in Scotland as in the UK as a whole. For example, the likelihood of being employed among STEP Ukraine participants in Scotland increased by 6.0 percentage points over the 20 months following the programme, compared to if they had not participated.

- It was unfortunately not possible to reach robust conclusions for Northern Ireland and Wales. This was due to lower sample sizes and more difficulty establishing robust control groups of non-participants (for more information, see Section 4.3 and Appendix 4).

4. Economic evaluation

- The cost-benefit analysis (CBA) of STEP Ukraine demonstrates that the programme delivered VfM, both from the perspective of the public finances and for society overall.

- The model compared the benefits of STEP Ukraine's impact on employment, such as increased tax revenues and reduced Universal Credit payments, against costs, including programme implementation costs and increased travel expenses.

- Our central scenario suggested that for every £1 invested in the programme, it generated gross savings of £2.04 (i.e., net savings of £1.04) for the public finances, via higher tax receipts and lower benefit payments arising from increased employment and earnings. The programme generated £3.67 in gross benefits (i.e., £2.67 in net benefits) for society overall, primarily from the value of increased economic output due to higher employment and earnings.

- Sensitivity analysis explored a range of plausible values around key assumptions required for the CBA. These result in benefit-cost ratios (BCRs) varying from 0.88 to 3.27 from a public finance perspective, and 1.59 to 5.84 from a social perspective.

Limitations of the research

The evaluation made use of multiple methods and drew on a variety of available data sources. However, there are several limitations to the data.

Process evaluation

- Time constraints and limited availability of programme partners restricted several planned data collection activities, including non-participating observations of programme activities and interviews with key English language training stakeholders.

- Partner facilitation of the survey questionnaire to delivery partners limited the evaluation team's oversight over its distribution. Coupled with low response rates, the findings do not reflect a representative picture of all STEP Ukraine implementers.

Language impact evaluation

- The counterfactual analysis relies on the assumption that earlier arrivals did not differ from later arrivals systematically on unobserved characteristics that affect the relationship between time in the UK and English language ability (e.g., unobserved ability, selective migration timing).

- To estimate the impact of STEP Ukraine on English language ability, we had to model the improvements in English language that we would have expected irrespective of the programme, due to increased time spent in the UK. This modelling exercise necessitated assumptions – in particular, that there is a linear relationship between time spent in the UK and English language ability.

- Around 20% of participants dropped out of the programme before taking an exit test, meaning that the results may be subject to attrition bias. For example, if participants drop out of the programme because it is not improving their English skills, then we would overstate the positive impact of the programme.

Employment impact evaluation

- The analysis window spans 20 months after the start of the programme. As a result, it is not possible to examine the longer-term impact of the programme.

- The analysis relies on the standard 'difference-in-differences' (DiD) assumption of parallel trends in outcomes between participants and non-participants. Our statistical tests generally demonstrate that the parallel trends assumption was, on average, satisfied during the months prior to the intervention. However, a 'stricter' version of the test generally shows that the parallel trends assumption did not hold in every individual month, as one would expect from a true randomised experiment.

- Unfortunately, due to the much smaller sample sizes, it was not possible to attain robust results for the impact of STEP Ukraine on employment outcomes in Wales and Northern Ireland.

Economic evaluation

- Some programme impacts could not be directly estimated from the impact evaluation, requiring assumptions (e.g., regarding impact persistence). Where possible, we applied standard assumptions from DWP guidance. Where uncertainty remained, informed assumptions were made and tested through sensitivity analysis.

- The model captures the main expected channels of benefits of the programme (employment, wages, and output), but may not account for all potential impact channels, such as improvements in health or wellbeing.

- The modelling reflects impacts on Universal Credit and taxation but excludes devolved benefits and non-Universal Credit DWP payments. While this omits some benefit types, Universal Credit is the primary income-related benefit in the UK and across all devolved administrations and captures a substantial share of employment-related fiscal effects.

Policy recommendations

In evaluating the implementation and process of the STEP Ukraine programme, we have identified several changes that are likely to improve the programme's effectiveness. While many of these recommendations may come with associated costs that need to be weighed against their potential benefits, they are worthy of consideration for enhancing outcomes for participants.

- **Design for diversity and flexibility**: Tailor eligibility, intensity, and support tracks to account for varying English proficiency, employment experience, and personal circumstances (e.g., caring responsibilities). For example, providing sessions in the evening may support those who have childcare or work responsibilities, preventing them from attending the programme during the day. In addition, the course could provide more entry points based on prior knowledge of the English language, allowing tutors to provide more tailored training to those with lower levels of baseline language proficiency.

- **Integrate holistic, trauma-informed support**: Embed mental health resources, digital skills training, and childcare support alongside employment and language strands.

- **Enhance operational efficiency and staff capacity**: Limit EA caseloads to 25-30 clients, invest in staff training, and provide structured breaks for reflection and improvement.

- **Strengthen cross-component coordination**: Facilitate regular communication and collaboration between English tutors and EAs to maximise programme synergy.

- **Build sectoral partnerships and alumni networks**: Foster connections with employers, professional associations, and peer mentors to support long-term employment outcomes.

- **Refine outcome monitoring**: Include measures of confidence, motivation, and non-employment outcomes; maintain flexible Key Performance Indicators (KPIs) that support meaningful progress.

- **Plan for sustainability**: Develop ongoing support mechanisms (e.g., alumni groups, drop-in sessions) to maintain engagement and integration post-programme in addition to the existing 12-month action plan.

- **Future programmes should implement a similar, successful English course**: The language impact evaluation suggests that the implemented English course should be considered a successful model for informing the design of future programmes that share this objective.

Conclusion

The STEP Ukraine programme demonstrates that rapid, integrated support combining intensive language learning with personalised employment guidance can produce measurable benefits for beneficiaries. A flexible, culturally attuned, and responsive approach is essential to meet the diverse needs of displaced populations. As refugee and forced migrant flows persist, the lessons from STEP Ukraine offer valuable direction for policymakers and practitioners seeking to design impactful and scalable integration programmes.

1 The STEP Ukraine Programme

This report presents the findings from the evaluation of the STEP Ukraine programme pilot, an initiative developed by the Ministry of Housing, Communities and Local Government (MHCLG)[1] to enhance English language proficiency and improve employment access for Ukrainian nationals seeking refuge in the UK following the outbreak of the war between Russia and Ukraine. This conflict has compelled many Ukrainians to flee their country in search of safety. The programme was designed to address the specific needs of these individuals, facilitating their integration into British society and enhancing their access to the labour market. The evaluation is being conducted independently by Alma Economics and RAND Europe on behalf of MHCLG.

The STEP Ukraine programme was delivered by World Jewish Relief (WJR), in partnership with the British Council. WJR is a registered charity that supports the resettlement of refugees and forced migrants in the UK. The programme was initially scheduled to run from August 2023 to August 2024. Participation was voluntary, with 17,282 signing up for the programme across the UK between 11th August 2023 and 10th August 2024, of whom 9,813 were onboarded[2]. Following a contract extension, delivery was continued from August 2024 to March 2025. In total, 45,512 Ukrainians across the UK accessed the programme across the initial stage and the extension.

The programme provided intensive support over a 12-week period, combining English language tuition with tailored employment support. At the outset, participants' needs, English proficiency, and career aspirations were assessed to align the support with their individual career goals. WJR Employment Advisers (EAs) offered practical guidance on CV development, job applications, and interview preparation, while the British Council delivered structured English language training. Language proficiency was assessed at the start, mid-point, and end of the programme to monitor progress.

Participation in STEP Ukraine was free of charge and targeted Ukrainian visa holders aged 18 or over who were unemployed or in low-paid employment. Eligible participants included those who entered the UK under the Homes for Ukraine Scheme, the Ukraine Family Scheme, or the Ukraine Extension Scheme.

STEP Ukraine took a holistic approach to supporting Ukrainian visa holders, addressing immediate needs such as English language acquisition alongside longer-term objectives of securing and sustaining employment. The programme comprised two core components:

- English language support: ten weeks of daily 2.5-hour English classes focusing on practical, work-related English skills delivered in small groups.
- Employment support: 12 weeks of employment support with a dedicated adviser offering support with CV writing, job applications, interview preparation, and job searching.

[1] Previously known as the Department for Levelling Up, Housing and Communities (DLUHC) from September 2021 to July 2024.

[2] STEP Ukraine Quarterly Reporting. 11 August 2023 – 10 August 2024 (MHCLG, 2024).

1.1 Characteristics of STEP Ukraine participants

The STEP Ukraine programme engaged a diverse participant group, reflecting both the scale of Ukrainian displacement into the UK and the varied socio-demographic characteristics of those seeking to enhance their labour market readiness. Monitoring data highlight patterns across length of stay, age, gender, and regional distribution. Table 1.1 summarises the main characteristics of the 9,813 participants who started the initial programme.

Participants exhibited considerable variation in their length of residence in the UK. The largest group had been in the UK for 13-18 months, with 7,124 sign-ups and 4,321 participants commencing the programme. Participants who had arrived more recently (0-6 months) also showed high engagement, with 3,283 sign-ups and 1,923 starters. The data suggest that both recent arrivals and those who had resettled sooner after the conflict began demonstrated strong motivation to access support.

The cohort was skewed towards middle-aged adults, with the 35-44 age group representing the largest segment (7,133 sign-ups, 4,320 starters). Participation among younger adults (18-24) and older adults (65+) was comparatively very limited, reflecting differences in the likelihood or capacity to actively seek labour market integration. In line with broader migration patterns, the programme attracted a predominantly female cohort: nearly four in five participants were women (13,686 sign-ups, 7,881 starters), while male participants accounted for 3,556 sign-ups and 1,910 starters.

Geographically, participants were broadly distributed across the UK, with concentrations in urban centres such as London and the South East, as well as in Scotland. This pattern reflects settlement arrangements under the Homes for Ukraine scheme and may also indicate the relative ease of accessing digital programme provision in urban areas.

Overall, STEP Ukraine successfully engaged a diverse population, predominantly 25-54-year-old females, but with notable variation in their length and area of residence in the UK.

Table 1.1 Profile of STEP Ukraine participants at a glance (August 2023-August 2024)

Length of time in the UK	Sign-ups (cumulative)	Started programme (cumulative)
0-6 months	3,283	1,923
7-12 months	2,644	1,595
13-18 months	7,124	4,321
19-24 months	3,363	1,594
Over 24 months	458	147
Unknown	410	233
Age range	**Sign-ups (cumulative)**	**Started programme (cumulative)**
18-24	1,401	640
25-34	3,729	2,046
35-44	7,133	4,320
45-54	3,511	2,102
55-64	1,143	561
65+	284	92
Sex	**Sign-ups (cumulative)**	**Started programme (cumulative)**
Male	3,556	1,910
Female	13,686	7,881
Other/unknown	40	24
Region or devolved administration	**Sign-ups (cumulative)**	**Started programme (cumulative)**
London	3,189	1,671
North East	665	366
North West	1,467	791
Yorkshire & Humber	1,012	627
East Midlands	928	516
West Midlands	1,059	620
East of England	1,089	636
South East (excl. London)	2,075	1,227
South West	2,142	1,216
Wales	865	472
Scotland	2,454	1,448
Northern Ireland	332	220
Unknown	5	3
Total	**17,282**	**9,813**

Source: STEP Ukraine Quarterly Reporting. 11 August 2023 – 10 August 2024 (MHCLG, 2024)

1.2 Objectives of the evaluation

The evaluation aimed to assess a novel model of support designed to improve the English language proficiency and employment outcomes of Ukrainian nationals who arrived in the UK following Russia's invasion of Ukraine. The need for rigorous evaluation was accentuated by the model's innovative delivery, which was fully remote and designed to offer a high degree of flexibility for participants.

The evaluation comprised four interrelated components:

- The process evaluation examined how STEP Ukraine was designed, delivered, and experienced by stakeholders and end-users. It also provided insights into how and why outcomes were achieved, complementing the findings of the impact evaluation.
- The English language impact evaluation focused on quantitatively estimating the impact of the programme on English skills. It compared participants' English language skills before and after the programme.
- The employment impact evaluation focused on quantitatively estimating the impact of the programme on participants' employment outcomes by comparing them to a control group of non-participating Ukrainian guests, including the likelihood of being in employment, the likelihood of being in sustained employment, and the monthly take-home pay.
- The economic evaluation estimated the VfM of the programme. It used the impact estimates for labour market outcomes as inputs to this analysis.

Although the programme was subsequently expanded to include additional cohorts of Ukrainian visa holders, this evaluation focuses on the initial cohorts of participants who engaged with the programme until August 2024.

2 The Process Evaluation

The objective of the process evaluation was to examine how STEP Ukraine was designed, delivered, and experienced by stakeholders and end-users. It was designed to complement the impact evaluation by providing insights into how and why the programme achieved its intended outcomes. The key findings and takeaways of the section are summarised in Box 2.1 below.

Box 2.1 Key findings from the process evaluation

Designing the programme:

- **STEP Ukraine's dual-focus model, which combined intensive online English training with personalised employment support, addressed both participants' professional aspirations and immediate integration needs, aiming to accelerate labour market entry and progression into higher-skilled roles.**
- **The fully online and adaptable delivery enabled the programme to reach participants across the UK**. This was particularly important under the Homes for Ukraine scheme, where participants were often hosted in rural or suburban areas with limited access to local services.
- **The three-stage structure (Entry, Delivery, Planned Exit) of STEP Ukraine provided a clear pathway through the programme from onboarding through to transition out of the programme**. This staged approach was intended to balance intensive, time-limited support with a structured exit strategy, though it also created pressure points at each handover stage.
- **A range of iterative adaptations during the design phase improved feasibility, enhancing the programme's practicality and participant retention**. KPI targets were revised to reflect achievable outcomes, and recruitment and onboarding processes were strengthened to reduce early-stage dropout rates.
- **Some learners struggled with the programme's intensity, exclusive use of English, and limited grammar support for beginners**. This finding was particularly true for lower-level learners and those facing trauma or competing responsibilities.

Delivering the programme:

- **The programme was launched within just six weeks of contract award, demonstrating operational agility and the ability to respond quickly to urgent participant needs**. This compressed timeline limited opportunities for detailed planning, comprehensive staff recruitment, and the finalisation of operational processes.
- Participants primarily accessed STEP Ukraine through informal Ukrainian networks and community intermediaries (e.g., Facebook, Telegram, community newsletters, and hosts). While this ensured strong reach within established networks, it limited accessibility for those less connected.

- The application process was generally straightforward, but delays in communication created anxiety and uncertainty. In addition, a perceived lack of transparency around the recruitment and onboarding process risked harming trust and resulted in disengagement from the programme.
- Employment support was highly individualised, with Ukrainian-speaking EAs offering culturally sensitive one-to-one guidance. This one-to-one guidance could have continued longer-term if follow-up support sessions were provided, and this would likely have benefited participants.
- Strategic partnerships expanded provision (e.g., Business in the Community (BITC), The Entrepreneurial Refugee Network (TERN)), though high caseloads, variable adviser practice, and mismatched expectations sometimes reduced the consistency and intensity of support.
- Employment support was most valued by participants when appropriately tailored to individual goals, including CV adaptation, job application coaching, and interview preparation.
- A lack of flexibility in scheduling appointments and language choice was a key barrier to participation (i.e., providing a choice of Ukrainian- or English-speaking EAs). Participants expressed difficulties in attending sessions where they had prior commitments, such as childcare, and participants with lower levels of language proficiency when joining the programme struggled to gain the full benefit of courses.
- Sector-specific employment support and recognition of Ukrainian qualifications were identified as gaps, which directly limited participants' ability to secure roles that were aligned with their prior careers.
- Integration between language and employment strands remained limited. While curriculum-level alignment existed, weak communication between EAs and English tutors limited cross-component synergy, constraining the potential of a fully coherent, mutually reinforcing programme.

2.1 Methods and data

Research questions

The process evaluation focused on the following key questions:

1. To what extent was the programme delivered as intended?
2. Who did the programme reach, who did it not reach, and why?
3. What aspects of the programme worked well, and which were less effective?

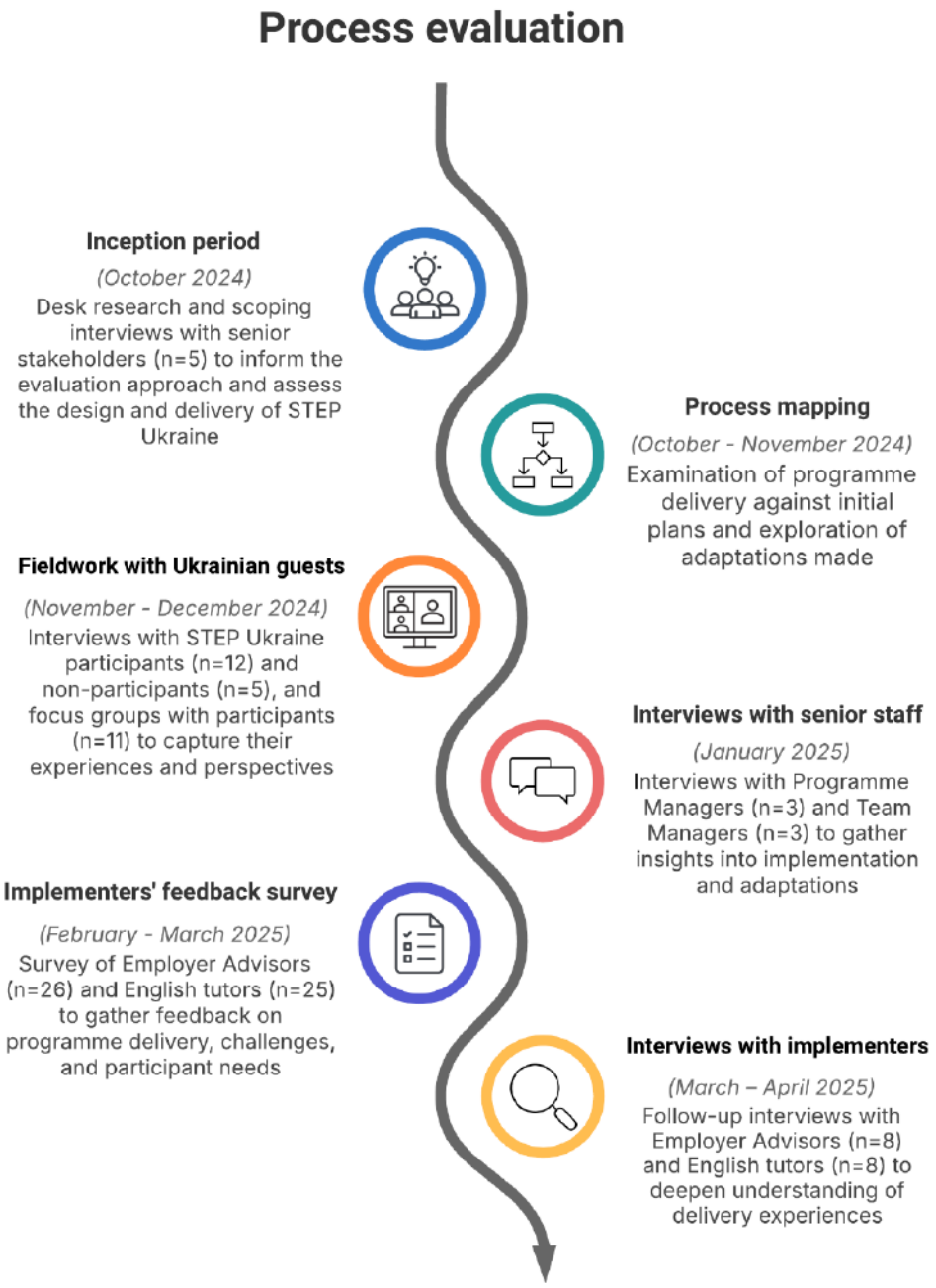
Our approach to the process evaluation was theory-based, participatory, and employed a mixed-methods design that triangulated evidence from multiple sources (desk research, surveys, interviews, focus groups, process mapping) and perspectives (Ukrainian visa holders, senior stakeholders, programme implementers) to investigate programme design, delivery, and experience. An overview of the methods and how these helped answer the evaluation questions is presented in Table A1.1 in Appendix 1.

The sections below outline the main data sources used, the strengths and limitations of the process evaluation, and the Theory of Change (ToC) underpinning the analysis.

Data sources

The process evaluation of STEP Ukraine drew on multiple complementary data sources to develop a comprehensive understanding of programme delivery, adaptations, and outcomes. Data collection took place between October 2024 and April 2025. The main components are summarised in Figure 2.1, and all interviews and focus groups are presented in Table 2.2 below.

Figure 2.1 Summary of the methods used for the process evaluation



Source: RAND Europe.

Desk research

Desk research provided the foundational evidence base for the evaluation. It included a review of programme documentation, operational reports, intervention materials (including the initial Theory of Change[3]), and research conducted by WJR and the British Council, as well as qualitative data from beneficiary surveys. This review informed subsequent analytical stages, including survey and interview design, and guided the process mapping exercise (see Appendix 1 for a full list of resources).

Scoping interviews with senior stakeholders

Initial scoping interviews were conducted with five senior strategic and operational leads to clarify the evaluation scope, confirm the direction of research, and identify key areas of focus. These interviews, conducted alongside discussions with programme and team managers, also informed the co-design of data collection tools, including the implementer survey and interview topic guides for EAs and English tutors.

Process mapping

Process mapping examined how STEP Ukraine was delivered in practice. Key steps included:

- Identifying stages of delivery, from participant recruitment to completion.
- Producing a visual flowchart outlining responsibilities, decision points, and potential outcomes.
- Detecting adaptations made during delivery using the ToC, an issues log maintained by delivery partners, and stakeholder interviews.
- Characterising adaptations with the Model for Adaptation Design and Impact (MADI) framework to assess their type (systematic or unsystematic, proactive or reactive) and potential impact (see Table A1.2 in Appendix 1).

Interviews and focus groups with Ukrainian visa holders

Semi-structured interviews and focus groups captured the perspectives of Ukrainian visa holders, including both participants and non-participants, to understand their experiences with STEP Ukraine and factors shaping language and employment outcomes. Participants were sampled to achieve diversity in gender, age, location, employment status, and English proficiency.

- Participants: 12 interviews and 11 focus group participants;
- Non-participants: 5 interviews.

Interviews were conducted online in English or Ukrainian, depending on participant preference, while focus groups were conducted in English. Accessibility measures

[3] This was developed by MHCLG as a Logic Model in early stages of programme design (not publicly available).

included translated materials (including privacy notices and information sheets), peer researchers trained to conduct interviews and focus groups, and flexible participation options (pre-recorded responses or written submissions). Non-participant interviews were included to provide a counterfactual perspective and deepen understanding of programme reach, accessibility, and barriers to engagement.

Interviews with programme and team managers

Interviews with Programme Managers (3) and Team Managers (3) explored programme implementation, responsibility allocation, decision-making processes, and adaptations to support delivery[4]. Key roles of the individuals interviewed included leading outreach, partnership and marketing efforts, coordinating support services, designing and delivering inductions, providing specialised support, and monitoring activities.

The interviews followed a semi-structured format, allowing interviewees to share their individual perspectives while generating comparable data for analysis. Conducted remotely via Microsoft Teams, interviews lasted around 60 minutes and were recorded for accuracy with consent from interviewees.

Implementers' feedback survey

The fieldwork with implementers followed a sequential explanatory design[5], whereby quantitative data was collected and analysed first, followed by qualitative data collection and analysis in quick succession. Initially, an online survey was administered to all implementing staff whose contacts were provided by WJR and the British Council, including 103 EAs and 65 English tutors. The survey collected data on programme delivery, challenges, participant support, and overall feedback. Although initially designed with closed-ended questions, open-ended items were incorporated to capture richer qualitative insights. The survey remained open for two weeks between February and March 2025. Table 2.1 summarises the number of responses received and the corresponding response rates across the two participant groups.

Table 2.1 Summary of survey participants and response rates

Participant group	Contacts received	Survey respondents	Response rate*
EAs	103	26	25%
English tutors	65	25	38%

**Note: Response rates are approximate, as there remains a small possibility that an individual may have completed the survey more than once.*

Source: RAND Europe.

[4] While the initial plan envisaged up to three interviews per organisation, five interviews were ultimately conducted with WJR staff and one with the British Council. The interview with the British Council was further supplemented by a detailed background document prepared by the organisation for the evaluation team, outlining the programme's phases, instructional and assessment approaches, and other relevant details.

[5] Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. *Field Methods, 18*(1), 3-20. https://doi.org/10.1177/1525822X05282260

Survey responses were analysed using descriptive statistics and cross-tabulations for closed-ended questions. Open-ended questions were analysed using thematic analysis (see below).

Follow-up interviews with implementers

Follow-up qualitative interviews with survey respondents provided in-depth insights into programme delivery. 16 interviews were conducted: 8 with EAs (split equally between English and Ukrainian delivery) and 8 with English tutors (covering all five levels of English provision). Interviews explored participant engagement, retention strategies, caseloads, staff workloads and training, coordination between employment and language support, and links with wider services. They followed a semi-structured format, guided by a topic guide developed in consultation with delivery partners and approved by MHCLG. They were conducted remotely via Microsoft Teams and lasted 45-60 minutes.

All qualitative data collected in the evaluation underwent thematic analysis, beginning with a thorough review to understand content and scope. The extracted data was organised into a pre-defined analytical framework aligned with the evaluation questions and the process mapping findings. The framework was iteratively refined based on emerging patterns, recurring themes, and insights identified during the analysis, ensuring it captured the complexity and nuances of the data. Themes were then categorised by applying codes to relevant portions of the dataset and compared across participant groups. Finally, the key themes and findings were examined in greater depth to provide detailed insights into how the programme was delivered and experienced in practice. Table 2.2 summarises all interviews and focus groups conducted as part of the process evaluation.

Table 2.2 Summary of interviews and focus groups

Interviewee/Data source	Role/function	Number of interviews/ participants	Report reference code	Timing
Senior Leads	Senior strategic and operational leads	5	SL1-SL5	Oct 2024
Ukrainian visa holders – participants	STEP Ukraine programme participants	12 (interviews) 11 (focus groups)	Participants	Oct 2024-Jan 2025
Ukrainian visa holders – non-participants	Eligible but non-enrolled beneficiaries	5	Non-participants	Oct 2024-Jan 2025
Programme Managers	Programme managers	3	PM1-PM3	Oct-Nov 2024
Team Managers	Team managers at delivery partners	3	TM1-TM3	Oct-Nov 2024
Employment Advisers	Frontline employment support staff	8 (interviews)	EA1-EA8	Feb-Mar 2025
English Tutors	Language provision staff	8 (interviews)	ET1-ET8	Feb-Mar 2025

Source: RAND Europe.

Communities Workshop

In addition to the core fieldwork, a Communities Workshop was conducted to gather broader stakeholder perspectives and validate emerging insights from the evaluation. While not formally part of the primary data collection, the workshop served multiple purposes: to raise awareness of STEP Ukraine, exchange experiences, identify lessons for future delivery, and discuss the provision of more distributed support across the UK.

The workshop engaged 12 participants representing a mix of stakeholders, Ukrainian guests, third-sector organisations, and local integration support providers in different areas and devolved administrations. Discussions were structured around different elements of the STEP Ukraine programme, remaining barriers, and unmet needs for Ukrainian visa holders, mapping available support against the Home Office's Indicators of Integration Framework[7], and discussing recommendations for future programmes.

Strengths and limitations of the data

This process evaluation drew on a broad evidence base, incorporating perspectives from Ukrainian participants, non-participants, and professionals involved in designing and delivering the STEP Ukraine programme. It was further strengthened by national-level stakeholder input and guidance from a steering group of subject matter experts.

The mixed-methods design sought to balance breadth and depth, combining qualitative and quantitative approaches. While these features enhanced robustness, several limitations should be noted:

- **Lack of observations**: Planned non-participatory observations of programme activities (e.g., English classes, employment workshops) could not be undertaken due to timing constraints, limiting insights into how components were delivered in practice.
- **Partner engagement delays**: Limited availability of WJR and the British Council delayed access to stakeholders, the development of research tools, and the start of fieldwork. WJR subsequently provided constructive input and access, while the British Council's involvement was restricted to one interview and a background document. This left gaps in the understanding of the English language training component.
- **Survey questionnaire changes**: The British Council requested the removal of some tutor survey questions (e.g., curriculum content, assessment design, pace, course structure), as they believed that providing this information would be against their internal codes of conduct. These were replaced with open-ended items, allowing data collection but limiting comparability with EA responses and reducing the depth of findings.
- **Survey administration limitations**: Delivery partners facilitated access to EAs and English tutors' details for survey participation, limiting the evaluation team's oversight of the distribution process and its control over who was or was not included in the survey pool. Response rates were low (25% for EAs; 38% for English tutors), so findings should be interpreted with caution as they are not representative of all implementers.

[7] Ndofor-Tah, Caroline; Strang, Alison; Phillimore, Jenny; Morrice, Linda; Michael, Lucy; Wood, Patrick; et al. (2019). Home Office Indicators of Integration framework 2019. University of Sussex. Report. https://hdl.handle.net/10779/uos.23468528.v1

Theory of Change of STEP Ukraine

The integrated design of intensive English language classes and personalised employment support was intended to address multiple resettlement needs in parallel. This is based on the premise that combined support would produce better outcomes than siloed interventions.

Figure 2.2 below presents a logic model of the STEP Ukraine programme, following the UK Magenta Book format. It outlines the programme's key inputs, activities, outputs, short- and medium-term outcomes, and anticipated long-term impacts with respect to English language and employment.

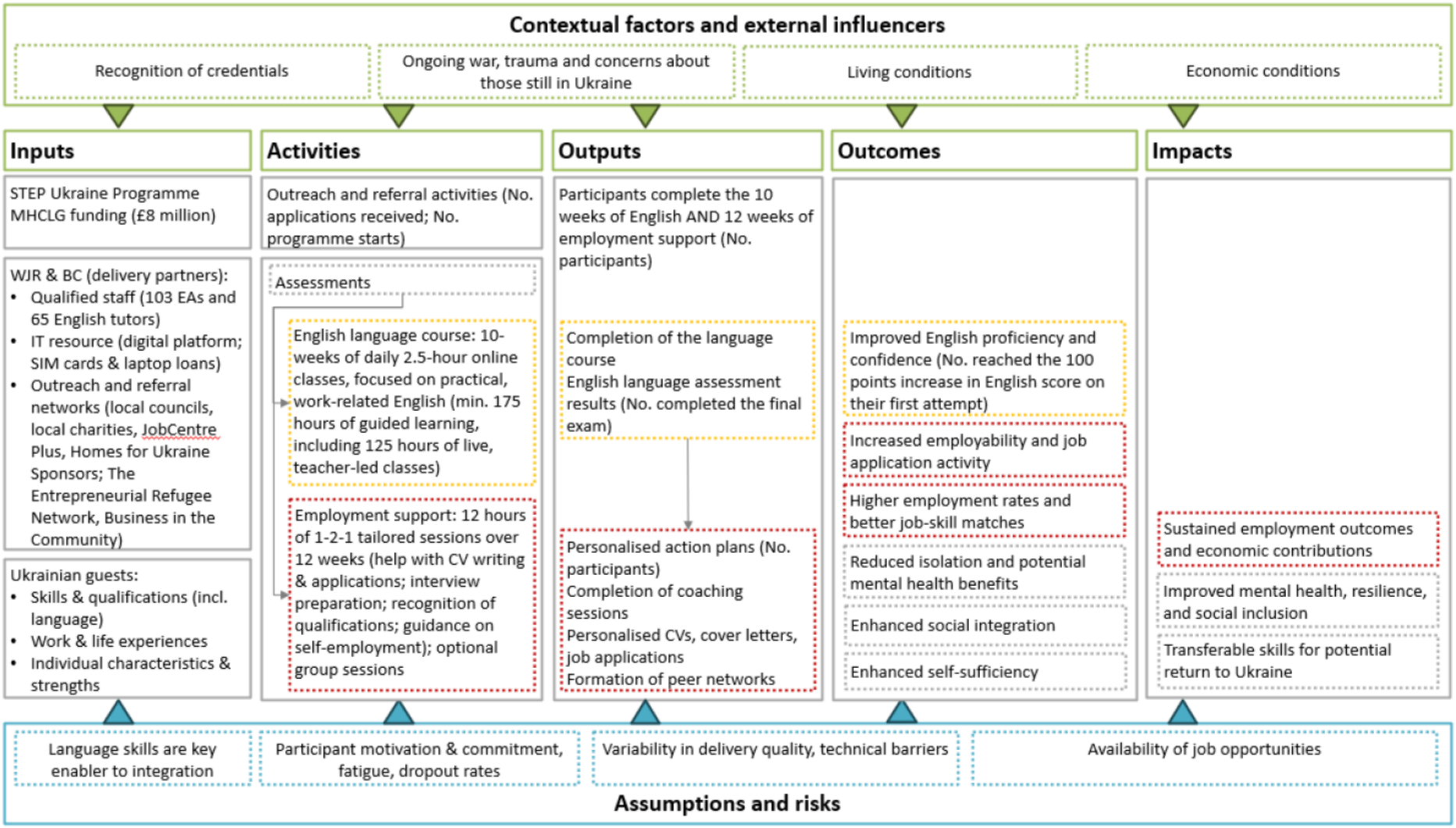
Participants were expected to complete English classes and coaching sessions, develop personalised CVs and cover letters aligned with UK standards, and create action plans with short- and long-term goals. English assessments were built in to track progress, with certificates available to reflect learning achievements – CEFR was picked as the accrediting service as it is internationally recognised and is used across the EU. Together, these outputs were designed to equip participants with stronger language abilities and the tools needed to pursue realistic employment opportunities.

The programme aimed to achieve early outcomes, such as measurable improvements in English proficiency, greater confidence, and enhanced understanding of how to search and apply for jobs in the UK. They were expected to become more independent in navigating recruitment processes. Over time, these skills were anticipated to translate into higher rates of employment and improved job matches, potentially leading to better earnings and job satisfaction.

The longer-term impacts envisaged included sustained participation in the labour market, financial independence, and reduced reliance on public services. Participants were expected to move from initial job placements into ongoing, stable work, thereby contributing to the UK economy and helping address labour shortages. Beyond these economic benefits, it was anticipated that improved financial stability and social integration would support greater wellbeing for participants and their families. Stronger community ties were also expected to enhance cohesion within host communities, while improved English proficiency and professional experience were seen as skills of lasting value for those choosing to return to Ukraine in the future.

This ToC was based on the assumption that language development would be a key enabler of integration and employability, while recognising that outcomes could be influenced by external factors such as housing instability, childcare responsibilities, or unrecognised qualifications. It also acknowledged risks, including variable quality of delivery, digital barriers, and the possibility of dropouts among those facing the greatest challenges. Nonetheless, the underlying programme logic was that by combining intensive language support with personalised employability coaching and wider referrals, participants would be equipped with the tools, confidence, and networks needed to achieve sustainable work and long-term integration.

Figure 2.2 Theory of Change for the STEP Ukraine programme



2.2 Setting up STEP Ukraine

This section examines the processes involved in setting up and launching the STEP Ukraine pilot, drawing on desk research, interviews with stakeholders and senior staff, and fieldwork with implementers. It explores the strategic planning that shaped the programme's dual-focus design, its integrated monitoring framework, the opportunities and challenges of a rapid mobilisation period, and the challenges, enablers, and adaptations made to the recruitment and onboarding process.

Strategic planning to meet the unique needs of Ukrainian visa holders

The STEP Ukraine pilot programme was launched through a competitive tender process led by MHCLG and represented a strategic attempt to address the unique circumstances of Ukrainian visa holders, including the short term of their visas, which allowed between 18-36 months of stay depending on application date, with no offer of indefinite leave to remain. Strategic and operational leads consistently described this group as "eager to work" and having "clear professional aspirations"[8]. This emphasis on employability shaped the programme's design, which was explicitly framed around facilitating rapid labour market entry or progression into higher-skilled and better-paid roles[9].

Evidence collected by MHCLG, however, revealed that employability support alone would not suffice. Despite generally high levels of education and IT proficiency, language barriers emerged as the primary constraint on employment outcomes, often outweighing participants' skills or motivation. As one stakeholder reflected, "*We had done in advance ONS research showing how English was a barrier and the classes that were available for participants were not good enough. They wanted something bespoke. … They wanted something quick and intensive*"[10]. This highlighted a critical tension between participants' aspirations and existing service provision, which follows the traditional academic term timetable. This prompted the adoption of a dual-focus model combining intensive English for Speakers of Other Languages (ESOL) instruction with personalised employment advice to support Ukrainians in becoming economically independent sooner than would be expected using traditional term-time provision.

This design marked a clear departure from existing ESOL provision, which was viewed as "not feasible" for this group due to transportation difficulties in rural areas, challenges in balancing caring responsibilities, and constraints arising from short visa durations[11]. STEP Ukraine was instead conceptualised as a fully online intervention, offering flexible scheduling. The new model aimed to facilitate flexible participation and prioritised accessibility and responsiveness to participants' needs.

As part of their bidding process, delivery partners played an important role in shaping this approach. WJR, leveraging 30 years of experience in Ukraine and involvement in both the Homes for Ukraine scheme and pilot employment programmes for

[8] PM1, PM2, SL2, SL3, SL4, TM2 (see Table 2.2).

[9] SL2, SL4, SL5 (see Table 2.2).

[10] SL4 (see Table 2.2).

[11] SL2 (see Table 2.2).

Ukrainian and Afghan clients in the UK, recognised that Ukrainian clients differed significantly from other displaced groups[12]. As one programme manager observed, "*For other refugees, working in the UK might seem so far away from where they currently are, whereas a lot of our Ukrainian clients were very keen to work… A lot of our Ukrainian clients, if they were a lawyer in Ukraine, they wanted to be a lawyer straight away in the UK*"[13]. This reinforced the need for tailored provision, with the challenging task of balancing participants' expectations for professional continuity against the structural realities of the UK labour market. The pilot work conducted by WJR informed the development of STEP Ukraine and the search for suitable providers like the British Council.

The Homes for Ukraine scheme added further complexity. Unlike traditional resettlement programmes where displaced populations are housed in areas with established refugee support infrastructure, Ukrainian guests were largely hosted by volunteers in more affluent, often rural or suburban, communities. As one programme manager explained, "*[W]ith the Ukrainian community, it was done through hosts, it was people volunteering… which meant that the infrastructure wasn't there*"[14]. This self-selecting resettlement pattern created new challenges, particularly limited access to specialist services, affordable housing, and public transport. It also introduced financial pressures, as participants needed higher-paid work to remain in these communities[15].

An initial ToC, developed internally by MHCLG, provided a strategic blueprint, integrating an online delivery model with the two core components of employment and ESOL support.[16] The ToC articulated the mechanisms through which STEP Ukraine could overcome structural barriers and tailor support to participant needs, including mitigating limited access to IT resources and reducing logistical burdens. However, the extent to which this strategic vision could be realised in practice depended on ongoing adaptability and the iterative refinement of delivery processes in response to emerging challenges.

An integrated monitoring framework

STEP Ukraine's online and flexible delivery model was explicitly designed to address structural and logistical barriers faced by Ukrainian visa holders. However, as an untested model tailored to a novel context, it carried inherent uncertainties regarding its feasibility and scalability. MHCLG framed the programme as a pilot from the outset, embedding a monitoring framework intended to assess immediate outputs and explore the programme's potential for long-term replication.

The programme's KPIs reflected this dual ambition of delivering tangible outcomes while generating evidence for future interventions. Table 2.3 summarises the established performance targets, and Box 2.2 presents a case study on the revision of KPIs and participant targets.

---

[12] PM1, PM3 (see Table 2.2).

[13] PM1 (see Table 2.2).

[14] PM1 (see Table 2.2).

[15] PM1 (see Table 2.2).

[16] SL2, SL4, SL5 (see Table 2.2).

Table 2.3 STEP Ukraine KPIs

KPI No.	Description	Target / Measure
1	Programme participation	Minimum 8,000 learners; aspirational target ~10,000 participants to guarantee programme viability and scalability.
2	Tailored support plans	90% of all students (excluding dropouts) to have a plan identifying employment barriers and steps to reduce them.
3	Employment support satisfaction	≥85% of course completers report sessions were helpful (via exit survey).
4	English language progress	All course completers to advance by at least one Common European Framework of Reference for Languages (CEFR)[17] level from baseline assessment.
5	Receipt of employment support	All students (excluding dropouts) to receive tailored employment support.
6	Employment outcomes	≥40% of unemployed students (excluding dropouts) in employment within 12 months; employment defined as paid contract or self-employment.
7	Earnings or employment progression	≥20% of low-paid employed participants (excluding dropouts) report increased earnings or employment more commensurate with skills within 12 months.
8	Mid-programme English course completion	≥35% of learners (excluding dropouts) complete English Learning component by month six.
9	Programme completion	All students (excluding dropouts) complete English Learning component by end of contract.
10	Delivery milestones	a. Launch programme sign-up within 1 week of contract award. b. Launch training course within 4 weeks of contract award. c. End course enrolment ≥10 weeks before contract end. d. Conduct exit survey within 2 weeks of course end. e. Interim reports quarterly; final report within 2 weeks of final course end. f. 12-month follow-up survey completed within 12 months of course start.
11	Webinars/ events	Organise at least 20 webinars/events providing insights into skills for career opportunities; hosted by 31st August 2024 *[NEW KPI].*

[17] Common European Framework of Reference for Languages (CEFR) is a framework for grouping language proficiency across 6 levels: A1, A2, B1, B2, C1, C2. These levels can be organised into three broader groups: Basic User (A1, A2), Independent User (B1, B2) and Proficient User (C1, C2). The international implementation and recognition of CEFR levels in language teaching and testing made this framework appropriate for STEP Ukraine, increasing the transferability of participants' attainment during the programme to other contexts outside the UK.

Box 2.2 Case study: Revision of KPIs and participant targets

Original challenge:

The programme was initially shaped by ambitious KPIs and targets. For example, the tender specification required all participants to improve by a full CEFR level, while subsequent expert advice indicated that sub-level progress was a more realistic goal. Attendance rates were set at 90%, and MHCLG anticipated delivery to 9,100 participants. However, the novelty of the programme made it difficult to assess whether these targets were achievable in practice.

While these KPIs provided a clear framework for measuring success, stakeholders subsequently reflected that several targets were overly ambitious relative to the operational realities and the pilot nature of the programme[18]. Initial expectations were later described as "over-ambitious" or even "ridiculous" by one key stakeholders[19]. Another stakeholder noted that the KPIs were "very result-driven," and that a revision around February 2024 was necessary to make them more realistic [20].

Stakeholders from MHCLG later acknowledged that their participant engagement projections had been optimistic, not accounting for difficulties in participant retention and survey completion rates[21]. They also expressed concern about the introduction of mid-point assessments by WJR and the British Council, cautioning that these could burden participants or inadvertently increase attrition if learners felt they had already achieved satisfactory progress.

Adaptation implemented:

Midway through delivery, MHCLG revised the KPI framework and payment schedule to better align with programme realities, including:

1. English language progress: The target was changed from improving English by one full CEFR level to a sublevel, which was in practice measured by 100-point improvements based on the English score system (0-600 points).
2. Participation and programme numbers: The initial discrepancy between WJR's minimum target of 8,000 participants and MHCLG's contractual target of 9,100 was addressed.
3. The definition of a 'dropout' was a matter of ongoing debate between MHCLG and delivery partners. Eventually, participants who withdrew from one of the programme's components were classified as 'partly engaged,' meaning that they would not necessarily be regarded as having fully withdrawn from the programme.

[18] SL2, SL4 (see Table 2.2).

[19] SL2 (see Table 2.2).

[20] SL4 (see Table 2.2).

Rationale and benefits:

The revision reflected recognition that the initial KPIs risked creating unrealistic benchmarks and "bad incentives" for delivery partners. By resetting the framework, MHCLG enabled a more balanced approach to performance measurement, avoiding undue pressure on specific outcomes and supporting a more accurate assessment of programme results. This adjustment also improved alignment between contractual obligations and operational capacity, enhancing collaboration between MHCLG and the delivery partners.

Impact:

The recalibrated KPIs and targets created a more credible framework for monitoring outcomes and assessing success. Stakeholders highlighted that the revised system encouraged adaptability, reduced delivery strain, and promoted a focus on meaningful participant progress rather than compliance with unattainable numerical thresholds.

Rapid mobilisation for a swift launch

The programme had a mobilisation period of six weeks from conception to implementation. Findings from the process evaluation indicate that the period following the contract award was marked by intense pressure and accelerated timelines, reflecting the urgent need to operationalise the programme and deliver support to Ukrainian visa holders quickly. Programme Managers described this period as "really quick" and "high pressure," highlighting the trade-off between speed and thorough preparation[22]. As one Team Manager reflected,

"there's been a lot of learning along the way. You know, nothing's really perfect, but we did also start this programme and like from conception to actual implementation, it was six weeks. We had no time to start a monster of a programme, so there was a lot of learning along the way"[23].

This rapid mobilisation required the swift recruitment of staff, including EAs and Team Managers, to ensure that the programme could commence on schedule. However, stakeholders noted that the accelerated pace may have constrained the ability to select personnel with the strongest track records or to provide comprehensive training and induction[24]. As a result, some operational processes were adapted iteratively during delivery rather than being fully defined from the outset.
Another factor influencing this early stage of the programme was MHCLG's deliberate decision not to prescribe detailed delivery processes in the tender specification (for example, by avoiding setting out a predefined student journey). This approach was intended to maintain flexibility and neutrality, giving bidders scope to propose their own delivery models, with the expectation that processes for managing dropouts, pausing,

[22] PM1, PM3, TM1 (see Table 2.2).
[23] TM1 (see Table 2.2).
[24] PM1 (see Table 2.2).

and restarting would be developed during mobilisation[25]. In practice, the initial STEP Ukraine delivery team was small, with four core managers taking on a wide range of responsibilities[26]. This distribution of tasks meant that specialist expertise was limited in some areas at the outset, which in turn constrained the extent to which certain operational processes could be finalised during mobilisation.

While the short mobilisation period achieved the practical objective of launching the programme quickly, it limited opportunities for deeper planning and collaboration among delivery partners, which may have strengthened the programme's foundations. Programme Managers and government officials alike reflected that the phase offered potential to "nail down specifics based on the proposed bid," a chance perceived as not fully utilised[27]. This iterative refinement during implementation highlights both the programme's adaptability and the risks associated with compressed preparation periods, particularly in terms of consistency, staff readiness, and alignment of operational processes with intended programme objectives.

Participant recruitment and onboarding

The recruitment and onboarding processes for STEP Ukraine were designed to identify eligible participants and prepare them for the programme's intensive dual-focus delivery model. While the process was structured around clear stages, evaluation evidence suggests that its implementation involved iterative adaptations in response to operational realities and participant needs.

Initially, participants primarily self-registered online, with some referrals facilitated by partners. Registration forms captured key demographic and logistical information, including self-identified English proficiency, employment status, childcare responsibilities, and preferred English class times[28]. Following registration, participants who confirmed readiness were allocated to an EA and invited to join a ten-week English language course delivered by the British Council. Early steps included a welcome pack in both English and Ukrainian, a short eligibility call, and the development of a tailored action plan by the EA.

Despite these structured stages, stakeholders highlighted significant operational challenges. Early allocation of participants to EAs was conducted manually using spreadsheets, which was labour-intensive and risked inefficiencies in matching participants' English proficiency and location to available staff. As one Team Manager observed, "*we were filtering thousands of names manually on one spreadsheet where everybody's doing the same thing… one little mistake and you've lost someone, pretty much"*[29]. Moreover, the English placement test was initially scheduled later in the process, after EA allocation, which sometimes led to mismatches in participant streaming and required reactive adjustments[30].

---

[25] SL4 (see Table 2.2).

[26] TM1 (see Table 2.2).

[27] SL4 (see Table 2.2).

[28] SL1, SL3, TM1, TM2 (see Table 2.2).

[29] TM1 (see Table 2.2).

[30] PM1 (see Table 2.2).

Several structural and behavioural barriers affected recruitment efficacy. A persistent challenge was participant attrition prior to programme commencement, driven partly by misunderstandings about the programme's intensity and requirements. Stakeholders noted that some participants expected STEP Ukraine to function as a recruitment agency rather than an employment support and language programme, reflecting a misalignment of expectations[31]. As a Programme Manager explains, "*clients often didn't take the 2.5-hour daily English classes seriously at first; they expected something more transactional*"[32].

Communication barriers further complicated recruitment. Participants sometimes failed to engage via phone or email, often due to cultural differences in communication preferences or use of non-primary email addresses[33]. Inaccurate self-assessments of English proficiency also hindered early allocation to appropriate language course levels. Furthermore, while most participants were digitally literate, a minority required laptop loans or SIM cards to fully engage with online delivery. Efforts were made to access hardship funds through local councils to provide laptops; however, this process could be time-consuming, sometimes taking four to six weeks to secure a single device[34]. These challenges were compounded by the programme's intensity, which could be particularly demanding for a cohort of participants experiencing war-related trauma[35].

Despite these obstacles, a number of factors proved instrumental in enabling recruitment. Word-of-mouth and social media outreach within Ukrainian communities were highly effective, with some STEP Ukraine alumni effectively acting like informal programme 'ambassadors,' offering information about the programme to those interested in joining[36]. Collaboration with Strategic Migration Partnerships, local authorities, JobCentres and charities extended programme visibility[37]. Partnering with the British Council as the language delivery partner has also been characterised by Programme Managers as lending credibility to the programme, enhancing participants' confidence in engagement and being seen as more "professional" by prospective employers[38]. Finally, a dedicated programme assistance team, including Ukrainian speakers, improved responsiveness to queries and reduced early dropout risk[39].

The recruitment process evolved substantially through adaptive learning. A key improvement involved centralising communications, shifting from dispersed EA-led emails to coordinated messaging, reducing the risk of participants missing information or emails being marked as spam[40]. The programme introduced the "Are You Ready?" email and eligibility call, described in Box 2.3, which acted as a proactive screening mechanism to confirm participants' commitment and

---

[31] SL4, PM1, TM1, TM2 (see Table 2.2).

[32] PM2 (see Table 2.2).

[33] TM2 (see Table 2.2).

[34] TM1 (see Table 2.2).

[35] PM1, PM2, SL3, TM1, TM2 (see Table 2.2).

[36] TM1, TM2 (see Table 2.2).

[37] PM1, TM2 (see Table 2.2).

[38] PM1, PM2 (see Table 2.2).

[39] PM1 (see Table 2.2).

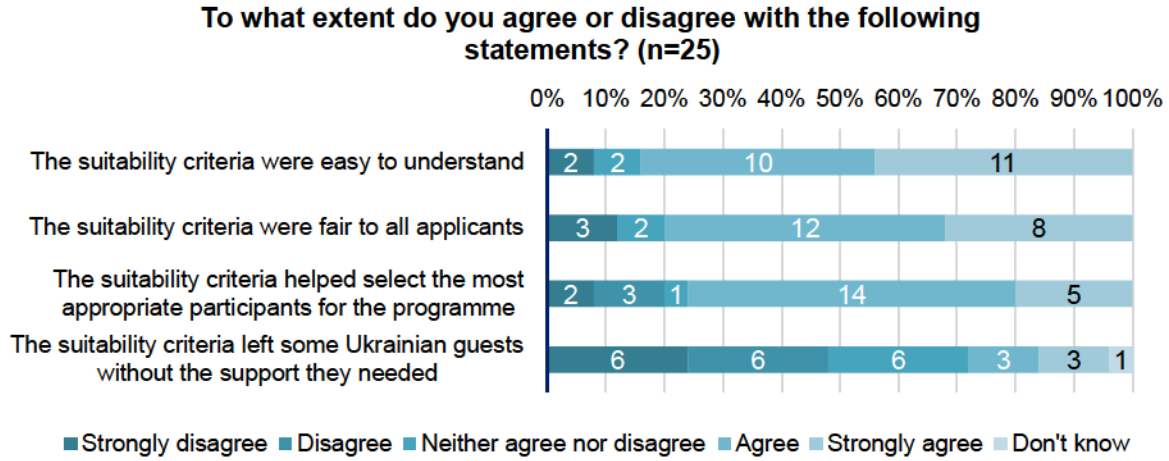
[40] PM1 (see Table 2.2).

understanding of programme requirements. This adaptation not only reduced early-stage dropouts but also improved the efficiency of the onboarding process by ensuring that EAs engaged with participants who were genuinely ready to participate and suitable for the programme.

Additionally, data management transitioned from Excel spreadsheets to a Customer Relationship Management (CRM) system, enhancing tracking of participant progress and operational oversight, though stakeholders acknowledged the system remained imperfect[41]. The timing of the English placement test was also adjusted earlier in the process to improve class streaming and reduce misalignment with EAs' caseloads.

Survey evidence supports the view that the programme's suitability criteria were broadly effective. 17 EAs among a total of 25 who responded to the survey said that they agreed or strongly agreed the criteria were fair (80%; 20/25), easy to understand (84%; 21/25), and helped select the most appropriate participants for the programme (19/25; 76%). Only 6 respondents (24%; 6/25) indicated that the criteria left some Ukrainian visa holders without the support they needed (see Figure 2.3). Surveyed EAs identified several circumstances contributing to exclusion, including insufficient time to commit to the programme, absence from the UK during the eligibility call, lack of biometric residence permits, and complete beginner-level English proficiency. These findings indicate that while the criteria largely facilitated equitable and effective participant selection, some structural barriers and contextual factors constrained access for certain groups.

Figure 2.3 EAs' perceptions of the suitability criteria



Source: RAND Europe, Survey with Employment Advisors.

[41] TM1 (see Table 2.2).

Box 2.3 Case study: Introducing the "Are you ready?" email and eligibility call

Original challenge:

Early stages of participant recruitment revealed a recurring issue: many individuals registered for STEP Ukraine but subsequently dropped out, forgot they had applied, or provided non-primary email addresses. This led to unfilled programme spaces being effectively lost and created significant challenges for EAs attempting to engage assigned clients.

Adaptation implemented:

To address this, WJR introduced a two-step intervention:

1. "Are you ready?" email: Sent approximately six weeks prior to programme commencement, this email required participants to actively confirm their interest and update contact details.
2. Eligibility call: Following confirmation, a 15-20-minute call with an EA verified visa status, age, and suitability for the programme, and ensured participants fully understood the dual focus on English language learning and employment support, its intensity, and the expected time commitment.

Rationale and benefits:

These steps were designed to secure client "buy-in" and encourage active engagement. By confirming interest and contact information, the email reduced the risk of unresponsive participants, while the eligibility call provided an opportunity for participants to understand programme expectations before formal onboarding. Collectively, this proactive screening ensured that only committed, eligible, and suitable individuals progressed, enhancing overall programme efficiency.

Impact:

The adaptation resulted in a significant reduction in early-stage dropouts, leading to higher retention among participants who commenced the programme. Staff workload was eased, as fewer unengaged participants reached later stages, and onboarding became more streamlined. Stakeholders noted that these measures were decisive in improving early engagement and securing a more committed cohort.

2.3 Delivery of STEP Ukraine

This section provides an examination of the implementation and delivery of STEP Ukraine, detailing how the programme's design was translated into practice to support Ukrainian visa holders. Drawing on desk research, stakeholder interviews, implementer surveys, follow-up interviews with implementers, and analyses of key enabling factors, challenges, and adaptations. It also discusses the integration of employment and language provision within a single delivery model and concludes with practical recommendations and insights to inform the design and delivery of future initiatives supporting displaced populations.

Core components of STEP Ukraine

The evaluation identified three interconnected stages in the delivery of STEP Ukraine: Entry, Delivery, and Planned Exit. These stages were designed to operationalise the programme's dual focus on personalised employment support and intensive English language training, yet the complexity and novelty of the approach introduced several implementation risks. A flowchart illustrating the STEP Ukraine process map is provided in Figure A1.1 in Appendix 1.

The main stages were as follows:

- Entry: The entry phase encompassed referral pathways, recruitment and registration, pre-intervention English assessment via the ''EnglishScore'' app, and the allocation of participants to an EA who conducted eligibility and suitability assessments alongside a pre-intervention survey. These steps were intended to ensure the successful onboarding of participants, reduce participant drop-out later in the programme, and provide a tailored start to the programme.

- Delivery: The main delivery phase consisted of 12 weeks of employment support provided by WJR and 10 weeks of intensive English language training delivered by the British Council. Employment support was highly personalised, with the provision of a personalised action plan and offering up to 12 hours of one-to-one sessions with an EA, supplemented by 'Working in the UK' group workshops, and employer-led seminars. Those interested in self-employment were instead referred to an alternative provision delivered by The Entrepreneurial Refugee Network (TERN). The English component was structured as a 10-week online course, delivering a minimum of 17.5 hours per week (totalling approximately 175 hours), with lessons stratified by proficiency level from A0/A1 to C1, and a mid-course and a post-intervention English test via the 'EnglishScore' app.

- Planned exit and follow-up: The exit stage involved the provision of an individualised exit plan and 12-month goals by the EA, potential signposting to other services, provision of an English certificate, and continued access to English Online resources for 6 months. A post-intervention survey and a 12-month follow-up survey were embedded to capture longer-term outcomes.

Delivery of employment support

The employment support component of STEP Ukraine was designed as a highly personalised intervention, offering up to 12 hours of one-to-one sessions with EAs over a 12-week period[42]. The overall structure balanced a mix of planned milestones and flexible delivery. The initial weeks were more structured, focusing on assessments and action planning, and the final week consolidated learning into a 12-month plan. The intervening weeks were deliberately flexible, enabling EAs to adapt support according to participants' evolving needs. This design recognised the heterogeneity of the Ukrainian cohort, which ranged from individuals requiring

[42] STEP Ukraine Logic Model (shared by MHCLG, not publicly available).

foundational CV support to highly skilled professionals navigating complex qualification recognition. However, it also placed significant responsibility on individual advisers to calibrate the intensity, scope, and relevance of support.

At the centre of the model was the personalised one-to-one relationship. Initial assessments identified participants' barriers to employment (such as childcare and other responsibilities, digital access, and IT literacy), allowing EAs to respond not only to work readiness but also to broader settlement challenges. Tailored action plans were co-produced to set short-, medium-, and long-term goals, blending pragmatic tasks (e.g., adapting Ukrainian CVs to UK norms) with confidence-building steps[43]. Support included services such as assistance with recognition of relevant professional qualifications; signposting to job boards and employment opportunities; support with CV writing and job applications; interview preparation, including role-playing cultural norms and providing actionable feedback; and guidance on self-employment. For those interested in self-employment, support included referrals to local entrepreneurship services, as well as training on key topics, such as business fundamentals, the UK tax system, and access to financial support and start-up capital.

"Everyone gets their own employment advisor, and it is very bespoke to the person's needs and desires. If somebody wants to rewrite their CV, the sessions will focus on rewriting the CV, and personal statements, and so on. But if somebody needs the transfer of their qualifications, which has been quite a big thing with Ukrainians, obviously then that's what we focus on. … The student, with the employment advisor, sets their own timelines and goals and so on." – Government official[44]

This individualised model aimed to empower participants by addressing their specific needs while equipping them with the tools required to navigate the UK labour market effectively[45]. However, from an implementation perspective, the reliance on EAs' discretion created some inconsistencies: some advisers offered holistic, future-oriented support, while others struggled to move beyond generic job search advice. Resources such as an EA handbook, templates, and an intranet were intended to standardise delivery and promote shared learning. While these provided a framework, the high degree of autonomy granted to EAs meant that practice varied considerably. The absence of systematic monitoring of support quality (beyond session attendance and participants' self-reported satisfaction) limited the programme's ability to ensure that personalised support translated into a consistent participant experience.

Optional group sessions were later introduced to complement the bespoke one-to-one approach. These centrally designed workshops addressed general employment topics and included employer seminars to provide insights into various sectors. Topics included CV writing, interview preparation, LinkedIn, confidence-building, and job searching, and created opportunities for participants to benefit from broader guidance[46].

---

[43] PM1, SL1, TM1, TM2 (see Table 2.2).

[44] SL4 (see Table 2.2).

[45] SL2, SL4, PM1, TM1, TM2 (see Table 2.2).

[46] SL1, PM1, SL3, TM1, TM2 (see Table 2.2).

The programme also incorporated features that extended beyond narrow employability, such as matching low-level English speakers with Ukrainian-speaking EAs, referrals to entrepreneurship training, and signposting to childcare, housing, or mental health services[47]. For highly skilled professionals, a specialised Team Manager provided overarching support to navigate issues such as professional body validation and understanding UK professional pathways[48]. These elements reflected a recognition that employment outcomes for displaced populations are contingent on broader integration factors.

Enablers, challenges, and adaptations to employment support

The delivery of employment support in STEP Ukraine combined several strengths with operational tensions, particularly between personalisation, scalability, and participant expectations. The programme demonstrated a high degree of adaptive capacity, responding iteratively to participant needs and systemic pressures, while highlighting the importance of clear design choices and aligned delivery mechanisms for future initiatives.

A defining strength of STEP Ukraine was its highly personalised model of one-to-one support. EAs conducted detailed initial assessments, developing tailored action plans that addressed participants' career aspirations, skills, and barriers to labour market participation. This level of support was described by some managers as "rare to have" and "a luxury," given the rarity of such bespoke provision in mainstream employment programmes[49]. Crucially, EAs adopted a forward-looking approach, equipping participants with tools and confidence to continue progressing independently post-programme:

"We create an action plan from the very beginning for the programme, for 12 weeks. And after that we created an exit plan, 12 months action plan, for next year for our clients … We teach our clients to do that without [the EA's] support in the future." – Team Manager[50]

The recruitment of Ukrainian-speaking EAs, including some with lived experience as visa holders, was another critical enabler. Language matching removed linguistic barriers and enhanced trust, while also allowing for culturally sensitive guidance. Having an EA with lived experience aimed to create among participants a sense of empathy and legitimacy that extended beyond employment advice to wider forms of pastoral support[51]. Several managers described this as a "game changer" in fostering engagement[52].

Strategic external partnerships further strengthened delivery. BITC organised a series of online seminars and workshops delivered by representatives from diverse industries, providing participants with insights into recruitment processes, workplace

[47] PM1, TM2, EA1, EA3, EA4 (see Table 2.2).
[48] PM3, TM2 (see Table 2.2).
[49] SL1, SL4, PM3 (see Table 2.2).
[50] TM2 (see Table 2.2).
[51] EA3, EA4, EA5 (see Table 2.2).
[52] PM1, TM2 (see Table 2.2).

expectations, and sector-specific knowledge[53]. They also ran a coaching programme that matched participants with mentors from their desired industries[54]. As one Team Manager noted:

"if we've got a client who's a financial advisor, an EA may know about employment but have no idea of how to advise someone … So, we'd match them with another financial advisor or someone who's in a similar kind of field to just mentor them." – Team Manager[55]

Similarly, TERN provided a specialised self-employment programme designed to equip participants with the knowledge and resources needed to launch their own businesses. The programme offered workshops covering essential aspects of entrepreneurship, such as how to start a business and secure loans to address financial challenges, as well as one-to-one sessions with business coaches[56]. These collaborations addressed gaps in EA expertise, such as industry-specific knowledge, and offered participants access to professional networks and training opportunities that would otherwise be out of reach[57].

Another enabler was the structured training and support framework provided to EAs. Survey data indicate that the majority of EAs found key components of this framework highly valuable: 20 out of 26 respondents (77%) rated the two-week induction as "very helpful," 19 out of 26 (73%) highlighted weekly team meetings as crucial for peer learning and problem-solving, and 18 out of 26 (69%) emphasised the importance of specialist support from managers for navigating complex cases (see Figure 2.4). Complementary mechanisms - such as intranet resources (17/26; 65%) and refresher inductions (15/26; 58%) - further supported ongoing professional development, while monthly programme-wide meetings (14/26; 54%) provided a forum for aligning practice across regions. The distribution of perceived usefulness across these mechanisms suggests that frequent, targeted, and context-specific support had the greatest impact on EA performance, whereas broader, less frequent interventions were comparatively less influential. Collectively, this multi-layered support infrastructure likely contributed to consistent delivery across a geographically dispersed team.

[53] TM1 (see Table 2.2).

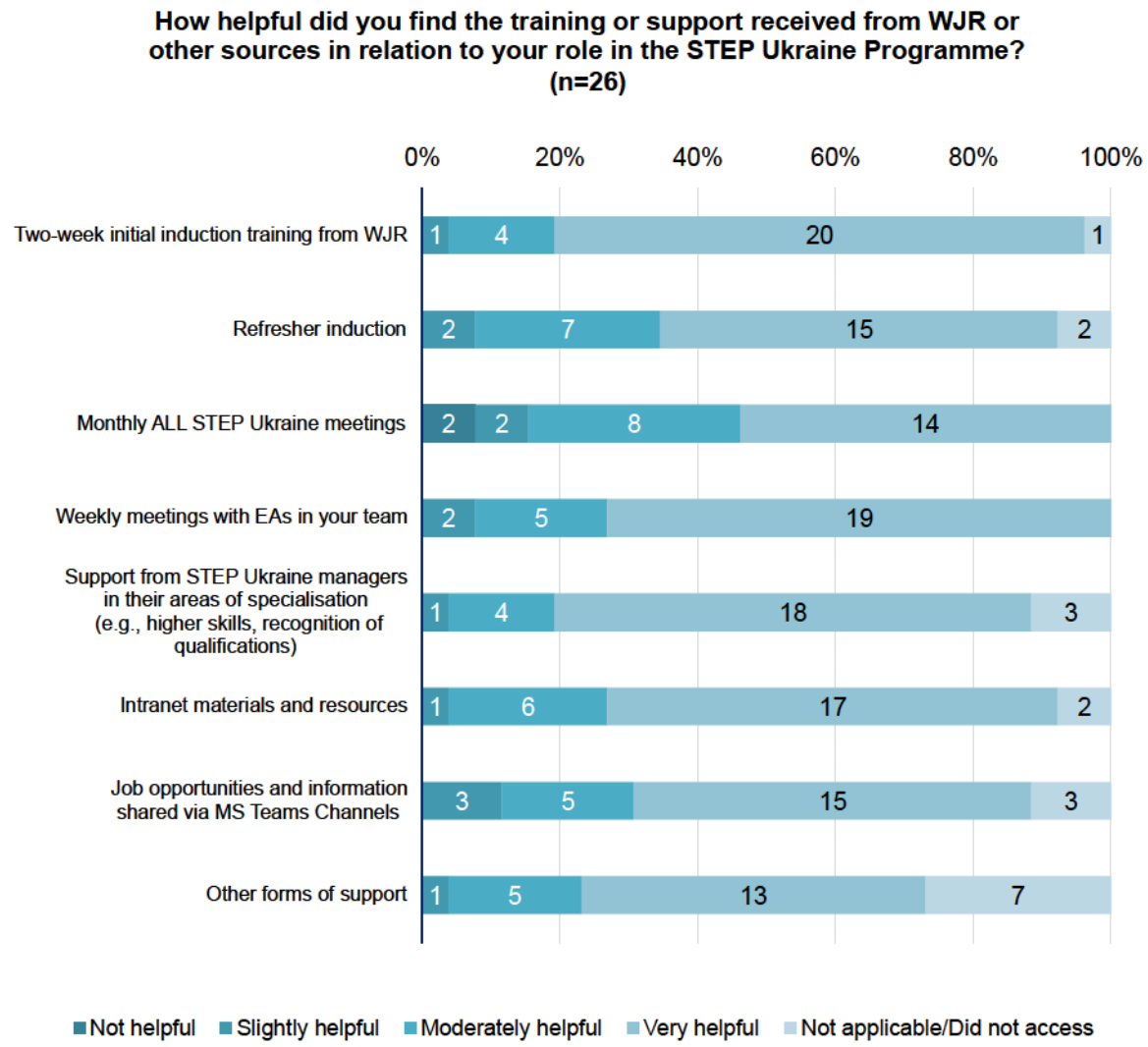
[54] PM1, TM1, EA2 (see Table 2.2).

[55] TM1 (see Table 2.2).

[56] PM1, TM1 (see Table 2.2).

[57] PM1, TM1 (see Table 2.2).

Figure 2.4 Helpfulness of the training or support received from WJR



Source: RAND Europe, Survey with Employment Advisers

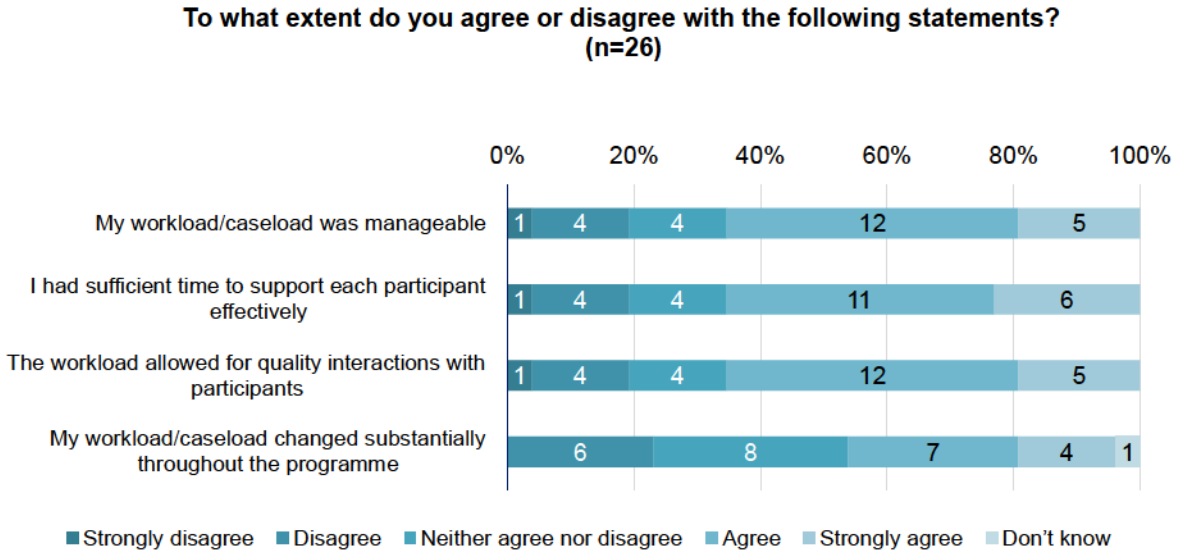
Practical enablers included the direct provision of IT equipment, such as SIM cards and laptop loans, which mitigated digital exclusion and ensured participants could access online resources and communication channels. Internal support structures (including regular team meetings, shared resource platforms, and the introduction of specialised Team Managers for different client needs) fostered collaboration, enhanced EA confidence, and enabled tailored guidance on complex career transitions.

Despite these strengths, several systemic and operational barriers constrained delivery. The most significant was the high caseloads borne by EAs, which rose from 25 to 35-36 clients. This increase reflected both the programme's success in reducing initial dropout rates, primarily due to improved onboarding, and sustained oversubscription[58]. Survey evidence suggested a mixed picture: 17 out of 26 EAs (65%) reported that their workload was manageable, that they had sufficient time to

[58] PM3, TM2 (see Table 2.2).

support participants effectively, and that they could maintain quality interactions (see Figure 2.5). Yet qualitative interviews with EAs and Programme Managers indicated that elevated caseloads risked diluting personalised support, particularly during more intensive phases of the programme, while also contributing to staff burnout and limiting opportunities for collaboration with other EAs[59].

Figure 2.5 Perceptions of caseloads among employment advisors



Source: RAND Europe, Survey with Employment Advisers

The introduction of group workshops was an adaptive response, aiming to reconcile efficiency with peer support by fostering collective learning (see Box 2.4). Yet the evidence on effectiveness was ambivalent. As reported by EAs, attendance was strongest when workshops aligned with participant priorities[60], but some EAs felt they displaced rather than complemented one-to-one engagement[61]. This raises questions about whether group-based provision should be seen as a sustainable enhancement to the delivery model or as a stopgap measure to compensate for capacity pressures. The introduction of centrally designed workshops highlighted an additional tension in the programme's design: while they sought to standardise quality, foster community, and relieve pressure on overburdened EAs, their optional nature and daytime scheduling limited reach, particularly for participants with low English proficiency or competing responsibilities. Together with employer seminars, which EAs described as facing inconsistent attendance, these group activities reflected the challenge of embedding additional commitments within already stretched participant schedules.

Another recurring challenge, as reported by EAs and Programme Managers, was the perceived mismatch between participant expectations and the programme's design. Staff reflected that some Ukrainian participants appeared to register with the hope that the programme would act as a recruitment agency, directly providing job

[59] EA4 (see Table 2.2).

[60] EA2, EA4, EA5, EA8 (see Table 2.2).

[61] EA1, EA7 (see Table 2.2).

placements, and were reportedly disappointed when this was not the case[62]. High-skilled professionals were seen to have "unrealistic" expectations of immediately resuming their previous professions in the UK, without fully appreciating the hurdles of qualification recognition or local labour market realities[63]. In addition, staff noted instances where participants sought to "cherry-pick" certain elements, such as employment or English support, rather than engaging with the integrated approach STEP Ukraine offered[64].

Such misaligned expectations, as noted by staff, often led to frustration or disengagement among participants. To address this issue, the programme enhanced its onboarding and assessment procedures (see Section 2.2), aiming to ensure that those who enrolled were genuinely committed and had the capacity to participate fully in all aspects of the programme. EAs received training to conduct thorough suitability assessments, helping to confirm that participants were willing and able to engage with both employment and English language support, and possessed the necessary time and motivation to benefit from the programme's integrated approach.

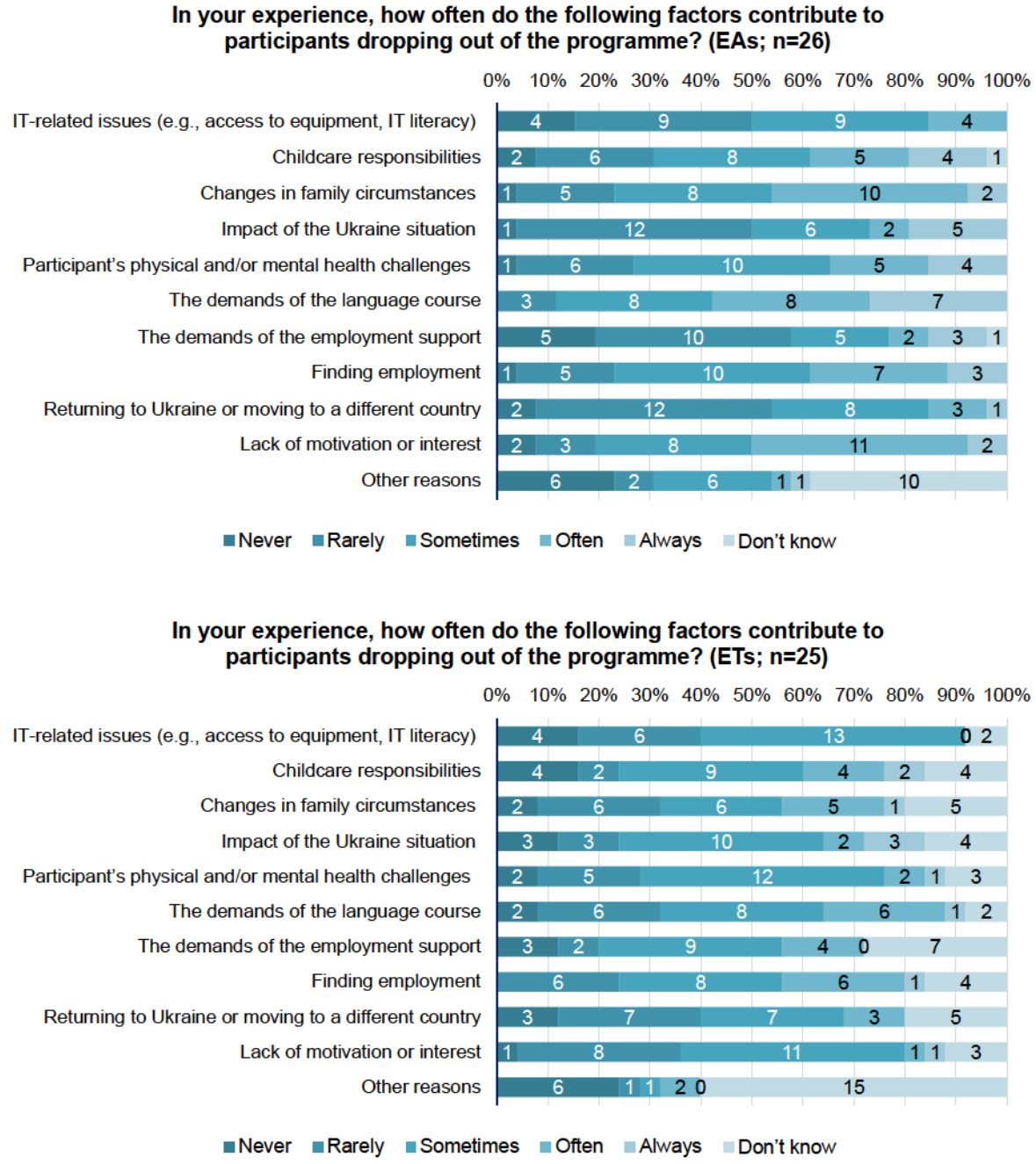
Furthermore, programme staff considered that participant disengagement was influenced by a combination of programme-specific demands, personal circumstances, and broader contextual factors, highlighting the complex interplay between participant needs and programme design (see Figure 2.6). Among 26 EAs who responded to the survey, 15 reported that the demands of the language course "often" or "always" contributed to dropout (58%), making it the most frequently cited factor, followed by 13 citing lack of motivation or interest (50%) and 12 noting changes in family circumstances (46%), while 10 highlighted participants leaving after finding employment (38%). Among 25 English tutors surveyed, the language course demands and employment attainment were both cited by 7 respondents (28%) as "often" or "always" contributing to dropout, with childcare and changes in family circumstances each noted by 6 (24%) and the Ukraine situation by 5 (20%). Interestingly, the fact that employment attainment during the programme could contribute to early exits indicates that while STEP Ukraine could facilitate labour market outcomes, its design was less suited to sustaining engagement post-placement and pointing to a tension between achieving short-term employment goals and maintaining participation.

[62] SL4 (see Table 2.2).

[63] PM1, PM2, ET8 (see Table 2.2).

[64] EA7, EA8 (see Table 2.2).

Figure 2.6 Implementers' perceptions of factors contributing to dropouts



Source: RAND Europe, Survey with Employment Advisers

EAs were acutely aware of how the ongoing conflict in Ukraine affected participants' emotional wellbeing and placed strong emphasis on adopting an empathetic and open-minded approach[65]. Building trust and rapport was recognised as essential, particularly when working with participants facing vulnerability or distress, and training was provided to equip EAs with tools and strategies for a supportive coaching environment. While these practices reflected sensitivity to participants' circumstances, the programme would have benefited from implementing a

[65] EA3 (see Table 2.2).

comprehensive trauma-informed approach. Established frameworks, such as SAMHSA, suggest that such an approach involves embedding specific principles – such as safety, trust, empowerment, and attention to cultural considerations – across policies, procedures, and programme practices (see Box 2.5 further below)[66].

Structural challenges persisted beyond programme completion. Formal post-programme support was limited, leaving participants without sustained guidance despite ongoing barriers to meaningful employment. Recognition of Ukrainian qualifications remained a recurrent obstacle. While EAs signposted participants to the UK National Information Centre (UK ENIC), specialist expertise for supporting high-skilled professions (such as law or medicine) was limited, constraining effective navigation of complex accreditation systems[67].

Box 2.4 Case study: Adopting EA-led group workshops to manage caseloads

Original challenge:

As the programme matured, EAs faced rising caseloads, which increased from around 25 to as high as 35-36 participants each. This constrained their capacity to provide tailored one-to-one support, while also limiting time for research, preparation, and proactive job-matching. Foundational employment topics, such as CV writing and job search strategies, were repeatedly covered in individual sessions, creating inefficiencies and placing additional strain on staff.

Adaptation implemented:

To address this, WJR and MHCLG jointly decided to incorporate EA-led group workshops into the programme. These workshops were designed to deliver general employment-related content to multiple participants simultaneously, alleviating pressure on individual case management. Initially, six centrally designed workshops ensured consistency and quality, covering topics such as CV writing, interview preparation, LinkedIn profiles, confidence and wellbeing, and job search strategies. Over time, more specialised sessions were added, including IELTS preparation and the use of AI in job searching.

[66] Substance Abuse and Mental Health Services Administration. SAMHSA's Concept of Trauma and Guidance for a Trauma-Informed Approach. HHS Publication No.(SMA) 14-4884. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2014.

[67] EA6, SL4 (see Table 2.2).

Rationale and benefits:

The workshops were introduced for three main reasons:

- Caseload management: By covering foundational material in groups, EAs were able to reserve one-to-one sessions for more complex, tailored support.
- Efficiency and consistency: Centralised design reduced duplication across EAs and ensured all participants received a minimum standard of guidance.
- Community building: Sessions aimed to foster a sense of peer support and collective motivation, echoing the positive social dynamics already established through English classes.

Impact:

Feedback on the workshops was mixed among EAs. While the workshops helped ease caseload pressures, some EAs chose to continue providing only one-to-one sessions. Attendance was high when workshop topics matched participants' needs, with some sessions described as "extremely effective" in boosting confidence and engagement. For EAs, the group format was seen as particularly valuable for participants who were further from the labour market or lacked clear career goals, offering both social support and practical guidance.

English language training delivery

The English language component of STEP Ukraine was designed as an intensive, structured intervention aimed at equipping participants with the language skills necessary to navigate the UK labour market. The course ran for ten weeks, with live, teacher-led classes delivered five days per week, each lasting 2.5 hours including a ten-minute break, alongside a recommended two hours of daily homework. This translated to a minimum of 175 hours of guided learning, including approximately 125 hours of live instruction[68].

The curriculum focused on "English for employability, English for work, and Social English." It was structured across proficiency levels ranging from CEFR levels A0/A1 to C1, distributing participants according to their initial assessment results. This training aimed to equip participants with practical language skills for professional and social contexts[69]. Lower-level classes focused on foundational skills, including basic workplace communication, practical writing, and everyday interactions, such as shopping, travel, and navigating social services. Higher-level classes emphasised professional mastery, including advanced writing, conflict resolution, networking, and leadership-oriented topics[70]. Sector-specific language instruction and thematic modules on employability, professional communication, and social English were incorporated to bridge the gap between general language competence and labour market applicability.

[68] British Council, background information document shared with the evaluation team.

[69] PM2 (see Table 3), British Council information pack.

[70] British Council, background information document shared with the evaluation team.

Pedagogically, the course employed the British Council's flipped-learning model. Learning was structured into three phases: preparatory online exercises to introduce grammar and vocabulary, interactive live sessions via Zoom focused on communication and practice, and follow-up materials that reinforced learning. In-class lessons followed a consistent structure, beginning with an opening activity to build confidence and rapport, followed by targeted grammar and vocabulary work, and concluding with applied tasks and feedback. This allowed live sessions to emphasise communication, interaction, and collaborative tasks in breakout rooms[71].

One English tutor interviewed described the class structure as beginning with a conversational activity to build rapport, moving into focused grammar and vocabulary linked to the homework, and then applying these skills through tasks and group feedback: "*Each stage engaged different parts of the brain… The first part was free, just trying to speak with humour and rapport, then came focused grammar and vocabulary, and finally tasks that combined both, which was effective*"[72]. Tutors were also given some flexibility to adapt lessons to participants' needs, particularly for lower-level learners or where sensitive topics arose, and to balance standardised content with responsiveness to emerging class dynamics[73].

Additional supports, such as webinars on grammar, pronunciation, online learning strategies, and time management, sought to maximise engagement and equip participants to manage the independent learning component effectively. This approach underscored the programme's recognition that language acquisition is not solely classroom-bound and that learner autonomy is critical for sustained skill development.

Assessment and progress monitoring were embedded into the course design. The 'EnglishScore' Core Skills Test, a mobile-based app with results mapped to the Common European Framework of Reference for Languages (CEFR) levels, was used pre-course, mid-course, and post-course to provide quantitative measures of proficiency. Alongside this, teacher-led qualitative assessments evaluating task completion, fluency, vocabulary, grammar, and pronunciation were also provided. Mid- and end-course progress reviews allowed participants to receive narrative feedback and targeted guidance, reinforcing motivation and enabling tutors to adjust support dynamically. Adjustments to assessment timing, such as conducting the post-course test during the final class, addressed practical barriers to completion and ensured that participants received timely validation of their learning. The final assessment led to the award of an English language certificate indicating the level achieved.

During onboarding, WJR offered students their choice of three class times (morning, afternoon, or evening) to accommodate different schedules. Changes to class time or proficiency level were generally allowed only within the first week and depended on teacher approval and availability[74]. In the extension cohorts, class sizes were intentionally kept small to ensure tutors could provide more personalised support to each student[75].

[71] PM2, ET6 (see Table 2.2); British Council, background information document shared with the evaluation team.

[72] ET6 (see Table 2.2).

[73] ET1, ET2, ET3, ET4, ET5, ET6, ET8, PM2 (see Table 2.2).

[74] EA6, ET5, ET8 (see Table 2.2).

[75] PM2; ET4 (see Table 2.2).

Enablers, challenges, and adaptations to English language training

Both English tutors and programme managers praised the flipped-learning approach for its effectiveness in online settings, especially given the participants' generally high levels of education and digital literacy[76]. Preparatory online exercises aimed to provide the foundation for lessons, while live sessions focused on communication and practice to consolidate learning. In these sessions, tutors facilitated interactive activities such as discussions, role-plays, and pair or group work in breakout rooms[77]. After class, participants were given access to review materials, including annotated teaching slides, which tutors noted supported reflection and reinforced key learning points.

The curriculum was explicitly employability-focused, covering practical skills such as career development, networking, and professional communication that aligned with the programme goal of improving labour market integration[78]. Supplementary resources (including bite-sized videos, webinars, and time management training) allowed learners to consolidate their skills outside of class. Course materials were highly standardised to ensure consistency across groups. However, some tutors, at their own discretion, incorporated cultural elements, such as British humour and customs, to make lessons more engaging and relatable[79].

To support continued language learning after the programme, participants were provided access to the British Council's English Online Self-Study course as part of the Council's social value commitment[80]. This included the course platform and resources such as the "Extra Study" section, which featured bite-sized videos on practical English usage. Access was initially intended to last up to three years after completing or passing the final test, enabling participants to revisit exercises and consolidate their learning over time. However, due to technical limitations and the need to manage platform capacity for new cohorts, this was later reduced to six months[81]. Access was also conditional on participants completing pre-class exercises and actively engaging in lessons. While the resource created opportunities for sustained language development, its effectiveness ultimately depended on participants' motivation and continued effort.

Tutors and Programme Managers also highlighted the importance of the sense of community fostered in online classes. They observed that group learning provided social and emotional support during a difficult time, counteracting isolation and helping participants build confidence[82]. It should be noted, however, that this assessment reflects the perspectives of staff rather than direct evidence from learners themselves.

---

[76] PM2, ET4 (see Table 2.2).

[77] ET7, ET4, ET8 (see Table 2.2).

[78] PM2, ET4 (see Table 2.2); British Council, background information document shared with the evaluation team.

[79] ET5 (see Table 2.2).

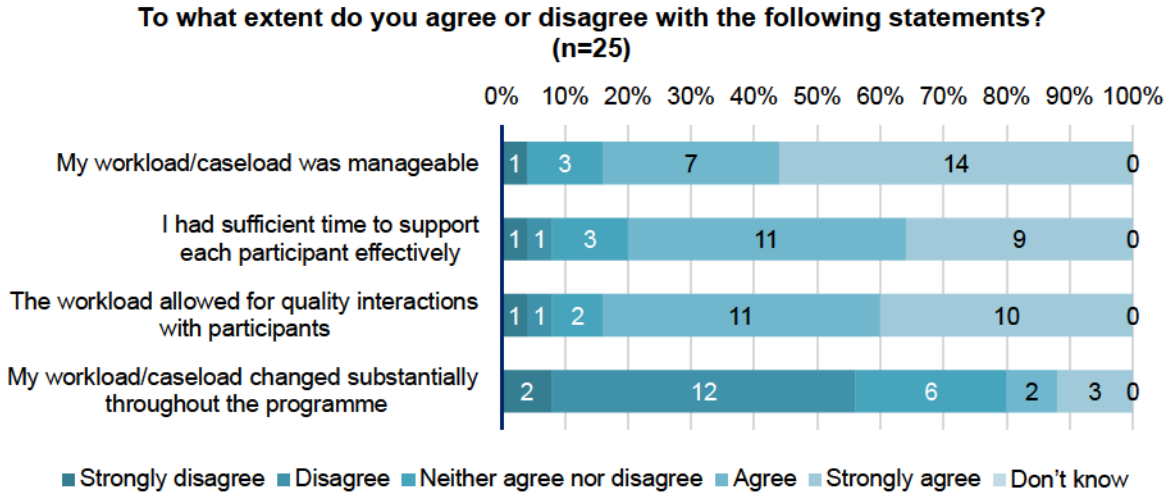
[80] PM2 (see Table 2.2), British Council information pack.

[81] ET6, ET7 (see Table 2.2).

[82] PM2, PM3, ET1, ET3, ET5, ET6, ET7, ET8, EA1 (see Table 2.2).

English tutors generally reported that their workload and caseload were manageable, supporting the delivery of high-quality learning experiences. For instance, 21 tutors among a total of 25 who responded to the survey (84%) agreed or strongly agreed that their workload was manageable, while 20 out of 25 (80%) agreed that they had sufficient time to support each participant effectively (see Figure 2.7). Similarly, 21 out of 25 tutors (84%) felt that their workload allowed for meaningful interactions with participants, suggesting that staffing levels and course design supported both individual attention and broader group facilitation. Nonetheless, interview feedback indicated that the course's intensity could be fatiguing, with two tutors recommending longer breaks to balance participant engagement with staff wellbeing[83]. These findings suggest that while the workload structure largely enabled effective teaching and personalised support, there is scope for minor adjustments to improve sustainability and reduce staff fatigue.

Figure 2.7 Perceptions of workload among English tutors



Source: RAND Europe, Survey with Employment Advisers

When asked about the usefulness of the training and support provided by the British Council and other sources, 23 tutors among a total of 25 who responded to the survey (92%) rated it as helpful or extremely helpful. One tutor did not respond, and one described the support as "mediocre" in an open-text comment[84]. Thirteen tutors (52%) indicated that no additional training or support was needed, reflecting overall satisfaction with the provision[85]. Among the 12 tutors who identified further needs, the most commonly cited areas were training related to trauma and conflict, including PTSD awareness – which was not seen as a required part of the contracted training as ONS did not indicate any potential wellbeing issues or concerns – and working with students affected by war (3 tutors), additional cultural awareness and support tailored to Ukrainian learners (3 tutors), and guidance on digital tools and platforms, including training on AI (3 tutors). Other suggestions included clearer exam preparation guidance (2 tutors) and improved orientation or onboarding materials in

[83] ET6, ET7 (see Table 2.2).

[84] RAND Survey with English tutors, survey question 8.

[85] RAND Survey with English tutors, survey question 9.

Ukrainian (1 tutor). Finally, one survey response reinforced earlier interview findings, noting that "*it would have been helpful to meet with some fellow teachers via video call, to discuss issues we had had*," highlighting the potential value of peer exchange and collaboration[86].

As in the case of EAs, English tutors also received training that introduced approaches for engaging with individuals who may have experienced distress, which provided them with practical strategies to create a safe and welcoming learning environment[87]. They were also encouraged to slightly adjust lesson content and classroom activities when necessary to avoid potentially distressing topics[88]. Examples of such adaptations included changing questions about family, which were described as "quite triggering for people who aren't with their family"[89], or altering verb tenses and imagery to avoid sensitive past experiences and disturbing reactions[90]. One English tutor mentioned adapting lessons "on many occasions" to ensure students felt understood and safe[91].

Despite these strengths, several challenges limited the accessibility and effectiveness of the English language component. Foremost was the intensity of provision. While the rigorous schedule allowed for rapid skill development, English tutors and EAs reported that many participants struggled to balance classes with employment, caring responsibilities, or the stressors of displacement. Stakeholders suggested that less intensive options (for example, two to four classes per week, or an extended programme duration) might have improved accessibility without compromising outcomes.

The appropriateness of materials was another recurring concern. While tutors generally valued the flipped model, they noted gaps in provision for lower-level learners, with insufficient focus on grammar and content sometimes seen as repetitive or overly geared towards high-level business scenarios[92]. Tutors reported that the course was particularly challenging for complete beginners (A0 level), who often struggled to keep pace, which could lead to demotivation and early withdrawal[93]. Similarly, content such as presentation skills was judged inappropriate for low-level learners, being perceived as irrelevant or unnecessarily difficult for their immediate needs[94]. One tutor observed that materials were "heavily geared to the British market," requiring further adjustments to better match participants' needs and stage of integration[95].

Some EAs and English tutors observed that grammatical instruction within the curriculum was lacking, prompting tutors to supplement lessons informally with their

---

[86] RAND Survey with English tutors, survey question 7.

[87] PM2, ET3, ET4, ET8 (see Table 2.2).

[88] PM2, ET8 (see Table 2.2).

[89] ET5 (see Table 2.2).

[90] ET8 (see Table 2.2).

[91] ET5 (see Table 2.2).

[92] ET6, ET7 (see Table 2.2).

[93] EA3, EA4, EA5, EA7, ET1, ET4, ET5, ET6, ET8 (see Table 2.2).

[94] ET3, ET4, ET5, ET6, EA3 (see Table 2.2).

[95] ET3 (see Table 2.2).

own resources[96]. Over time, the British Council responded by allowing greater flexibility for tutors to adapt the curriculum to better suit the needs of each group. Tutors incorporated additional scaffolding and introduced more grammar content for lower-level learners. New support mechanisms, such as dedicated grammar and pronunciation webinars, time management workshops, and test familiarisation sessions, were also implemented to help participants handle the course's demands and alleviate assessment-related anxiety.

Another factor that warrants consideration relates to the use of the 'EnglishScore' test and the potential for multiple test attempts. Participants were encouraged at different stages of the course to take the test in order to familiarise themselves with the platform and the types of questions. Staff emphasised that their role was to help learners approach the test in conditions that would reflect their "true performance" rather than be distorted by technical or procedural unfamiliarity. As one Programme Manager explained, *"when it comes to any test performance, what we want is the test to capture the true performance of the student… what we don't want is externalities like them being thrown off because of the format of the test or the technical apparatus of the test"*[97]. Programme Managers highlighted that participants were subject to the platform's standard rules, which limited them to two test attempts per day and refreshed the item bank to avoid duplication. Although these restrictions safeguarded validity to some extent, the centrality of test improvements as a programme KPI may have created unintended incentives to promote retakes.

Additionally, while the mobile-based 'EnglishScore' test could be a useful tool for quantitatively tracking learner progress, Programme Managers and English tutors noted that it did not always capture the full range of learning outcomes, particularly for higher-level learners[98]. Gains in areas such as confidence, fluency, and vocabulary development were sometimes underrepresented in test results. To address this limitation, the programme introduced teacher-led progress reviews at both mid-course and end-of-course stages[99]. These reviews provided narrative and graded feedback on key dimensions of learning (including task completion, grammar, vocabulary, fluency, and pronunciation), which allowed for a more contextualised and comprehensive assessment of learner progress. As one Programme Manager explained, while the 'EnglishScore' test was required, "*we know that's not always the best way to capture the impact of a programme… progress reviews allowed teachers who knew the students to test them in context and provide a more rounded review*," which was not only more accurate but also motivating for participants[100]. These adaptations helped balance the requirements of standardised testing with a richer, practice-based understanding of learner achievement.

Beyond curriculum and assessment issues, a key factor affecting the effectiveness of English language training was the extent to which the programme could accommodate the psychological and emotional circumstances of participants. English tutors consistently highlighted that the psychological impact of war and

[96] EA1, EA2, EA3, ET5, ET7 (see Table 2.2).

[97] PM2 (see Table 2.2).

[98] PM1, PM2, ET2, ET4, ET6, ET7 (see Table 2.2).

[99] PM2 (see Table 2.2), British Council information pack.

[100] PM2 (see Table 2.2).

displacement constrained participants' ability to engage fully with the programme, particularly the intensive English language training[101]. Some tutors and EAs observed that trauma could manifest in difficulties with focus, attendance, and readiness to pursue integration, with some participants experiencing depression and struggling with the idea of creating a life in the UK without their families[102].

Programme managers and implementers attempted to mitigate this through flexible course design, careful topic selection, reviewing course content to avoid triggering material, shifting discussions towards forward-looking topics, and signposting to specialist services[103]. However, several EAs and tutors expressed that more could have been done to address mental health needs directly. Some felt there were gaps in training and support for dealing with complex emotional and psychological issues and would have welcomed further capacity-building in this area[104]. Additionally, the evidence suggests that without further flexibility in attendance requirements and stronger, embedded mechanisms for mental health and trauma support, structural and psychological barriers are likely to continue limiting the effectiveness of language training[105]. This points to the importance of integrating specialist provision or formal partnerships with trauma-focused organisations, so that language acquisition efforts are not undermined by unaddressed wellbeing needs (see Box 2.5).

Box 2.5 The Importance of a trauma-informed approach for STEP Ukraine participants

Rationale:

Ukrainian refugees participating in STEP Ukraine face complex, overlapping challenges affecting their mental health, social integration, and learning outcomes. Trauma from conflict, displacement, and ongoing concerns for loved ones can impair memory, concentration, and motivation, reducing the effectiveness of English language learning and employment support[106]. Mental health stigma, cultural differences in understanding wellbeing, and barriers to accessing healthcare exacerbate these challenges[107]. Despite these findings, it is important to note that the original design did not incorporate trauma-informed approaches, as these were not raised as a key concern during the ONS survey. This discrepancy may highlight the differences in mental health concerns that are self-reported and those that are observed by teachers and advisers.

---

[101] PM2, PM3, TM2, EA3, EA5, ET1, ET3, ET4, ET5, ET6, ET8 (see Table 2.2).

[102] EA5, EA8, TM2, ET3, ET4, EA5, EA8 (see Table 2.2).

[103] PM2, ET4, ET6, ET8, TM1, TM2, EA2, EA5 (see Table 2.2).

[104] EA3, EA8, ET5 (see Table 2.2).

[105] Kim, S. H. O., Ehrich, J., & Ficorilli, L. (2012). Perceptions of settlement well-being, language proficiency, and employment: An investigation of immigrant adult language learners in Australia. International Journal of Intercultural Relations, 36(1), 41-52.

[106] Kaufman, K. R., Bhui, K., Katona, C. (2022). Mental health responses in countries hosting refugees from Ukraine. BJPsych Open. 2022 Apr 1;8(3):1-10. doi: 10.1192/bjo.2022.55. Epub ahead of print. PMID: 35361299; PMCID: PMC9059731.

[107] Kienzler, H. (2024) Community integration, quality of life, thriving, and mental health among refugees and asylum seekers. A London service provider perspective. Front. Public Health 12:1358250. doi: 10.3389/fpubh.2024.1358250.

Kyiv International Institute of Sociology (2018) Mental health in Donetsk and Luhansk oblasts. https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2024.1358250/full, accessed 31 March 2026.

Evidence of impact:

- Social factors, including language proficiency, employment, and accommodation satisfaction, are modifiable determinants of refugee wellbeing[108].
- Rapid labour market entry without sufficient language skills can undermine confidence and health, while extensive language and integration courses may temporarily reduce employment rates due to "lock-in effects"[109].
- Forced migrants benefit from structured learning environments that provide routine, social interaction, and opportunities to express emotions, which contribute positively to mental health and social integration[110].

Principles and recommendations:

A robust trauma-informed approach recognises trauma's prevalence, identifies its signs, and integrates this understanding into all programme policies and practices to prevent re-traumatisation[111]. Core principles include:

1. Safety – creating predictable and supportive learning and employment environments.
2. Trustworthiness and Transparency – ensuring clear communication and reliable support.
3. Peer Support – fostering social connections and shared learning.
4. Collaboration and Mutuality – engaging participants in co-designing interventions.
5. Empowerment, Voice, and Choice – respecting autonomy and cultural perspectives.
6. Cultural, Historical, and Gender Sensitivity – tailoring support to refugee-specific experiences and norms.

Expected outcomes:

- Enhanced engagement in English language classes and employment support.
- Improved social and emotional wellbeing.
- Greater resilience and motivation for labour market integration.
- Reduced risk of re-traumatisation and isolation.

---

[108] Campbell, M. R., Mann, K. D., Moffatt, S., Dave, M., & Pearce, M. S. (2018). Social determinants of emotional well-being in new refugees in the UK. Public Health, 164, 72-81.

[109] Kosyakova, Y., Gatskova, K., Koch, T., Adunts, D., Braunfels, J., Goßner, L., ... & Vandenhirtz, M. (2024). Labour market integration of Ukrainian refugees: An international perspective (No. 16/2024en). IAB-Forschungsbericht.

Morrice, L., Tip, L. K., Collyer, M., & Brown, R. (2021). 'You can't have a good integration when you don't have a good communication': English-language learning among resettled refugees in England. Journal of Refugee Studies, 34(1), 681-699.

[110] Savychenko O, and Portnytska N (2018) Pilot study of stigmatization of mental health problems in the Ukrainian educational environment. MHGC Proceedings, 80-82.

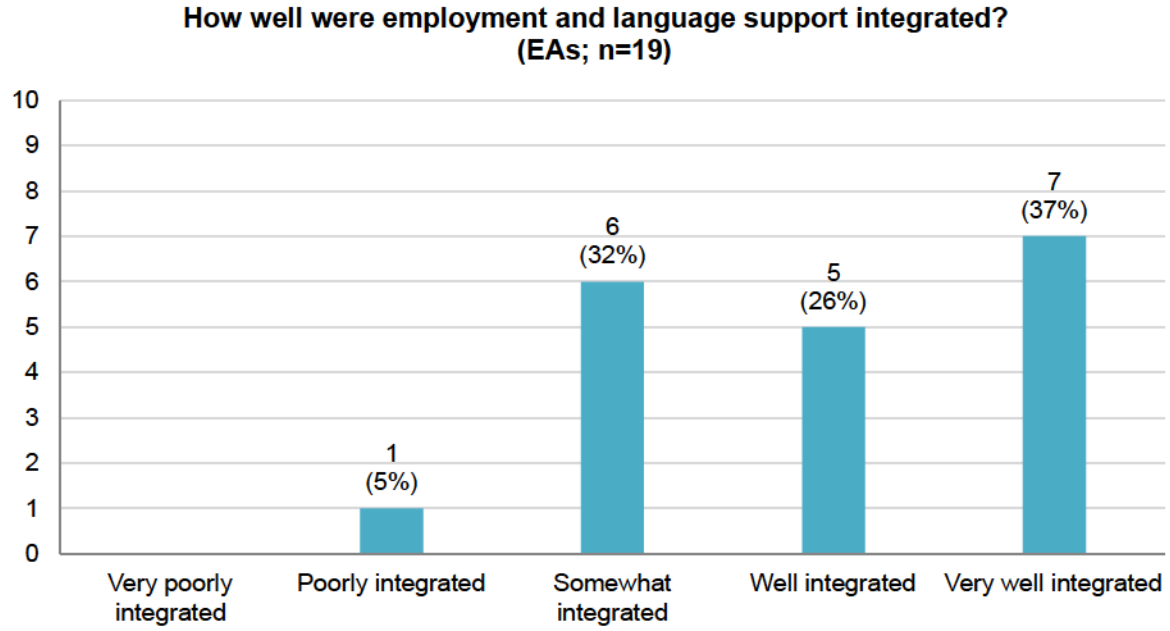
[111] Substance Abuse and Mental Health Services Administration. SAMHSA's Concept of Trauma and Guidance for a Trauma-Informed Approach. HHS Publication No.(SMA) 14-4884. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2014.

Towards an integrated delivery model

The STEP Ukraine programme was designed on the premise that English language training and employment support would operate as mutually reinforcing interventions, with integration expected to improve participants' employment prospects. In practice, however, the level of integration was uneven, and the evidence suggests that structural barriers and weak communication channels limited the extent to which the two components functioned as a coherent whole.

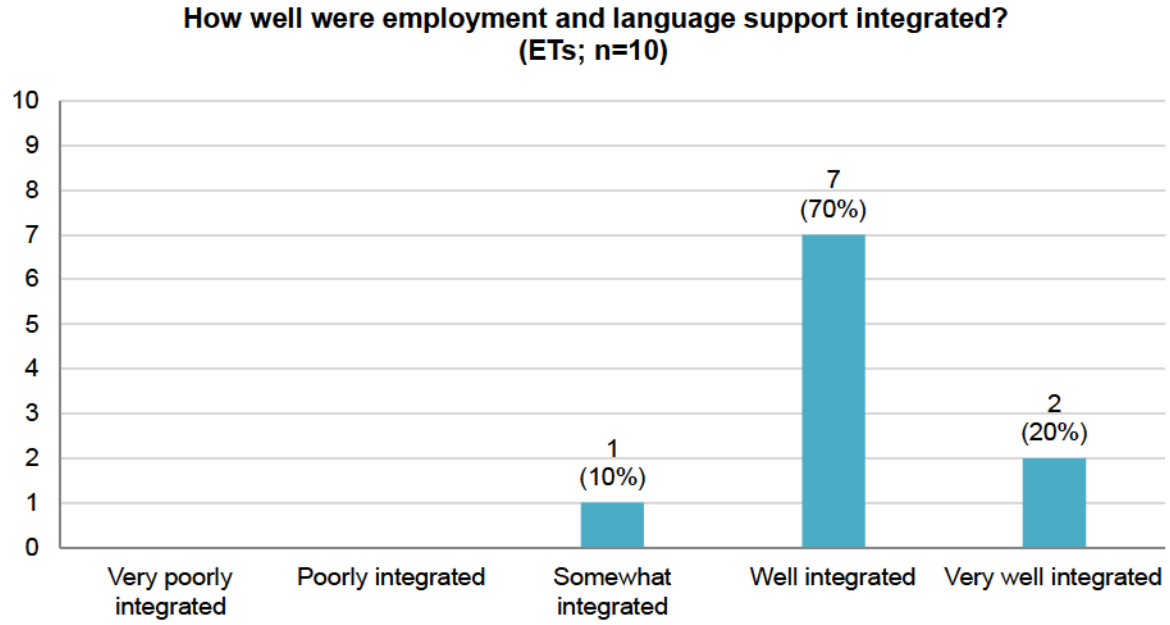
Survey data point to a significant divergence in perceptions between front-line staff groups (see Figure 2.8). Among 26 EAs responding to the survey, 19 (73%) reported that they believed employment and language support were integrated[112]. By contrast, only 10 out of 25 English tutors (40%) shared this view[113]. The discrepancy raises questions about whether "integration" was understood differently across staff groups: for EAs, having access to attendance registers and occasional programme documentation from the British Council may have constituted integration, whereas for tutors, the lack of meaningful interaction with EAs reflected a siloed delivery model.

Figure 2.8 Perception of the integration of employment and language support



[112] 19 reported Yes, 5 No, and 2 "Don't know"

[113] 10 reported Yes, 2 No, and 13 "Don't know"



Source: RAND Europe, Survey with Employment Advisers and English tutors

The interviews reinforce this interpretation. One tutor stated, "*I really didn't know what the employment support was. Maybe we could have gotten some more background on it. I had no idea what they were being provided*"[114], while another described the two components as "very siloed"[115]. Some of the tutors interviewed found themselves ill-equipped to respond when participants mentioned their employment support in class[116]. EAs, meanwhile, were formally tasked with monitoring attendance and motivating engagement in English classes[117]. Several EAs, however, admitted to having little understanding of what the English classes entailed in practice, including the course delivery model, the online platform used, or teaching methods[118]. This gap undermined their ability to support learners effectively, particularly when participants raised practical or pedagogical questions about the classes.

Challenges in communication between EAs and the British Council further compounded this problem. Although some EAs reported timely and appropriate communication[119], others described the British Council as slow to respond or difficult to reach[120]. One EA noted that the British Council sometimes appeared to misunderstand her requests or send boilerplate responses to them, obliging the adviser to escalate relatively simple issues to the Team Manager[121]. At the outset, there was also confusion about whether EAs or programme assistants were

[114] ET3 (see Table 2.2).
[115] ET7 (see Table 2.2).
[116] ET3, ET4, ET7 (see Table 2.2).
[117] EA1 (see Table 2.2).
[118] EA2, EA7, EA8 (see Table 2.2).
[119] EA1, EA6, EA7 (see Table 2.2).
[120] EA4, EA7, EA8 (see Table 2.2).
[121] EA4 (see Table 2.2).

responsible for handling participants' requests regarding class changes[122]. While this was eventually clarified, the episode illustrates how weak communication protocols reduced the efficiency of integration and placed unnecessary burdens on EAs.

Despite this limited collaboration between English tutors and EAs, at the curriculum level, there was stronger evidence of alignment. Both components of the programme emphasised employability skills, with English classes foregrounding business English and professional communication, while EAs reinforced language acquisition by arranging mock interviews and encouraging practical application in job search activities. When asked whether integration between the two components had helped improve job placement rates, 19 of 26 EAs (73%)[123] and 2 of 25 English tutors (48%)[124] agreed or strongly agreed. Some EAs interviewed also noted that being informed of the weekly language topics would have allowed them to better connect classroom content with their one-to-one sessions, an indication that modest improvements in information-sharing could have enhanced cross-component synergy[125].

2.4 Participant experiences and outcomes

This section provides a detailed assessment of participant experiences and outcomes associated with STEP Ukraine, drawing on interviews with both programme participants (12) and non-participants (5), as well as insights from focus groups (11). The analysis begins by examining perspectives on the application and recruitment process. Subsequent sections offer a review of participants' perceptions of the programme's employment support and English language training components, followed by reflections from non-participants regarding their views of the intervention. The section concludes with key lessons learned and recommendations to inform the design and implementation of future programmes.

Application to the STEP Ukraine Programme

Evidence from interviews indicates that participants predominantly learned about STEP Ukraine through informal Ukrainian networks, including social media platforms such as Facebook and Telegram, as well as through newsletters from community groups and hosts under the Homes for Ukraine scheme. This reliance on peer networks and community intermediaries highlights both the programme's reach within existing social structures and potential limitations in accessibility for individuals less embedded in these networks.

Participants' motivations for applying were closely aligned with their personal employment and life objectives in the UK. Where participants had clear aims for employment in the UK, they generally knew what support they needed to achieve their goals, reflecting a strategic approach to navigating the UK support system and labour market. While this could be a combination of employment support and

[122] EA8 (see Table 2.2).

[123] RAND Survey with English tutors, survey question 11.

[124] RAND Survey with employment advisers, survey question 10.

[125] EA2 (see Table 2.2).

learning English, participants tended to prioritise one over the other. Participants who wanted to secure a job related to their career in Ukraine reported that they had specific employment support needs they wanted help with, such as successfully applying for a job. Conversely, participants with lower English proficiency tended to prioritise language development before pursuing work opportunities. As one participant explained,

"When I came to the UK I didn't have the required level of English language [to find a job according to my previous work experience] ... so I understand that I need to advance my English language skills and this is something that I work towards at the moment." (Male, 51-60 years old, intermediate level English, employed).

The application process itself was generally perceived as straightforward, and participants typically felt confident completing it online. However, some participants required support from friends and family to complete the application. Moreover, a small number of participants expressed dissatisfaction regarding communication and transparency about their application. The protracted period between application and confirmation, often lasting between one and two months, left some participants anxious and uncertain about their status. One participant described this as "*a little bit worrying for me*," noting, "*There was maybe about two months or so where I haven't heard anything about this programme and I was not sure whether I'm still registered or not or whether I got lost somewhere in the system*"[126]. Such lapses in communication risk eroding trust and may deter future applicants, particularly those already experiencing uncertainty and instability.

Perceptions on employment support

Most STEP Ukraine participants consulted valued EAs' efforts to translate their existing skills and experience into the UK labour market context. They highlighted the following activities as helping them to familiarise themselves with the UK employment landscape and coach them to successfully apply for jobs:

- Drafting or tailoring existing CVs and cover letters to ensure they meet UK employers' expectations.
- Signposting to recruitment websites.
- Support with completing job applications.
- Preparing for job interviews through advice, preparing question responses, and mock interviews.

A key benefit of employment support noted by participants was to understand recruitment processes in the UK and how to make their application stand out when applying for jobs. This support included EAs providing guidance on emphasising key words in CVs according to specific job specifications, tailoring applications to individual job specifications and ensuring that CVs, cover letters, and application forms were written according to UK employers' expectations. As one participant explained, "*I*

[126] Male, 41-50 years old, intermediate level English, employed.

obviously knew how to make a CV, but it was more for kind of Ukrainian context, but I didn't know how to do it in the UK context, so I would say that it was very useful"[127].

Participants consulted particularly appreciated EAs who took the time to understand their specific priorities and employment support needs, and who personalised their assistance accordingly. Mock interviews based on actual job opportunities that simulated live interview conditions, for example, helped participants build confidence in responding spontaneously in English. As one participant shared, "*It's difficult to think quickly in a language that isn't your native one... and it helps me to be more confident*" [128].

Participants also emphasised that the flexibility in scheduling meetings with their EAs was a key enabler of engagement with the programme. This adaptability allowed them to balance employment support with intensive English classes, as well as other employment and caring responsibilities, highlighting how logistical flexibility can mitigate participation barriers. The option to work with either English- or Ukrainian-speaking EAs further illustrates the importance of tailoring provision to individual language needs. Participants with lower English proficiency valued Ukrainian-speaking advisers for in-depth discussions about their employment plan, whereas those with higher proficiency often preferred English-speaking advisers to use the meetings as an opportunity to practise language skills.

Despite the considerable strengths of the STEP Ukraine programme, several challenges were identified regarding the alignment of provision with participants' needs. One key issue related to the concurrent scheduling of English language classes and employment support, which constrained participants' ability to benefit fully from both elements. Participants with lower levels of English proficiency, in particular, reported that it was difficult to set meaningful employment goals before their language skills had improved. As one participant observed, "*To get a better job, you need English language. So, it would be helpful to get support from a work coach at a later stage, when you're better at English language*" [129].

Similarly, participants with advanced English proficiency found that the combined workload of language classes and employment support limited the time they could dedicate to job search activities, often delaying their engagement with job applications until after the programme concluded. For these individuals, enhancing English proficiency was seen as a prerequisite for pursuing roles aligned with their aspirations, particularly where specific positions required a high level of language competency. Many participants only began focused job searching once they had completed the language component of the programme, by which time the employment support had also ended: "*I was really focused on language, and I didn't apply for any jobs at all during the programme. So, I started thinking about applying and all the employment stuff after the course*" [130].

---

[127] Male, 51-60 years old, intermediate English, employed.

[128] Female, 18-30 years old, advanced English, not in employment.

[129] Female, 18-30 years old, elementary English, not in employment.

[130] Female, 41-50 years old, advanced English, employed.

Some participants also reported that employment support did not adequately address sector-specific requirements and felt that support provided by EAs was too general for their situations. Limitations included having qualifications from Ukraine recognised in the UK or supporting participants to understand how to access roles in specific sectors: "*I have to prove my diploma with the Royal Institute of British Architects (RIBA) – I have to pass an exam. It's too hard for me because it's a very different background, I don't have enough experience in architecture*"[131].

Some participants also reported dissatisfaction with programme administration. Uncertainty around scheduling and logistical arrangements occasionally led to disengagement, with one participant noting that confusion over meeting links resulted in the breakdown of contact: *"[The Employment Adviser] sent me the [meeting] links and I cannot go for this link…maybe because of that, we stopped our conversations"*[132]. Participants further expressed concern that their long-term employment goals were overlooked when they were not currently active in the labour market due to childcare, education, or other commitments. These accounts suggest a possible tension between short-term operational pressures and the programme's broader strategic objectives and highlight the importance of reliable systems and proactive adviser engagement in supporting meaningful and sustainable employment outcomes.

Progress in employment outcomes was less immediate and visible compared to language acquisition. Participants highlighted that attaining roles aligned with their skills and career ambitions required not only English proficiency but also time, long-term guidance, and systemic understanding of UK labour markets. For participants consulted who lacked well-defined employment goals, uncertainty about their UK trajectory inhibited engagement with EAs and the development of actionable employment plans: "*I didn't have a specific plan, and I didn't know what job opportunities I was looking for. I wanted to do something like volunteering in the medical sphere, but I didn't mention this to my tutor*"[133].

Even among participants who engaged actively, employment support did not always bridge the gap between skills development and job attainment. Some reported mismatches between interview preparation and recruiter expectations, as well as limited feedback from employers, which undermined their confidence. As one participant reflected: "*I'm not confident when speaking still now. I have quite good English and [employers] always say that I speak quite well, but I don't feel confident in interviews because it's really important and stressful for me*"[134].

While the programme contributed positively to knowledge, skills, and confidence, securing employment often proved a longer and more uncertain process. Factors such as language proficiency, recognition of overseas qualifications, and the non-linear nature of job search limited immediate outcomes. These findings highlight the

[131] Female, 51-60 years old, upper intermediate English, employed.

[132] Female, 61+ years old, upper intermediate English, not in employment.

[133] Female, 18-30 years old, elementary English, not in employment.

[134] Female, 18-30 years old, advanced English, not in employment.

importance of strengthening support around job search persistence and resilience to help participants remain engaged and avoid discouragement when facing setbacks.

English language learning

The accounts of participants consulted reveal a complex balance between effective practice and persistent challenges in delivery. At its best, the provision created an immersive and engaging learning environment. Tutors who adapted lessons to participants' needs, provided constructive feedback, and fostered open communication were consistently praised. As one participant explained, the difference between an engaged and a disengaged tutor was stark: *"We spoke with [the teacher] about stuff besides the class, about life, about habits, about common stuff. But the [new teacher] who in front of and just go through the schedule, the topic. Monotone voice"*[135].

Participants highlighted that tutors who demonstrated strong classroom management and set high expectations helped sustain motivation and commitment to learning English. Those with positive experiences of the language component also recognised that it combined class-based interaction with self-directed study through homework and activities on the online portal. Many kept notes from these activities to use as a reference during and after the course. Participants further valued continued access to the portal beyond the programme, which enabled them to progress at their own pace and focus on areas of language relevant to everyday needs.

The strong emphasis on live communication was especially well-received. Participants consulted reported that practising in a supportive environment boosted their confidence, with several crediting the approach for accelerating improvements in listening and speaking skills. One participant described the impact on daily comprehension: *"After the programme, I remember I just walk on the street and [hear] some words and 'oh I know these words'… I start feel more confident to be honest, because… I better understand people"*[136]. However, while communicative practice aligned with the programme's aims of labour market integration, some learners, particularly those at intermediate or advanced levels, argued that insufficient attention was given to grammar and structural accuracy. For them, progress required deeper technical instruction rather than conversational fluency: *"How can you develop your language if no one corrects your mistakes or help you understand how to avoid those mistakes"*[137]. Similarly, another participant stated: "*I didn't like that we had activity but we didn't have like grammar information like a workbook, it just straight away you do this exercises without preparing to it or repeating*"[138]. This central tension reflected the challenging balance between differing pedagogical priorities for various learners: functional confidence versus linguistic precision.

The British Council's approach of delivering ESOL exclusively in English also divided students' opinions. While many acknowledged it improved their fluency, participants with lower proficiency found it alienating and, at times, a barrier to learning. *"For periods of time, we would just be sitting there and not even understanding what the*

[135] Female, 51-60, upper intermediate English, employed.

[136] Female, 18-30, intermediate English, employed.

[137] Female, 31-40, upper intermediate English, employed.

[138] Female, 18-30 years old, advanced level English, not in employment.

teacher wants… because it was only in English. There was no translation into Ukrainian at all"[139]*.* This highlights a structural risk in one-size-fits-all immersion models: they may accelerate progress for some learners while excluding others, particularly those least prepared for full English-only delivery.

Breakout groups also received varied feedback from participants. Some valued the opportunity to practise peer-to-peer communication highly, with one student remarking: "*We were always speaking in these groups. It was amazing*"[140]. However, the self-directed nature of these groups often resulted in uneven participation and persistent errors. As another participant noted, *"We heard our mistakes all time, but [the teacher] didn't help us"*[141]*.* Without active teacher monitoring, tailored group allocation, or provision of systematic feedback from English tutors, participants felt that the effectiveness of group work was diminished, and learning opportunities were not fully realised.

Workload and course intensity also emerged as critical challenges. Many participants struggled to keep pace with the syllabus and felt they did not have sufficient time to absorb what they had learned or practice it in class. They also expressed that there was insufficient overlap between lessons that would have allowed learners to build on what they had learned previously as they progress through the course: *"For every topic, we need at least two or three times more [time]... every day we started from the next lesson. We didn't continue the previous one, even though don't cover everything"*[142].

Participants who were balancing study alongside work or caring responsibilities tended to find the intensity of the English language classes particularly demanding. They described how this made it difficult to make room for other priorities, such as working or job searching. One participant explained, "*I was basically doing the course in the daytime. Then after lunch I was going for work, then coming back in the evening from work and doing homework. And it was basically the time where there was nothing else happening like I just needed to concentrate just on those things*"[143]. Several participants suggested that adjusting the course's intensity or introducing more flexible pacing could better accommodate those juggling multiple commitments.

Despite these challenges, the majority of participants understood that the English language learning component of STEP Ukraine was an intensive course, which required a large amount of effort to generate tangible progress. They typically reported tangible progress, particularly in confidence to communicate in everyday and professional contexts. Gains were most pronounced among those starting with lower levels of English. These participants valued being able to identify milestones in their progress in learning English, which motivated them to continue improving their

[139] Female, 41-50, elementary English, employed.

[140] Female, 61+ years old, upper intermediate level English, not in employment.

[141] Female, 41-50, intermediate English, student.

[142] Female, 51-60 years old, upper intermediate English, employed

[143] Male, 41-50 years old, intermediate English, employed.

communication skills. As one explained: *"It was hard for me to start speaking. Now, I can ask, 'Please tell me slowly. Please repeat me.' It's very, very good"*[144].

Even more advanced learners noted improved willingness to speak and reduced fear of errors: *"I really increase my level of English, even though my English wasn't on the lower level"*[145]. This included situations such as speaking to healthcare professionals during health appointments, speaking to customer service staff from utility service providers, while out shopping and speaking to neighbours. Participants with a higher level of English who wanted to improve their speaking skills felt the programme reduced their fear of making grammatical mistakes while talking and therefore increased their confidence to speak.

Views of non-participants

Interviews with Ukrainian visa holders who did not participate in STEP Ukraine suggest that awareness of the programme was widespread and broadly comparable to that of participants. Most had first heard of it through social media, Ukrainian community groups, or their hosts under the Homes for Ukraine scheme. This awareness often extended to both components of STEP Ukraine, although in practice, individuals were selective in their interest, focusing more on the element (language or employment support that aligned with their immediate needs. This indicates that while programme visibility was strong, the perceived relevance of different components varied, which may have shaped decisions about engagement with the programme.

Importantly, all five non-participants interviewed had applied to STEP Ukraine but ultimately did not participate, either because they were not accepted, as they did not meet the eligibility criteria or because they later chose not to enrol. For those not accepted, a recurring concern was a lack of clarity in the application process, particularly regarding acceptance status and start dates. As one interviewee reflected: *"So I thought that I'll fill in the application form and then there will be some feedback. You know, somebody will get in touch with me"*[146]. Such experiences prompted individuals to seek alternative support elsewhere, typically English language provision.

Among those who were accepted but decided not to proceed, the main issue was the perceived intensity of the programme relative to their personal circumstances. The programme design, while ambitious, did not always align with the complex realities of potential participants' lives and competing demands. Interviewees noted that they did not feel confident they could find the time needed to attend classes and complete independent study while also working, looking after children, and managing other commitments, such as learning to drive. One non-participant felt that the inflexibility of the class scheduling was a barrier because they would have needed to choose different time slots for when they attended class, but they were told by the programme organisers that this would not be possible. As they put it: *"it was impossible for me to take part in this programme, mainly because I just had no time.*

[144] Female, 41-50, intermediate English, student.

[145] Female, 61+ years old, upper intermediate level English, not in employment.

[146] Female, 41-50 years old, elementary level of English, not in employment.

All of my work shifts… were completely coinciding with the programme"[147]*.* These reflections underline how fixed structures within the programme could inadvertently exclude those with the highest need for flexibility, raising questions about the balance between programme coherence and adaptability.

Non-participants also reported that they typically took part in alternative support to STEP Ukraine. There was an emphasis on alternative English language provision, with interview participants identifying improving their communication skills as a priority for positive employment outcomes. Alternative support was often perceived as more flexible and therefore more manageable, particularly where providers allowed learners to adapt schedules around work and family responsibilities. One interviewee noted: *"The main point was that [another English language programme] was flexible and the second point was that it was provide for us as a charity, so I didn't need to pay for it"*[148]*.* Others with lower levels of English reported preferring Ukrainian-speaking tutors, sometimes paying privately, which suggests that linguistic and cultural familiarity was also a significant factor influencing take-up.

2.5 Conclusions

The evaluation of the STEP Ukraine programme underscores the role of tailored interventions in supporting Ukrainian visa holders' integration into the UK labour market. The programme's innovative dual-focus model, combining intensive English language training with personalised employment support, addressed participants' professional aspirations while accounting for structural constraints, such as limited transport, caring responsibilities, and short visa durations.

Several factors were instrumental in enabling the programme's successful implementation. Operational agility and rapid mobilisation allowed the programme to launch within six weeks of contract award, providing timely support to Ukrainian arrivals. The flexible, fully online delivery model facilitated participation across a geographically dispersed cohort. Personalised, culturally sensitive support – through Ukrainian-speaking EAs, some with lived experience as visa holders – enhanced trust and engagement. Iterative adaptations in recruitment, onboarding, and delivery mechanisms, alongside strategic partnerships with organisations such as BITC and TERN, further strengthened feasibility, retention, and access to sector-specific guidance. The embedding of a monitoring framework from the outset supported ongoing assessment of outputs and informed evidence-based improvements.

Despite these enablers, the evaluation identified several challenges. The compressed mobilisation timeline limited opportunities for detailed planning, comprehensive staff recruitment, and the finalisation of operational processes. High caseloads for Employment Advisers, rising from 25 to 35-36 clients, reduced the intensity and consistency of personalised support and limited opportunities for collaboration and shared learning. Variability in practice and mismatched expectations occasionally affected participant experience, particularly where the programme could not function as a recruitment service. The intensity and

[147] Female, 41-50, elementary English, employed.

[148] Female, 31-40, upper intermediate English, employed.

accessibility of the English language component created barriers for learners with lower proficiency, competing responsibilities, or caring commitments. Limited integration and communication between Employment Advisers and English tutors constrained the potential for a fully integrated programme, while structural and psychological barriers, including recognition of Ukrainian qualifications and the ongoing impact of war and displacement, continued to affect engagement.

Participants generally reported positive experiences, with employment support being perceived as most effective when tailored to individual goals, including CV adaptation, application guidance, and interview preparation. Flexibility in scheduling and the option to work with Ukrainian- or English-speaking Employment Advisers facilitated engagement. English language training was reported to improve confidence and communication skills, though intensity and the exclusive use of English posed challenges for some learners. Gaps in sector-specific employment support and qualification recognition limited alignment with participants' prior careers, while non-participants highlighted programme intensity, rigid scheduling, and limited flexibility as reasons for seeking alternative English provision.

Overall, STEP Ukraine demonstrates the feasibility of rapidly deployed, dual-focus support combining language and employment components for Ukrainian visa holders resettling in the UK. The programme's experience highlights the value of flexible, culturally sensitive, and responsive interventions, while also underscoring the need to balance rapid mobilisation with sufficient operational planning, integrated delivery, and ongoing support mechanisms. The findings provide evidence to inform future initiatives, ensuring they are both effective and adaptable to the diverse circumstances of participants.

Following on from the initial process evaluation, which was undertaken in Phase 1 of the evaluation, the following employment impact evaluation (Chapter 4 and economic evaluation (Chapter 5 were undertaken in the following contracting year. During this second contracting year, a refreshed version of the STEP programme has been introduced and is aimed at Ukrainian Guests and Hong Kong British Nationals. Although no formal process evaluation of the refreshed STEP programme has been undertaken, it is our understanding that it was designed to reflect changes based on lessons learned from this process evaluation.

Lessons learned from STEP Ukraine: Recommendations for future programmes

STEP Ukraine provides an evidence base for designing and delivering integrated employment and English language support for displaced populations in complex and evolving contexts. The following recommendations, structured across programme design, operational delivery, support mechanisms, partnerships, monitoring, and holistic support, are intended to inform the development of future initiatives and enhance accessibility, effectiveness, and sustainability.

Programme design and participant engagement

- To strengthen future programmes, it is essential to implement thorough eligibility and suitability assessments during participant registration. These should take into account each individual's time constraints, English proficiency, caring responsibilities, and motivation for both employment and language learning. Clear, upfront communication about programme intensity, its dual-focus nature, and the difference from direct job placement services will help set realistic expectations and reduce early dropouts.

- Adopting differentiated programme tracks can better support participants with varying work, family, or health commitments. For instance, a "two-speed" model could offer both a standard, full-intensity track and a lighter, more flexible option with fewer classes or reduced weekly contact hours. This approach allows participants to engage at a pace that suits their circumstances, ensuring greater accessibility for those balancing employment or caring duties.

Operational efficiencies

- Strengthen onboarding and expectation-setting: Implement stricter suitability checks and clearly communicate the intensity and commitment required for both English and employment components.

- To ensure high-quality, personalised support, EAs' caseloads should be limited to around 25-30 clients at a time. Integrating individual guidance with group sessions can enhance efficiency and boost overall engagement, particularly for foundational employment topics. Offering sessions at various times and in multiple languages is crucial, especially for participants who have limited English proficiency or demanding schedules, helping to maintain accessibility and consistent participation.

- Schedule planned breaks between programme cohorts to give staff time for reflection, strategic planning, and completion of administrative tasks, thereby maintaining both quality and continuity of the programme. Focus on hiring staff with backgrounds in refugee support, employment services, or language education, and ensure that all new team members receive comprehensive onboarding and role-specific training, including translated resources as needed.

- Implement a centralised and intuitive CRM system to simplify data collection, minimise duplication of effort, and enhance real-time monitoring of outcomes. Such a system is integral for tracking participants' progress, supporting effective programme management, and enabling timely adjustments based on operational insights.

- Enhance holistic support: Integrate direct mental health and trauma support, provide better guidance on childcare support, offer digital skills training, and formalise post-programme support (for instance, via an alumni network).

Employment support

- Clearly communicate that employment support aims to empower participants with the skills and knowledge needed to navigate the UK job market independently. Emphasise the development of self-directed job search strategies, tailoring CVs and applications, and building effective interview techniques.

- Encourage EAs to collaborate with participants in setting realistic, motivating goals for the short, medium, and long term, taking into account their previous qualifications and local labour market conditions. Support should be individually tailored to match each participant's ambitions and the specific requirements of their chosen sectors.

- Strengthen opportunities for professional networking by organising sector-focused events, workshops, establishing mentorship programmes, and forming partnerships with local employers and community groups, including for internships and apprenticeships. These structured initiatives can help participants connect with industry professionals, fostering smoother labour market integration and career advancement.

English language learning

- Design dedicated learning paths for lower-level English learners (A0-A1), incorporating focused grammar support, more structured exercises and discussion, and tailored feedback that addresses individual needs. Consider separating beginner groups into different levels to ensure teaching methods are well-matched to learners' proficiency and foster greater engagement.

- Communicate the expectations and structure of a full-immersion approach clearly during the application process, so participants can make informed decisions. For peer-led sessions, provide explicit instructions regarding participation, collaborative learning, and group dynamics. Assign staff or teaching assistants to actively monitor these discussions, offering real-time correction and guidance to address persistent language errors.

- Adopt a variety of instructional techniques, such as situational role-plays and practical language exercises, directly connected to participants' daily lives and employment goals. Provide the option of less frequent classes extended over a longer period, introduce mid-course breaks, and allow for flexible attendance policies to better accommodate participants with conflicting responsibilities.

- To maximise the effectiveness of breakout groups, set clear expectations around participation and turn-taking, ensuring all voices are heard and that no one participant dominates the discussion. Whenever possible, allocate teaching staff or assistants to actively support these groups, giving immediate feedback and correcting mistakes as they arise.

- Strengthen integration and communication: Improve direct contact between English tutors and EAs, provide feedback mechanisms between tutors and EAs, streamline attendance monitoring, and better coordinate class and employment support schedules.

- Support staff: Provide trauma-informed and context-specific training, offer Ukrainian-language support, and foster stronger peer networks to reduce isolation in online teaching.

Leveraging community and partnerships

- Strengthen trust and ensure culturally sensitive guidance by employing support staff and EAs who speak Ukrainian. This approach proved invaluable in STEP Ukraine and should be a cornerstone of future programmes.

- Promote community building through group activities, interactive online classes, and peer support mechanisms. Strategic use of word-of-mouth within the Ukrainian community has been particularly effective for recruitment and can continue to drive engagement and participation.

- Expand collaborations with local employers and intermediary organisations to provide practical employment opportunities, including internships, apprenticeships, or entry-level roles. Consider mechanisms to incentivise participation and secure sector-specific opportunities aligned with participants' prior experiences and skills.

Monitoring and evaluation

- Future monitoring frameworks should consider including metrics to assess non-employment outcomes, such as increased confidence, motivation, hope, and understanding of UK workplace norms, as these are key drivers of long-term outcomes.

- Simplifying post-programme surveys and making them more user-friendly could encourage higher response rates and yield deeper insights into long-term outcomes.

- Regular attendance proved to be a strong indicator of improved language and employment results; therefore, making attendance a central measure of success and actively supporting participants to remain engaged is highly recommended.

- Early targets in STEP Ukraine were often considered overly ambitious, especially regarding anticipated improvements in English proficiency. This sometimes created pressure to meet numerical goals at the expense of genuine progress. Introducing flexibility into KPI frameworks and allowing for adjustments over time will better support effective and responsive programme delivery.

Holistic support beyond core services

- Some participants experienced a sense of isolation after completing the programme. Establishing ongoing support, such as alumni networks, regular drop-in sessions, or continued access to EAs, can help extend the programme's benefits, foster lasting connections, and support participants' ongoing progress with employment and language development.

- Integrating specialist trauma-informed support and culturally sensitive mental health resources, co-designed and co-developed in active partnership with refugee communities and relevant experts, can improve wellbeing for participants affected by conflict and displacement. This approach recognises trauma symptoms, incorporates trauma-informed practices into the programme, and reduces the risk of re-traumatisation[149].

- Childcare responsibilities, especially among single mothers, limited access and sustained engagement. Embedding childcare support or establishing strong referral pathways to local childcare services can help address this barrier. Incorporating dedicated childcare support or establishing robust referral pathways to local childcare providers could help alleviate this challenge.

- Similarly, while housing support falls outside the direct scope of employment and language initiatives, building partnerships with housing organisations or offering basic housing advice could further address participants' needs and promote greater stability.

- STEP Ukraine relied significantly on participants' digital skills. To support full engagement, future programmes should provide comprehensive digital literacy training, offering introductory workshops or onboarding sessions for online platforms like Zoom and learning management systems. Equipping participants with these essential skills will enable them to access all aspects of the programme and make the most of available learning opportunities.

[149] Substance Abuse and Mental Health Services Administration. SAMHSA's Concept of Trauma and Guidance for a Trauma-Informed Approach. HHS Publication No.(SMA) 14-4884. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2014.

3 The Language Impact Evaluation

The objective of the language impact evaluation was to identify the impact of the STEP Ukraine programme on participants' English language skills. The research question that we sought to address was: What was the impact of STEP Ukraine funding on the English language outcomes of the programme's beneficiaries? The key findings and takeaways are summarised in Box 3.1.

Box 3.1 Key findings from the English language impact evaluation

- **STEP participants showed significant improvements in English test scores between programme entry and exit**, with average increases ranging from 20.8% (more conservative measure) to 29.2% (less conservative measure).
- **After accounting for English improvements that would have occurred without the programme, STEP participants' scores improved by 17.6% to 25.9% more than expected**. English skills tend to increase naturally with length of stay in the UK; analysis suggests participants would have experienced a 7.5-point increase in English test scores over the three-month programme period without STEP. The estimated programme effects are statistically significant, meaning they are unlikely to have been observed by chance.
- **Demographic factors are significantly associated with English language ability at programme entry**. Older individuals, men, those with lower educational attainment, and people with childcare needs tended to demonstrate weaker English skills before starting STEP.
- **Limitations of the counterfactual analysis should be noted**. The approach assumes a linear relationship between time in the UK and English ability, assumes no systematic differences between earlier and later arrivals on unobserved characteristics, and may be affected by attrition bias.

3.1 Methods and data

Data sources

We evaluated the impact of the STEP Ukraine programme on English language skills by using test scores and programme monitoring data that track the English proficiency of participants over time and include key demographic characteristics (like the participant's gender and region/country of residence).

The data sources used to carry out the impact evaluation of English language outcomes were:

- Pre- and post-intervention English proficiency assessments collected before and after participation in the programme.
- Other programme data, including demographics, attendance levels, dropout rates, and the length of time spent in the UK.
- A cross-sectional survey of Ukrainian visa holders in the UK conducted by the Office of National Statistics (ONS) in April 2024, with questions regarding STEP participation, employment outcomes, and self-assessed English ability.

Pre-post analysis of impacts on English language skills

The pre-post analysis of impacts on English skills consisted of conducting simple before-after comparisons of English test scores administered at the point of entry into the programme and at exit from the programme (the pre- and post-intervention English proficiency assessments). These comparisons were carried out individually for each student cohort of the programme that was within the scope of the evaluation – those who started the programme between August 2023 and August 2024 – as well as for all participants across cohorts. Care was taken to disaggregate by cohort in case changes in programme delivery over time altered the impacts of the programme. Standard t-tests were conducted to determine the statistical significance of the observed differences between average test scores at entry and exit.

Counterfactual analysis

While the above approach to pre-post analysis can identify changes in outcomes from before to after programme participation, it does not allow us to determine the causal impact of the programme. English language skills are likely to improve as migrants spend longer in the UK, whether or not they are subject to a specific intervention such as STEP. It is possible that this is particularly true of those Ukrainians who chose to participate in STEP – for example, because they may have high levels of motivation for developing good English skills and meeting or expanding their employment potential in the UK. Therefore, a credible strategy for identifying causal effects requires a method for estimating English language improvements we would expect in the absence of the programme (the "counterfactual"), accounting for the possibility of self-selection into the programme of specific types of individuals.

To address this challenge, we looked at the rate at which English language skills were acquired by STEP Ukraine participants in the absence of language training but after moving to the UK. We did this by conducting regression analysis of participants' entry test scores on their length of time spent in the UK, to estimate the rate of "natural progression" in English proficiency for programme participants – that is, how much participants' English level would have progressed in the absence of the programme (i.e., a counterfactual estimate). We produced this estimate while controlling for variation in entry scores across cohorts, as well as the following participant characteristics: age, gender, region or country of residence, months spent in the UK, whether the participant was employed, and whether they had childcare needs, all at the point of programme entry.

The estimated model regresses the English test score of each participant at programme entry on a time variable corresponding to the number of months spent in the UK when they received their entry test score. The estimated coefficient on this time variable tells us the relationship between the amount of time spent in the UK and a participant's English test score at programme entry. Using this estimated relationship, we were able to subtract the English language improvement that would have occurred in the absence of the programme from our pre-post estimate of the change in English ability between programme entry and exit. We refer to this adjustment as our "counterfactual adjustment." Intuitively, the improvement in English test scores over and above that which we would expect due to natural progression is attributed to the causal impact of STEP.

This approach does not require controlling for potential interventions or training received elsewhere, as the BAU scenario, which we intended to capture through our counterfactual estimate, assumes that potential STEP Ukraine participants are able to access other interventions and language support, e.g., learning English via apps, self-study from books.

As a robustness check for whether the counterfactual estimate obtained from the test score data was broadly valid, we used the ONS survey data to obtain a separate estimate of the relationship between months since arriving in the UK and English language ability, this time as measured by self-assessed English proficiency on a 1-5 Likert scale.

Relationship between demographic factors and English language ability

We also explored whether STEP participants or members of the population eligible for STEP exhibited systematic variation in their English language ability in the absence of the programme based on demographic or regional factors. For this, we carried out multivariate regression analysis of entry test scores, controlling for demographic variables from programme monitoring data. We also conducted analogous regression analysis using the ONS survey data, regressing self-assessed English proficiency on demographic variables. The included control variables in the analysis of test scores were STEP cohort, age, gender, region or country of residence, time spent in the UK (in months, whether the participant was employed, and whether they had childcare needs, all at the point of programme entry. The included controls in the analysis of self-assessed English proficiency were whether the respondent is 30 or older, gender, whether they have children aged 0-17, highest qualification, current living arrangements, whether they have been experiencing physical or mental ill health lasting 12 months or more, and region/country of residence.

3.2 Impact on English skills

We first present a "before-and-after" (pre-post comparison of STEP participants' English skills. We then undertake modelling to estimate the rate of English progression that we would have expected among this group in the absence of the STEP programme (the "counterfactual". Finally, we use a pseudo difference-in-differences approach in order to attribute the remaining observed progression, over and above the counterfactual, to the causal impact of STEP Ukraine.

Pre-post analysis of test scores

Our data source for the pre-post analysis of participants' English skills was the English proficiency assessments administered at entry to and exit from STEP. Descriptive statistics relating to this dataset are presented in Table 3.1 below. We disaggregate by cohort in case changes in programme delivery over time altered the impacts of the programme.

Table 3.1 Descriptive statistics for English proficiency assessments by cohort*

Cohort	Aug 23	Sep 23	Oct 23	Nov 23	Dec 23	Jan 24	Feb 24	Mar 24	Apr 24	May 24	Jun 24	Jul 24	All
Sample size	82	133	851	632	1,338	566	1,156	1,588	588	1,230	1,078	584	9,826
Entry test taken (% of sample)	100	100	100	100	100	100	100	100	100	100	100	100	100
Exit test taken (% of sample)	84.1	75.9	81.1	79.1	78.8	78.1	82.0	78.5	78.2	80.1	77.0	82.2	79.4
Average attendance (% of booked group classes)	63.0	57.1	67.5	64.2	64.6	64.2	67.2	60.2	57.9	59.7	56.2	61.5	62.2
Average age	41.6	39.2	40.8	40.7	41.2	41.6	41.6	40.6	41.3	40.1	40.9	42.2	41.0
% male	17.1	10.5	18.6	17.1	18.9	18.4	18.9	20.0	22.3	19.2	20.1	23.6	19.4
% female	81.7	88.7	81.2	82.8	80.9	81.6	80.9	80.0	77.6	80.8	79.8	76.4	80.4
% unemployed	59.3	55.6	52.7	47.9	51.4	52.5	52.2	50.9	50.7	50.7	51.6	80.0	53.1
% with childcare needs	24.7	33.8	30.9	27.8	30.5	28.3	27.7	29.2	23.0	22.6	25.6	23.8	27.3

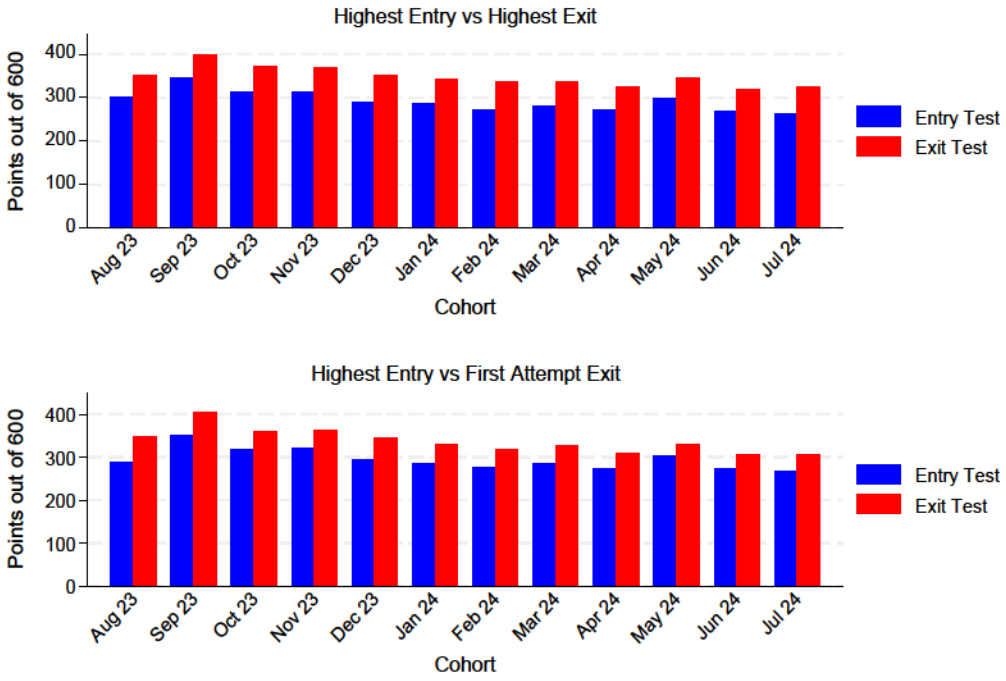
*Note - This is for the highest test scores only – the sample changes when we use only first-attempt scores due to data loss from merging datasets. This is because the first-attempt scores were collected in a separate dataset, which needed to be obtained from the British Council, and the lack of a consistent identifier used throughout the programme's rollout complicated merging across datasets.

Comparing scores from the English tests administered at programme entry and exit allowed us to determine pre-post changes in English proficiency for STEP

participants. As multiple test attempts were allowed for both the entry and exit tests, the figures in this section include two separate charts – one that compares the highest entry test score with the highest exit test score attained by participants, and another that compares participants' highest entry test scores with their first attempt exit test scores. The highest-scoring attempt obtained in the entry test is what was used for placement at the appropriate programme level. We judge that it is probably more appropriate to compare this with the highest exit score than the first attempt exit score, since to measure the change between two time points, it is most robust to use an analogous measure at both points in time. That said, we include the first attempt exit scores too as a more conservative scenario.

The comparison of entry and exit test scores (Figure 3.1 below suggests that programme participants in all cohorts showed a noticeable improvement in their English test scores going from the entry test to the exit test, using the average of the four component tests (grammar, vocabulary, listening, and reading in either case. We can see that a visible improvement remains even after using the participants' first attempt test scores for the exit test. All tests were scored out of 600.

Figu



The comparison of changes in average entry and exit test scores (see Figure 3.2) confirms that the choice of exit score does not change our overall conclusions from the pre-post analysis of test scores. It shows average test score differences going from entry to exit as a percentage change from participants' entry test scores, distinguishing between cohorts (see Figure A2.1 in Appendix 2 for the same results shown as standard deviations from average entry test scores). We can see that the average percentage change in test scores, regardless of which exit score is used, is considerable. This is confirmed by Table 3.2, which shows the average difference from entry to exit in terms of points, percentage, and standard deviations. We find that even the most conservative pre-post comparison (first attempt exit versus highest entry) shows an average improvement of 20.8% in participants' English scores from programme entry to exit. This corresponds to about one-third of a standard deviation.

Figure 3.2 % change in average entry versus exit test score by cohort

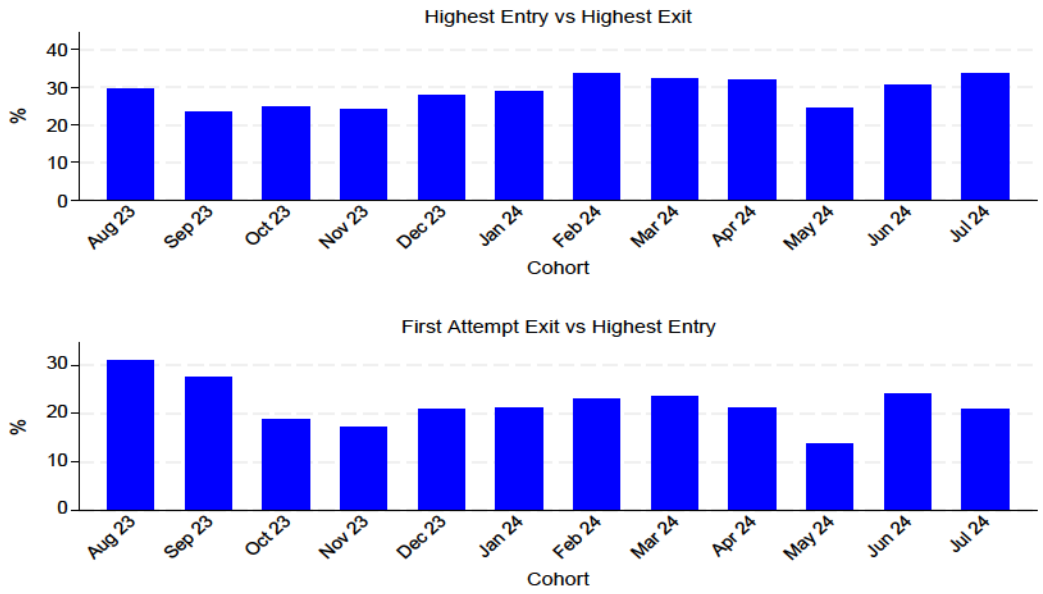


Table 3.2 Average test score difference in points, percentage, and standard deviation (all cohorts)

Test score comparison	Points	Percentage	Standard deviations
Highest exit vs highest entry	54.4	29.2%	0.45
First attempt exit vs highest entry	38.0	20.8%	0.31

This improvement in test scores is statistically significant at the 1% level for every cohort and for the participants as a whole. This holds regardless of whether the exit score used is the first attempt or the highest of multiple attempts (see Tables A2.1 and A2.2 in Appendix 2).

In short, the pre-post analysis suggests that STEP participants' English proficiency increased considerably following the 12-week course. The next step in the impact evaluation, covered in the next section, is to estimate how much of this change can be attributed to the STEP programme.

Counterfactual analysis of test scores

Figure 3.3 shows that, for each cohort, the average entry score is higher for participants who have spent longer in the UK. This positive correlation between months spent in the UK and average entry score can be used to distinguish the programme's impact on participants' English skills from the improvement they would have experienced regardless, simply from living in the UK and independently accessing available resources and support outside of the programme.

Specifically, by regressing participants' entry test scores on their length of time spent in the UK, we can attempt to estimate the rate of "natural progression" in English proficiency for programme participants. This represents an estimate of how much their English level would have progressed in the absence of the programme (i.e., a counterfactual estimate). We can refine this estimate further by controlling for variation in entry scores across cohorts, shown in Figure 3.4 below, as well as the following participant characteristics on which data is available: (i) STEP cohort, (ii) age, (iii) gender (whether they are male), (iv) their region/country of residence, (v) whether they are unemployed, and (vi) whether they have childcare needs. The results of this analysis are presented in Table 3.3 and the full regression output is provided in Table A5 in Appendix 2.

Figure 3.3 Line of best fit for relationship between average entry test score (highest) and months in the UK, by cohort

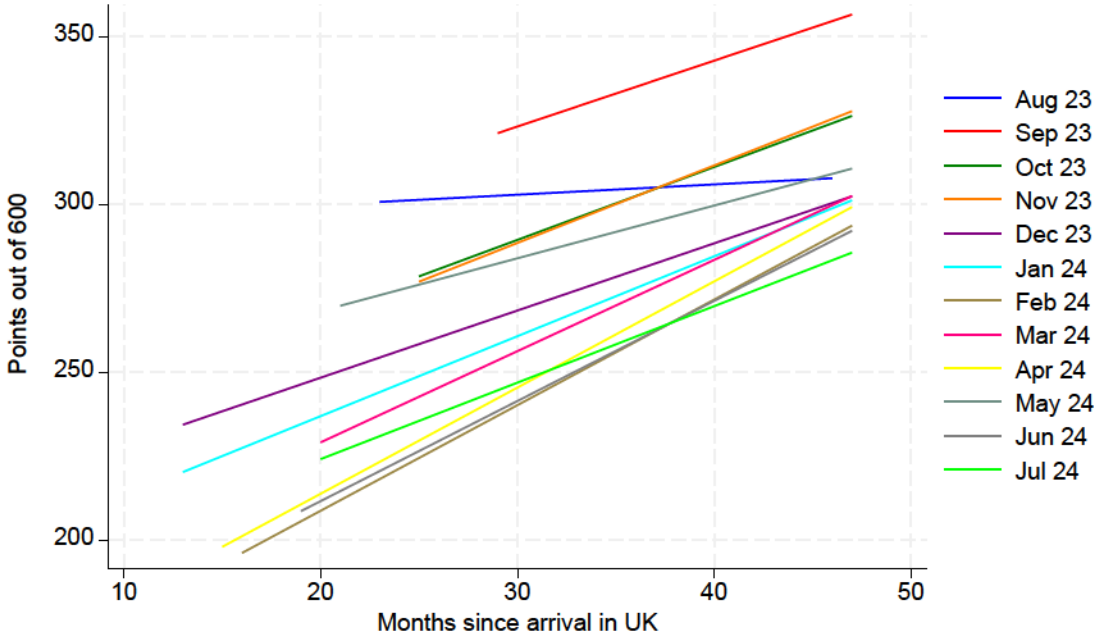
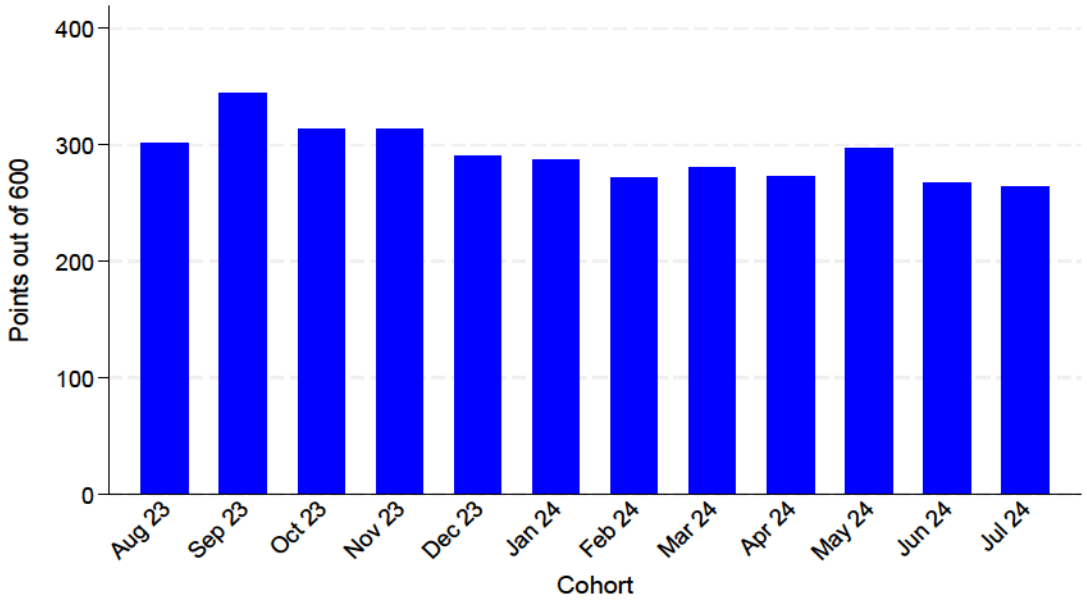


Figure 3.4 Average entry test score (highest) by cohort



The estimate we are primarily interested in – that of the coefficient on months since arriving in the UK – does not change considerably after controlling for age, gender, region/country of residence, employment status, and childcare needs. This suggests that our estimate of the relationship between average entry scores and the number of months spent in the UK is relatively robust to potential confounding factors.

Based on the regression controlling for cohort and participant characteristics, we find that for each additional month spent in the UK, a programme participant can expect their entry test score to be more than 2.5 points higher (out of the 600 total). This amounts to an average increase of 7.5 points over a period of 3 months, which is the length of the STEP Ukraine programme. We can interpret this 7.5-point increase as a plausible estimate of the counterfactual increase in English skills that a programme participant would have experienced if they had spent those 3 months in the UK but without participating in the STEP programme.

Table 3.3 Estimates from regressing entry test average score (highest) on months in UK

Specification	Coefficient on months in UK
Without controls	2.75**
With cohort controls	2.52**
With full set of controls	2.51**

Note: Statistical significance is denoted as follows: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Using a procedure resembling difference-in-differences, Table 3.4 below applies this 7.5-point increase as a counterfactual adjustment to the pre-post results presented earlier. We do this by subtracting 7.5 points from the pre-post improvement in test scores. The residual improvement in test scores is what we attribute to the causal impact of the STEP programme. We estimate this effect to be at least 17.6%, using the first attempt score. This is statistically significant both across and within cohorts (see Tables A2.4 and A2.5 in Appendix 2). Note that statistical significance is influenced by sample size and in itself does not imply causality.

Table 3.4 Average adjusted test score difference in points and percentage change (all cohorts)

Test score comparison	Average points difference	Average percentage change	Average standard deviation change
Highest exit (adjusted) vs highest entry	46.9	25.9%	0.39
First attempt exit (adjusted) vs highest entry	30.5	17.6%	0.25

Robustness check using ONS survey data

As a robustness check for whether the counterfactual estimate of 7.5 points is broadly valid, we used the ONS survey data to obtain an estimated linear trend representing the relationship between self-assessed English proficiency on a 1-5 Likert scale and months since arriving in the UK. Descriptive statistics for the ONS survey dataset are provided in Table 3.5.

Table 3.5 Descriptive statistics for ONS survey data by STEP participation

Sample	STEP participants	STEP non-participants	All
Sample size	1,003	6,221	7,224
% aged 18-29	8.9	19.2	17.8
% aged 30-49	76.6	64.9	66.5
% aged 50-69	14.4	13.8	13.9
% aged 70 and over	0.2	2.0	1.8
% male	16.0	27.7	26.1
% unemployed	23.1	11.8	13.5
% with childcare needs	5.8	5.1	5.2

Note: Respondents reporting that they have left the UK permanently or were never in the UK are excluded.

Table 3.6 presents the key results from regressing self-assessed English proficiency on months since arriving in the UK, as well as respondent characteristics on which data was collected. The following respondent characteristics were included as control variables in the regression specification with controls: (i) whether the respondent is 30 or older, (ii) gender (whether they are male), (iii) whether they have children aged 0-17, (iv) the level of their highest qualification, (v) their current living arrangements, (vi) whether they have been experiencing physical or mental ill health lasting 12 months or more, and (vi) their region/country of residence. The full regression output is provided in Table A8 in Appendix 2.

We find that for each additional month spent in the UK, respondents' self-assessed English score increases, on average, by 0.01 points out of 5 (significant at the 5% level). Scaling this result so that the maximum number of points is the same as the English tests used in the STEP programme (600, instead of 5), we estimate that respondents' self-assessed English proficiency would have increased by 4.5 points out of 600 over 3 months. While this figure is smaller than the 7.5-point counterfactual increase estimated using programme data, it confirms our overall conclusion that STEP participants experienced a statistically significant improvement in their English test scores, and the magnitude of the improvement was at least 17.6% (using the first attempt exit test score) relative to their entry test score.

Table 3.6 Estimates from regressing self-assessed English level (minimum) on months in UK

Specification	Coefficient on months in UK
Without controls	0.02**
With controls	0.01*

Note: Statistical significance is denoted as follows: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Relationship between demographic factors and English language ability

The above analysis highlighted several additional factors that appear to influence the English language abilities of Ukrainian refugees in the UK. These insights are drawn from the two regression models with controls presented in Tables 3.3 and 3.6, each examining a different measure of English proficiency in relation to a range of demographic characteristics, as well as the region/country of the UK where individuals currently reside. The measure of English language ability used in the first regression was the highest entry score test result among STEP Ukraine participants, while the second regression used ONS survey respondents' self-assessed English level (the lowest among their self-assessed speaking, reading and writing levels).

In both regression models, youth was associated with higher levels of English language ability. These effects were highly significant, remaining robust even at the 1% level. Specifically, the results indicated that for each additional year of age, STEP participants score on average 2.7 points lower on the entry test. Similarly, the regression using ONS data found that respondents over the age of 30 report a 0.44-point decrease in their self-assessed English proficiency on a 5-point scale, compared to younger respondents.

Gender appears to be an important factor influencing language proficiency, with females demonstrating notably stronger English skills. The regression results showed that female STEP participants, on average, scored 11.5 points higher on their highest entry test compared to their male counterparts. In addition, females in the ONS survey rated their self-assessed English ability 0.13 points higher on a 5-point scale than males. Both of these impacts were found to be statistically significant at the 1% level.

We also observed that respondents with children tended to demonstrate lower English language proficiency than those without children or whose children are already adults. On entry tests, participants with childcare needs scored on average 14.7 points lower than their counterparts without such needs, a difference that is statistically significant at the 1% level. Similarly, the analysis of ONS survey data showed that having children under 18 years of age was associated with a 0.06-point reduction in language ability on a 5-point scale, though this effect was only statistically significant at the 10% level.

The region/country of the UK in which a respondent lives also appeared to account for variation in English language ability. Results from the STEP Ukraine entry tests showed that, compared to respondents in North England (the reference group, those from the Midlands and from South and East England scored significantly higher, by 17.3 and 13.2 points, respectively. In contrast, participants living in Northern Ireland scored on average 21 points lower than those in North England. Meanwhile, the scores of respondents from Scotland and Wales did not differ significantly from those of participants in North England. The regression using ONS survey data, which also uses the North of England as the reference category, presented a somewhat different picture. The estimates suggested that both the Midlands (-0.13 and Wales (-0.21 had lower scores on the 5-point scale compared with the North of England, although the effect for the Midlands was only statistically significant at the 10% level. For all other regions/countries, including the South and East of England, Northern Ireland, and Scotland, there was no statistically significant difference relative to the North of England.

Results from the STEP Ukraine entry test indicated no statistically significant difference in scores between employed and unemployed participants. Regression analysis from the ONS survey showed, as might be expected, a strong positive association between education level and English proficiency. Relative to individuals with degree-level qualifications, those with vocational or job-specific qualifications scored on average 0.77 points lower, while those with school-level qualifications scored 0.36 points lower. Participants with no qualifications reported average scores of just 0.94 on a 5-point scale. All of these effects were statistically significant at the 1% level.

Regarding living arrangements, the ONS survey regression results showed no statistically significant differences in English language ability between individuals living with a sponsor, with family or friends, or in temporary accommodation or experiencing homelessness. However, the results did indicate a statistically significant increase of 0.14 points in self-reported English ability among those who were renting compared with those living with a sponsor.

The ONS survey regression results indicated that individuals with a physical or mental health condition lasting 12 months or longer scored, on average, 0.01 points lower on the 5-point scale. However, this effect is small in magnitude and only statistically significant at the 1% level.

3.3 Conclusions

The evidence from this impact evaluation strongly indicates that the STEP Ukraine programme succeeded in improving English language ability among participants. Our analysis went beyond simply comparing test scores before and after the STEP programme. We used the observed relationship between participants' initial English language test scores and time spent in the UK to understand how participants' language skills would have improved even if they did not participate in the programme. This approach allowed us to estimate the improvement in English language ability that can be specifically attributed to the STEP Ukraine programme.

We estimate that the STEP programme improved participants' English skills over the course of the programme by 17.6% according to our conservative estimate and by 25.9% on average according to our less conservative estimate. These improvements were statistically significant, providing strong evidence that the STEP programme enhanced English language proficiency beyond what would have occurred naturally through independent progression. Strengthening English proficiency was a core objective of STEP, and the results clearly demonstrate that the programme achieved that.

Our analysis also highlighted the influence of demographic and regional factors on language ability pre-entry to the programme. Younger participants, women, individuals with higher levels of education, and those without childcare generally demonstrated stronger English language skills prior to participation in the programme. There was also considerable, albeit more mixed, variation in language proficiency depending on the region/country of the UK in which participants reside.

The STEP Ukraine programme, as delivered, clearly improved participants' English language skills beyond what would be expected through natural progression alone. The implemented English course should therefore be considered a successful model for informing the design of future programmes that share this objective.

Limitations of the analysis

Notable limitations to the analysis of the English language impacts are as follows:

- The counterfactual analysis relies on the assumption that earlier arrivals did not differ from later arrivals systematically on unobserved characteristics that affect the relationship between time in the UK and English language ability (e.g., unobserved ability, selective migration timing).
- Our pseudo-difference-in-differences approach relies on assumptions such as a linear relationship between time spent in the UK and the English language ability, which may not perfectly represent what would have occurred in the programme's absence.
- Around 20% of participants dropped out of the programme before taking an exit test, meaning that the results may be subject to attrition bias. If individuals who dropped out of the programme differ systematically from those who remained in terms of the impact the STEP Ukraine programme had on them, then results may be biased. For example, if participants drop out of the programme because it is not improving their English skills, then we would overstate the positive impact of the programme.

4 The Employment Impact Evaluation

The primary objective of the employment impact evaluation was to assess the causal impact of the STEP Ukraine programme on participants' employment outcomes. The key findings and takeaways of the employment impact evaluation are summarised in Box 4.1.

Box 4.1 Key findings and takeaways of the employment impact evaluation

- **STEP participants were 6.6 percentage points more likely to be in employment over the 20 months following entry to the programme**, when compared to Ukrainian guests who did not participate in the programme, but who otherwise share similar observed characteristics.
- **Estimated employment reached a steady state of approximately 8 percentage points from 7 months after treatment**, but impacts did emerge shortly after participation and grew steadily in the months afterwards.
- **STEP Ukraine participants were 7.7 percentage points more likely to be in sustained employment for at least 3 consecutive months** and 6.5 percentage points more likely to remain employed for 6 consecutive months.
- **STEP Ukraine participants experienced approximately 10.7% higher wages over the 20 months after the intervention**, assuming that they were already employed prior to the start of the programme. This estimate will include impacts from either increased hours of work or hourly wages (the data available does not allow us to distinguish between the two).
- **STEP Ukraine participants in Scotland were 6 percentage points more likely to be employed over the 20 months following the start of the programme**, compared to if they had not participated. Due to data constraints, we could only partially answer the question of whether impacts varied across the devolved administrations. The estimated impacts of STEP on employment are broadly the same in Scotland as in the UK as a whole. Unfortunately, it was not possible to draw conclusions for Wales and Northern Ireland, due to low sample sizes and more difficulty establishing robust control groups than in the UK-level analysis[154]. Note that this does not mean that we found the programme to be ineffective in Northern Ireland or Wales, but that we are not able to draw conclusions based on the statistical analysis undertaken in this study.

[154] For more information on the analysis of employment impacts in Northern Ireland, Scotland, and Wales, please see section 4.3.

4.1 Methods and data

We evaluated the impact of STEP Ukraine on employment outcomes of participants using a quasi-experimental design, comparing the evolution of outcomes of participants to a counterfactual group of Ukrainian nationals in the UK who did not participate in the programme. The employment impact evaluation addressed the following research questions:

1. What is the impact of STEP Ukraine on employment probability? Does participation in the programme increase the likelihood that participants are employed?

2. What is the impact of STEP Ukraine on employment sustainability? Does the programme help participants maintain continuous employment over time (e.g., remaining employed for at least 3, 6, or 12 consecutive months)?

3. What is the impact of STEP Ukraine on earnings among those in employment? For participants who were already employed at the time of programme participation, does STEP Ukraine lead to higher monthly earnings compared to non-participants?

4. Do the impacts of STEP Ukraine vary across devolved administrations (Scotland, Wales, and Northern Ireland)?

Data sources

The evaluation of the impact of STEP Ukraine on the employment outcomes of Ukrainian guests used three data sources. The primary data source was HMRC Pay As You Earn (PAYE), which contains information on monthly earnings from employment, as well as the region/country of the individual during that month. The secondary data source used includes the Migrant Worker Scan (MWS) and the STEP Ukraine programme monitoring data held by WJR.

STEP Ukraine participants commencing the programme in August 2023 to July 2024 were identified in HMRC data using National Insurance Numbers (NINOs) and other identifying information available in the programme monitoring data collected by WJR. This identified the 'treated group' for the quasi-experimental analysis – those who participated in the STEP Ukraine programme. To identify the control group for the analysis, the MWS was used to identify Ukrainian guests in the UK who did not take part in STEP Ukraine. The resulting dataset was an individual-level monthly panel dataset, containing both STEP Ukraine participants and a counterfactual group of non-participating Ukrainian guests. For each individual, the panel ran from March 2022 or from the individual's date of entry into the UK until September 2025.

HMRC PAYE data

For both the participant and non-participant groups, the panel dataset included monthly gross earnings, monthly employment status (a binary variable, identified based on non-zero earnings for a given month), a binary treatment indicator for whether the individual was a STEP Ukraine participant, date of birth (DoB), sex,

nationality, date of arrival in the UK, and geographic location (a categorical variable of 12 regions/countries across the UK). For STEP Ukraine participants, the data also included the month in which they started the programme.

Geographic location data was only available in HMRC PAYE data for individuals with at least one PAYE record during the timeframe of the panel (March 2022–September 2025). These variables were not available for those who had not been employed in the UK at some point since receiving a NINO. For STEP Ukraine participants, programme monitoring data was used to address the systematically missing geographic location data. There was no secondary data source that could address the missing geographic location data for the control group. Unfortunately, this means that there is no geographic information for those non-participants who were never employed during the period covered by the dataset.

Programme monitoring data

The programme monitoring data collected by WJR contained the following data used to match and prepare the dataset for analysis:

- NINOs of 5,837 out of 9,826 STEP participants (59.4%);
- First and last names of all 9,826 STEP participants (100%);
- DoBs of 9,799 out of 9,826 STEP participants (99.7%);
- Month of programme start of all 9,826 STEP participants (100%);
- Geographic information, collected by WJR at programme registration, of all 9,826 STEP participants (100%).

Geographic information in the programme monitoring data was used to fill in missing geographic location data of those participants who do not have a PAYE record in the panel (i.e., participants who were never in employment).

Migrant Worker Scan

The MWS is data held by HMRC on all overseas nationals living in the UK who have registered for and been allocated a NINO. This dataset was used to identify non-participating Ukrainian Nationals living in the UK, in order to develop a counterfactual group of non-participants for the analysis.

Data linkage and preparation

To enable the linking of STEP Ukraine participants, the NINO, first and last name, date of birth, cohort, and geographic location for each participant were shared with HMRC. In addition, HMRC used the MWS to gather data on non-participant Ukrainian Guests living in the UK. This enabled HMRC to produce a dataset for the control group, providing information on employment and pay outcomes.

The PAYE data of both participants and non-participants was matched by HMRC using a combination of exact matching, based on NINOs, and fuzzy matching[155], based on DoB, first and last name, for those individuals lacking NINOs. Even when NINOs were available, fuzzy matching based on other identifying information was used to ensure a correct match in case of any errors in the NINO. Once matched, personal identifying information (NINOs, first and last name) was removed from the dataset before it was provided to us for analysis.

The HMRC developed a rating system for classifying the quality of the matches, with Green and Amber matches giving sufficient confidence that the data was correctly matched (see Table 4.1). 'Red' matches were discarded from the dataset, as we believed there to be chance that the data would be incorrectly matched.

Table 4.1 Match rating system for HMRC PAYE data linkage

Match rating	% of STEP participant sample used in analysis	Description
Green	88.9	Either matched on: - NINO, first name, last name, and DoB; or - first name, last name and DoB.
Amber	11.1	Either matched on: - NINO and 1 or 2 of first name, last name and DoB; or - 2 of first name, last name and DoB.
Red (discarded)	0	Either: - matched on NINO only, or - no match was possible using the data provided.

The sample used in the analysis only contained those who were of working age, 18-65, and those who feasibly could have participated in the programme. Individuals who were of UK State Pension age were removed, as were records for individuals under 18. Individuals with a missing arrival date or a date of arrival in the UK after the final cohort of STEP Ukraine included in this analysis (August 2024) were also removed. STEP participants with a date of arrival in the UK after the month their cohort began the programme were dropped due to the inconsistency in the data. Unfortunately, we are not able to distinguish between individuals where the information which made them ineligible was correct and those where it was misreported. In total, this resulted in 26,520 observations being removed from the raw data received from HMRC before the full analysis was conducted, equating to approximately 16% of the raw sample.

Due to the nature of real-time PAYE data, in a very small minority of cases, gross monthly pay was negative. This was due to the correction of errors in the dataset

[155] Fuzzy matching is a technique used to link similar or related data when it is not possible to identify an exact match with numerical data. In this case, because programme monitoring data only contained NINOs for 59.4 percent of STEP Ukraine participants, fuzzy matching is used to link HMRC PAYE and programme monitoring data based on a probabilistic combination of other identifying information.

happening in real time. For example, when overpayments in a previous month that were subsequently corrected in the following period. In these instances, negative pay was coded to 0, and these individuals were considered not employed in that period.

Baseline descriptive statistics

After the preparation steps outlined above, the total sample for the employment impact evaluation consisted of 134,244 individuals, including 8,268 STEP Ukraine participants and 125,976 non-participating Ukrainian nationals living in the UK.[156] The 'treated' group refers to those who participated in STEP Ukraine, and the 'control' refers to Ukrainian guests who did not participate.

Table 4.2 summarises the key demographic characteristics of the treated and control groups used in the employment impact evaluation, based on their ages in July 2023, the month before the first cohort started the programme. Because the control group was not randomised, there is no guarantee that the treatment and control groups have the same characteristics, and the Table shows some key differences. STEP Ukraine participants were 80.9% female, and 45.2% were aged 35-44. In terms of their geographical location, the highest shares were located in London (15.1% and the South East of England (16.9%. While a considerable share of STEP Ukraine participants was located in Scotland (16.0%, much smaller numbers were located in Wales (5.0% and Northern Ireland (2.3%.

Table 4.2 Descriptive characteristics of the employment impact evaluation sample, by group

Descriptive characteristic	Category	Treated group	Treated group	Control group	Control group
-	-	*STEP Ukraine Participants*	*%*	*Non-Participants*	*%*
Sex	Men	1,578	19.1	43,796	34.8
Sex	Women	6,690	80.9	82,180	65.2
Age group	24 or under	590	7.1	27,526	21.9
Age group	25-34	1,831	22.1	34,127	27.1
Age group	35-44	3,739	45.2	35,357	28.1
Age group	45-54	1,691	20.5	19,026	15.1
Age group	55-65	417	5.0	9,940	7.9
Time since arrival in the UK (July 2023)	Arrived after July 2023	858	10.4	14,341	11.4
Time since arrival in the UK (July 2023)	0-6 months	770	9.3	13,101	10.4

[156] Around 1,500 participants were dropped from the sample due to incomplete information about age or incompatible dates of arrival.

Descriptive characteristic	Category	Treated group	Treated group	Control group	Control group
Time since arrival in the UK (July 2023)	7-12 months	1,921	23.2	28,696	22.9
Time since arrival in the UK (July 2023)	13-18 months	4,383	53.0	60,419	48.0
Time since arrival in the UK (July 2023)	19-24 months	63	0.8	3,207	2.6
Time since arrival in the UK (July 2023)	Over 24 months	273	3.3	6,212	4.9
Region or country	East	632	7.6	*N/A*	*N/A*
Region or country	East Midlands	480	5.8	*N/A*	*N/A*
Region or country	London	1,246	15.1	*N/A*	*N/A*
Region or country	North East	226	2.7	*N/A*	*N/A*
Region or country	North West	598	7.2	*N/A*	*N/A*
Region or country	Northern Ireland	187	2.3	*N/A*	*N/A*
Region or country	Scotland	1,153	14.0	*N/A*	*N/A*
Region or country	South East	1,397	16.9	*N/A*	*N/A*
Region or country Region or country	South West	919	11.1	*N/A*	*N/A*
Region or country	Wales	416	5.0	*N/A*	*N/A*
Region or country	West Midlands	493	6.0	*N/A*	*N/A*
Region or country	Yorkshire and the Humber	521	6.3	*N/A*	*N/A*

The baseline employment characteristics of the treated and control groups prior to STEP Ukraine's start are summarised in Table 4.3. To accurately compare the treatment and control groups in descriptive statistics, we have to pick an individual point in time to make the comparison. Otherwise, the characteristics of the treatment group would be taken over at different points of the year, while the control group would

be taken only at one point in time. Therefore, these statistics are again based on the characteristics observed in July 2023, just before the first cohort of STEP Ukraine started the programme. As not all the individuals included in the full analysis sample had arrived in the UK or were aged 18 in July 2023, these descriptive characteristics reflect a smaller sample of 7,262 treated individuals and 106,983 control individuals (i.e., only those who were in the UK and aged 18 in July 2023. The full sample used in the impact evaluation analysis includes individuals who entered the UK between July 2023 and August 2024 or aged into eligibility during that period.

Table 4.3 presents mean employment rates. These are calculated from an individual employment indicator where 0 indicates that an individual is not employed and 1 indicates that an individual is employed. It also presents average monthly earnings, including in the sample individuals with gross pay equal to 0. At baseline, employment rates are higher in the treated group, with a mean of 0.494. As 80.9% of the treated group, the average reflects the trends among female participants. The employment rate is higher among treated women (0.501 compared with men in the treated group (0.458. However, among the control group, individuals' monthly wages are higher than those of the treated group, with average monthly earnings of £799.18 compared to £761.09. This could be for a range of reasons, including the fact that those with higher earnings are less likely to seek out support because they already have more of the skills or English language ability that the programme is designed to improve. On average, wages are lower for women in both the treated and control groups compared to their male counterparts. This may be driven by higher rates of part-time work among women, although, as we do not observe hours of work in the HMRC data, we cannot distinguish the impacts of lower hours from lower hourly wages. As average monthly earnings include those who are not employed, at baseline, we see that while employment rates are higher among STEP Ukraine participants compared to non-participating Ukrainian nationals in the UK, wages are considerably lower.

Table 4.3 Employment characteristics of treated and control groups at baseline (July 2023)

Employment characteristic	Total treated group	Men treated group	Women treated group
Employment (mean, 1=employed)	0.494	0.458	0.501
Monthly earnings (mean, £)	761.09	977.19	717.98
Employment characteristic	**Total control group**	**Men control group**	**Women control group**
Employment (mean, 1=employed)	0.396	0.335	0.426
Monthly earnings (mean, £)	799.18	857.47	771.24

Employment impact evaluation methodology

The employment impact evaluation examines whether STEP Ukraine improved participants' employment outcomes using a quasi-experimental design (QED. This type of methodology was required because the programme did not include a randomised control group (i.e., participation was voluntary and not withheld from potential participants. Therefore, we effectively compared the evolution of outcomes for people who took part in STEP Ukraine with those of Ukrainian guests who did not take part in the programme, but who were demographically similar. This comparison allows us to causally attribute the change in employment outcomes to participation in the STEP Ukraine programme, under the assumption that there were no other changes that impacted the treatment and control groups differently.

The approach we used is called 'Difference-in-Differences' (DiD because it compares changes in the outcome before and after the introduction of STEP Ukraine, for the treated group (STEP participants and the control group (non-participants. However, traditional DiD approaches are not able to robustly account for the fact that STEP Ukraine participants joined the programme in different months (i.e., in monthly cohorts. Therefore, we utilised an alternative DiD approach, which is robust to a staggered introduction of an intervention as suggested by Callaway and Sant'Anna.[157]

This method constructs separate control groups for each monthly cohort of STEP participants, using individuals who had not yet joined, or who never joined the STEP Ukraine programme. In addition, the method applies statistical weighting to help ensure that there are similarities between the treated and control groups in terms of observable characteristics (age, sex, and the number of days since arrival in the UK.

A key requirement of the DiD approach we used is that the treated and control groups must have followed similar employment trends before STEP Ukraine began. We undertook both visual and statistical assessments of whether this requirement holds for our analysis and provide results for these tests throughout the upcoming section, highlighting that, overall, our results are robust to this requirement.

Our evaluation focussed on three key employment outcomes, in line with first three of our research questions: (i employment probability identified by whether someone had positive earnings in any given month; (ii sustained employment defined as holding continuous employment over a 3-, 6-, or 12-month period; and (iii monthly wages (conditional on employment using national logarithms so that estimates can be interpreted approximately as percentage effects on wages.

The analysis was constrained by the limited range of variables available in the administrative data. For example, geographic location could not be used in the main analysis because it was unavailable for non-participants (i.e., the control group who were never employed during the period studied; however, sensitivity analysis, which included geographic location or the subset of individuals where it was available, produced similar results.

---

[157] Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2), 200-230. https://doi.org/10.1016/j.jeconom.2020.12.001

We provide full technical details, including detailed information about the statistical specification, weighting approach, outcome definitions, and parallel trends tests, in Appendix 4.

4.2 Impacts at the UK level

Employment probability

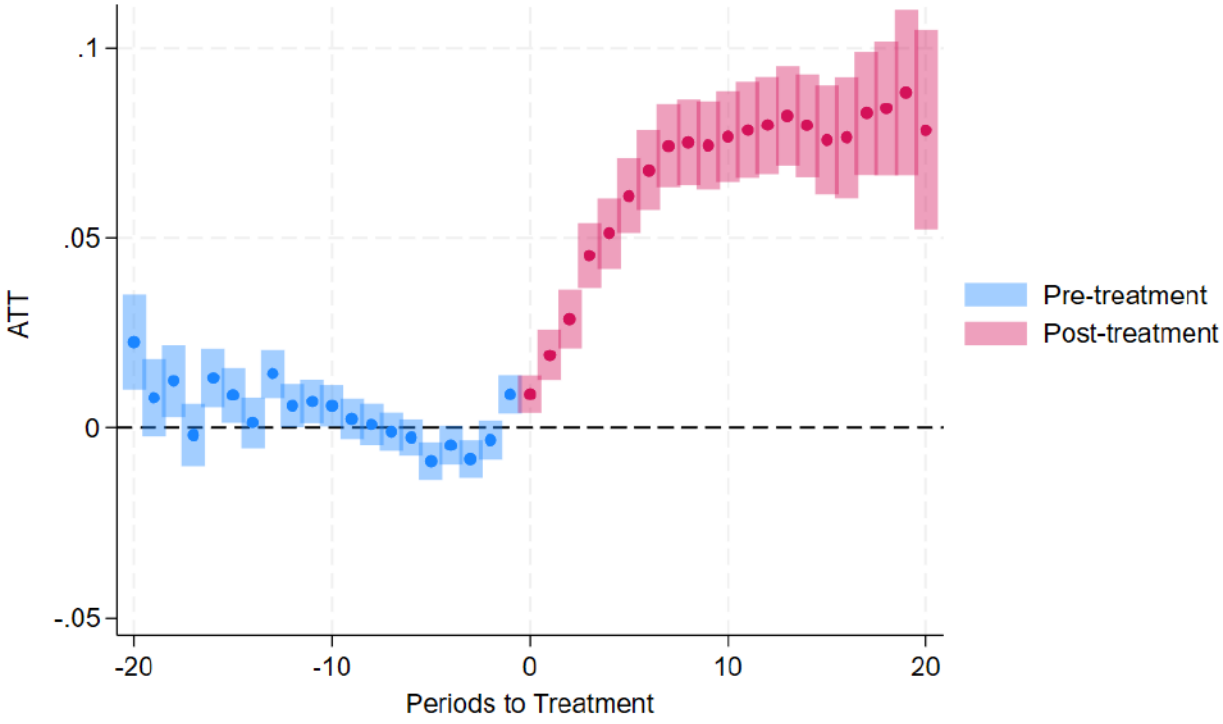
We first focus on the impact of STEP Ukraine on employment probability; here, a positive average treatment effect on the treated group indicates that STEP Ukraine participants were more likely to become employed following their participation in STEP Ukraine.

Impact on employment probability

Figure 4.1 presents the average treatment effect of STEP Ukraine on participants' likelihood of being in employment each month, for the 20 months prior to and after starting the programme. The average treatment effect on employment probability increases steadily during the first 7 months following programme start and then stabilises. At that time, 7 months after starting STEP Ukraine, participants are approximately 8 percentage points more likely to be in employment than they would have been in the absence of the programme. The time profile of these effects suggests that, broadly speaking, increases in employability arising from the STEP programme take up to 7 months to result in an actual move into employment. After that point, everyone whom the programme helps to secure employment has done so, and the impact therefore stabilises. If we take all 20 post-treatment periods together, including the 7-month period over which impacts are still building, then the average treatment effect is around 6.6 percentage points (see Table 4.4.

The estimated average treatment effect coefficients show the difference in the outcome between STEP Ukraine participants and the weighted comparison group (the 'counterfactual' of non-participation in the programme, at each month relative to the programme start. The coefficients presented on the Y-axis correspond to the percentage point change in employment probability among the treated group.

Figure 4.1 Average effect of STEP Ukraine on employment probability of participants



Note: Event study plots show the average treatment effect on the treated and confidence interval at each month, relative to the programme start. For example, Period to Treatment 0 is August 2023 for the first student cohort, and July 2024 for the last cohort included in this evaluation.

Table 4.4 Post-treatment average effects on employment probability

Average effect on employment probability	Coefficient	Standard error
Post-treatment average effect (0 to 20 months)	0.066***	0.005

Note: Statistical significance is denoted as follows: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Employment probability pre-treatment trends

A visual inspection of the pre-treatment trends, in blue, shows evidence of some differences in trends between treatment and controls in some individual periods prior to treatment (see Figure 4.1). This is mostly the case in earlier periods, more than a year prior to treatment, although estimates are also less precise at that point. In the months immediately preceding treatment, there is also tentative evidence of a phenomenon known as “Ashenfelter’s dip” – there is a small deterioration in the employment rate of the treatment group relative to the control. This dip may reflect that some participants enter the STEP programme in response to a deterioration in their labour market prospects. Overall, there is little sign of a substantial systematic difference in trends prior to treatment.

Our formal statistical pre-treatment trends tests, presented in Table 4.5, essentially reflect these points. The details of the tests are included in the notes to the table, but the key points are as follows. The strictest pre-treatment trends test (that there is never any difference in trends in any individual pre-treatment period is rejected. The less strict test (that there are no systematic differences in trends on average over the pre-treatment period passes when taking the 12 months prior to treatment. It fails when including all 20 months prior to treatment, but even then, the economic magnitude of the difference in trends is very small relative to post-treatment differences.

Table 4.5 Pre-treatment trend analysis for employment probability

Employment probability pre-treatment trend test: Chi^2	**P-value**	**-**
20 months before treatment	0.000	-
12 months before treatment	0.000	-
Employment probability pre-treatment trend test: Average	**Coefficient**	**Standard error**
20 months before treatment	0.004***	0.001
12 months before treatment	0.000	0.001

Note: Estimates of the Wald Chi^2 statistic provided in the top panel test the null hypothesis that all pre-intervention ATTs are equal to zero. A significant Chi^2 statistic indicates that there was a significant difference between the treatment and control groups in at least one pre-intervention period. Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

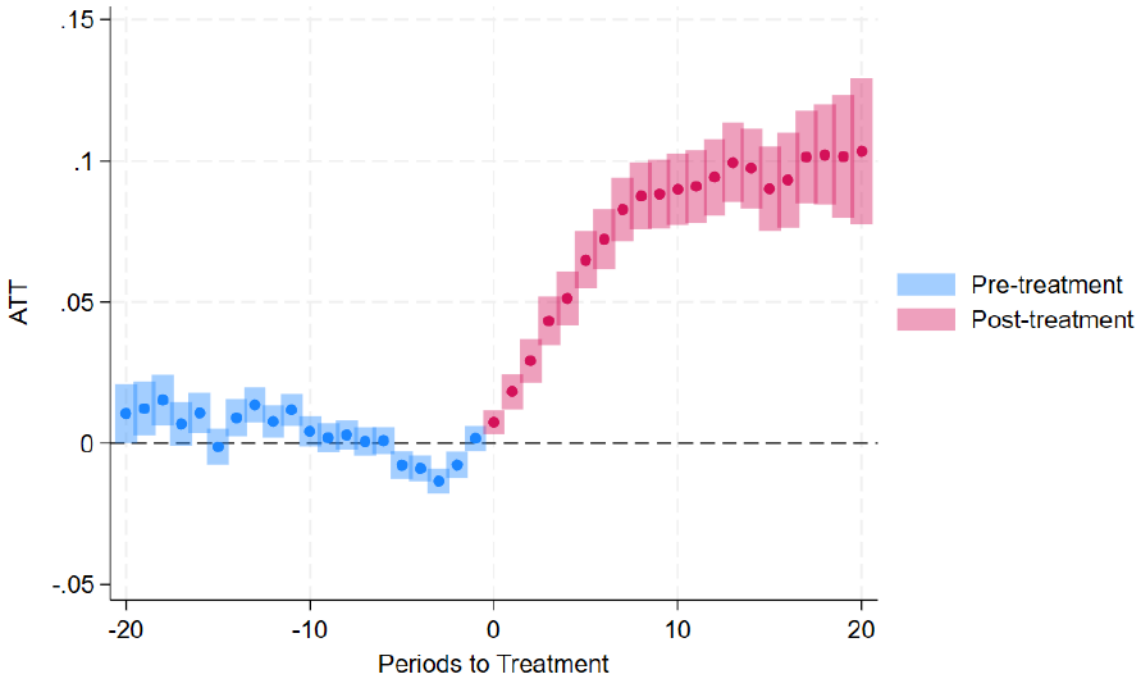
Employment sustainability

Next, we turn our attention to whether the employment of STEP Ukraine participants was sustainable. We test whether there is a higher probability of STEP Ukraine participants being employed for a period of at least 3, 6, and 12 consecutive months after they started the programme. It is important to note that the matched dataset we received from HMRC did not include any identifiable information about individuals' employment in order to maintain confidentiality. Due to the lack of information about job title and employer from this data, we cannot confirm that an individual appearing as employed in consecutive months remains in the same employment for the spell.

Impact on 3-month employment sustainability

Figure 4.2 shows the average effect of participating in STEP Ukraine on the treated group's likelihood of being in 3 months of sustained employment. Probability increases steadily in the 8 months after starting the programme. The average effect then starts to flatten out while still gradually increasing, showing a sustained effect across the 20 months after the start of the programme. In the 18 months after starting STEP Ukraine, the likelihood of participants being in 3 months of continuous employment is 10.2 percentage points higher than the counterfactual.

Figure 4.2 Average effect of STEP Ukraine on employment sustainability (3 months) of participants



The average impact of STEP Ukraine on employment sustainability over the full 20-month post-treatment period is summarised in Table 4.6. During the full 20 months following the start of the programme, the programme increases the participants' likelihood of being in 3 months or more of sustained employment by 7.7 percentage points compared to if they had not participated.

Table 4.6 Post-treatment average effects on employment sustainability

Average effect on employment sustainability	Coefficient	Standard error
3-month employment sustainability (0 to 20 months)	0.077***	0.005

Note: Statistical significance is denoted as follows: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Impact on 6-month employment sustainability

The effect of STEP Ukraine on participants' likelihood of being in 6 months or more of sustained employment is shown in Figure 4.3. Similar to the 3-month employment sustainability results (see Figure 4.2), we see a steady increase after the start of the programme. As would be expected with using the stricter definition of employment sustainability, the incline for the 6-month employment sustainability outcome is more gradual. Around 14 months after the start of the programme, the positive effect starts to stabilise. By 18 months after the start of STEP Ukraine, participants are 10.2 percentage points more likely to be in 6 months or more of sustained employment compared to the counterfactual. The average impact of STEP Ukraine on 6 months of sustained employment over the full 20-month post-treatment period is summarised in Table 4.7. During the full 20 months following the start of the programme, the programme increases the participants' likelihood of being in 6 months or more of sustained employment by 6.5 percentage points.

Figure 4.3 Average effect of STEP Ukraine on employment sustainability (6 months) of participants

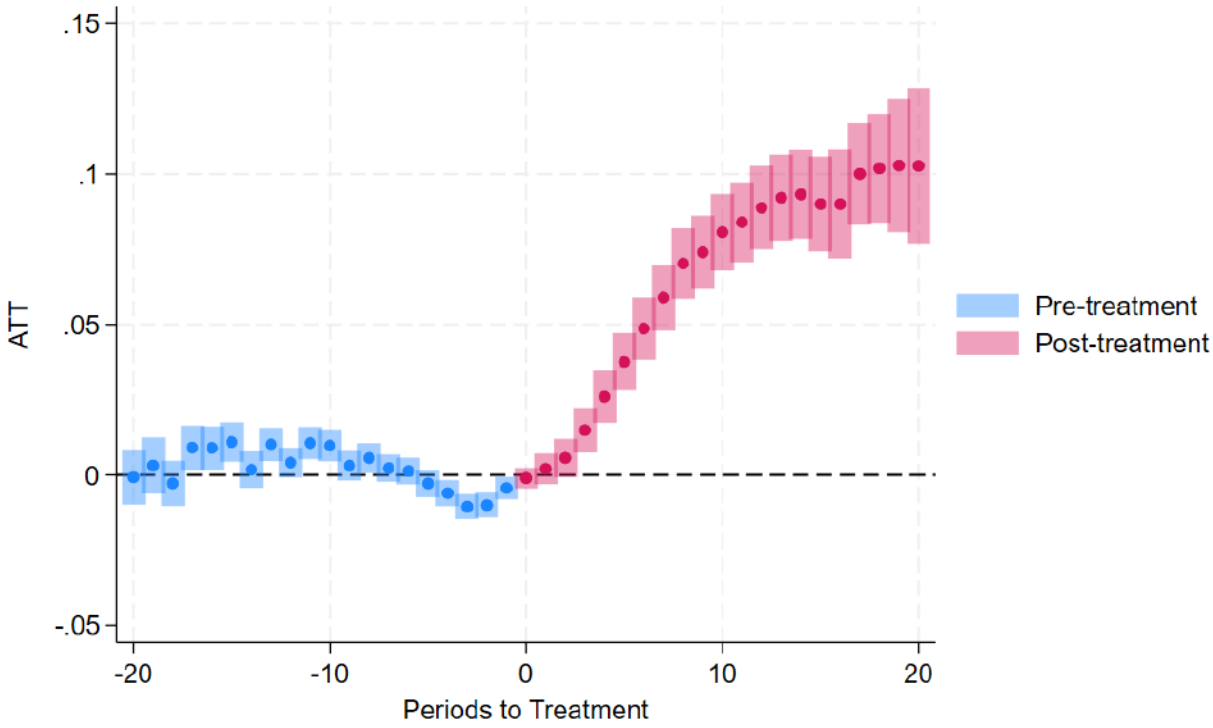


Table 4.7 Post-treatment average effects on 6-month employment sustainability

Average effect on employment sustainability	Coefficient	Standard error
6-month employment sustainability (0 to 20 months)	0.065***	0.005

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

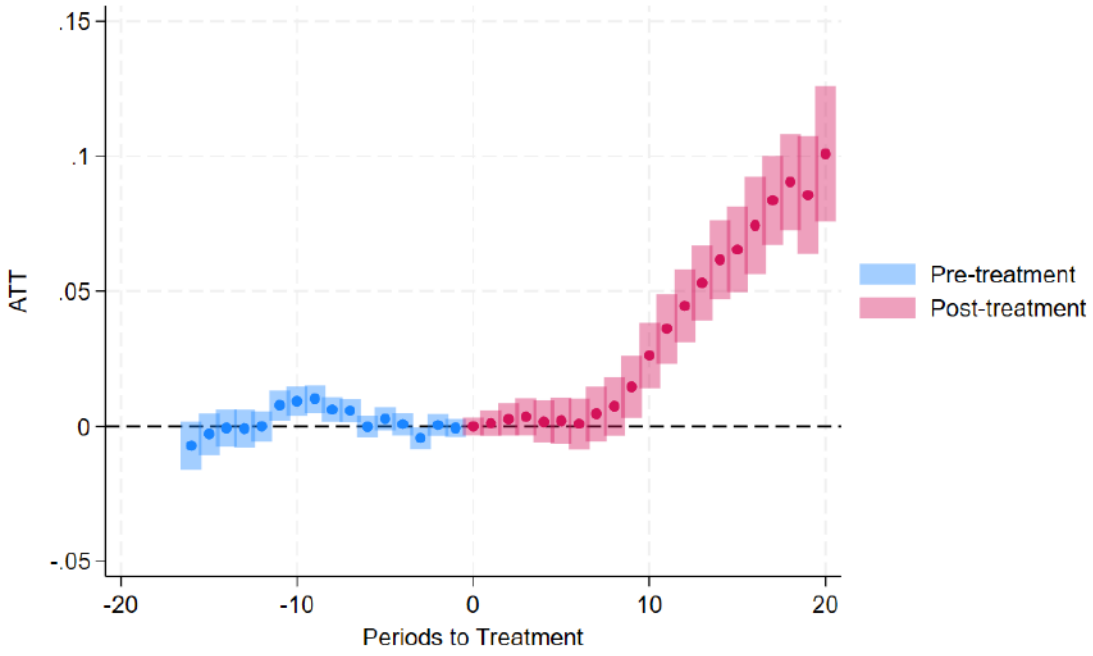
Impact on 12-month employment sustainability

Unfortunately, as shown in Table 4.9, we are unable to establish parallel trends between the treatment and control groups for the 12-month sustained employment outcome. Both the stricter and weaker tests of pre-treatment trends are rejected – i.e., our results suggest that there are differences both in the trends in individual pre-treatment periods, and in the trends on average over the pre-treatment window. As a result, we strongly advise caution over the interpretation of the following results, which come with the caveat that the underlying parallel trends assumption does not hold.

Figure 4.4 presents the effect of STEP Ukraine on longer-term employment sustainability, defined by 12 or more months of sustained employment. The effect of the programme on this outcome starts to increase after 7 months, with an expected longer lead time observed for the longer sustained employment period analysed. From 7 months after the programme's start, the effect on employment sustainability starts to increase steadily up to 18 months. The panel duration and the longer lead time required to observe this outcome mean it is not possible to observe whether the

effect of STEP Ukraine on 12 months of sustained employment levels out. However, the programme clearly has a strong positive effect. At 20 months after the start of the programme, STEP Ukraine participants are 10.1 percentage points more likely to be in 12 months or more of sustained employment compared to the counterfactual.

Figure 4.4 Average effect of STEP Ukraine on employment sustainability (12 months) of participants



The average impact of STEP Ukraine on employment sustainability over the full 20-month post-treatment period is summarised in Table 4.8. The programme increased participants' likelihood of being in 12 months or more of sustained employment by 3.6 percentage points.

Table 4.8 Post-treatment average effects on 12-month employment sustainability

Average effect on employment sustainability	Coefficient	Standard error
12-month employment sustainability (0 to 20 months)	0.036***	0.005

Note: Statistical significance is denoted as follows: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Employment sustainability pre-treatment trends

A visual check of the pre-treatment trends for the 3-month (see Figure 4.2) and 6-month (see Figure 4.3) sustained employment outcomes shows a similar pattern to the pre-treatment trends observed for employment probability (see Figure 4.1). Both pre-treatment trends have an "Ashenfelter's dip," but there is no substantial systematic difference in pre-treatment trends between the treated and control groups. Figure 4.4, showing pre-treatment trends for 12-month employment sustainability, displays a consistent pre-treatment trend, especially in the 6 months leading up to the introduction of the STEP Ukraine programme.

We also tested the pre-treatment trends of the 3-, 6-, and 12-month sustained employment outcomes statistically (see Table 4.9. The stricter test indicates that we do not have parallel trends for any of the sustained employment outcomes, suggesting that there is a difference in the trends of the treatment and control groups in some pre-treatment periods. The less strict parallel trend test holds for the 3- and 6-month sustained employment outcome, although not for the 12-month sustained outcome. Therefore, we advise caution over the interpretation of the 12-month sustained employment outcome results (due to lack of evidence of parallel trends from both tests, whereas the 3- and 6-month sustained employment results are likely to be more robust (as they pass the weaker parallel trend test.

Table 4.9 Pre-treatment trend analysis for employment sustainability

3-month sustainability pre-treatment trend test: Chi^2		**P-value**
20 months before treatment		0.000
12 months before treatment		0.000
3-month sustainability pre-treatment trend test: Average	**Coefficient**	**Standard error**
20 months before treatment	0.004***	0.000
12 months before treatment	-0.001	0.001
6-month sustainability pre-treatment trend test: Chi^2		**P-value**
20 months before treatment		0.000
12 months before treatment		0.000
6-month sustainability pre-treatment trend test: Average	**Coefficient**	**Standard error**
20 months before treatment	0.004***	0.001
12 months before treatment	0.000	0.001
12-month sustainability pre-treatment trend test: Chi^2		**P-value**
20 months before treatment		0.000
12 months before treatment		0.000
12-month sustainability pre-treatment trend test: Average	**Coefficient**	**Standard error**
20 months before treatment	0.002**	0.001
12 months before treatment	0.003***	0.001

Note: Statistical significance is denoted as follows: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Monthly wages

Impact on monthly wages

Figure 4.5 presents the average effect on the treated sample's wages conditional on being employed, for those who were employed at baseline, using inflation-adjusted real pay in logs. Real pay in logs provides a normalised outcome, which can be interpreted as an approximate percentage change in monthly wages. For example, a 0.1 change in the log pay outcome is roughly equivalent to a 10% change in monthly wages. We cannot distinguish between hours of work and pay per hour in the HMRC data, so any impacts could come either from changes in hours (e.g., movements between part-time and full-time work or progression in hourly pay or a combination of the two. On average, our estimates of post-treatment average effects (see Table 4.10 suggest that participation in STEP among those already in employment increased monthly wages by approximately 10.7% in the 20 months post-intervention. There is less evidence than for employment that the effect had reached its peak within this window.

Figure 4.5 Average effect of STEP Ukraine on monthly wages (real pay in logs) of participants

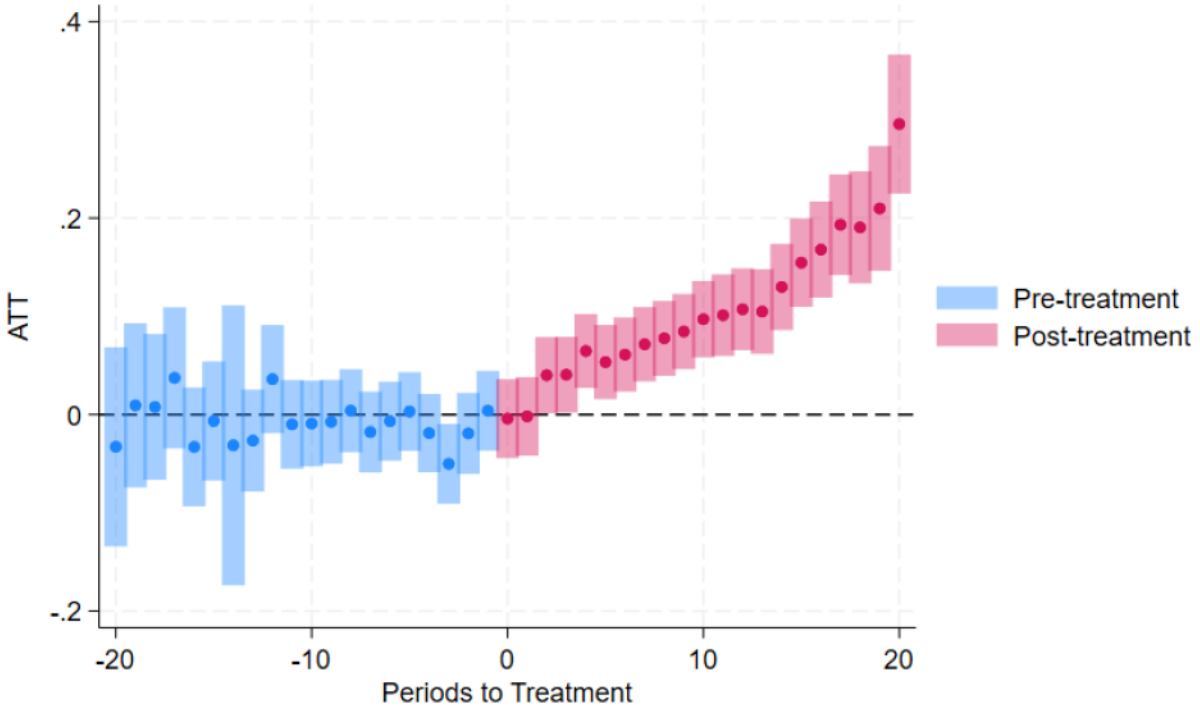


Table 4.10 Post-treatment average effects on monthly wages (real pay in logs)

Average effect on employment probability	Coefficient	Standard error
Post-treatment average effect (0 to 20 months)	0.107***	0.015

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Monthly wages pre-treatment trends

The visual check on pre-treatment trends for the monthly wages outcome (see Figure 4.5) suggests that the parallel trends assumption holds, as there are only very small movements around 0 (which indicates no difference between the treatment and control group). We again tested the parallel trends using both our stricter and weaker statistical tests for the monthly wages outcome (see Table 4.11). Our stricter test reveals that there are some differences between the treatment and control group in some individual pre-treatment periods (Figure 4.5 suggests that this is likely to be 3 months before the treatment began). However, our weaker test reveals that there are likely parallel trends over a longer pre-treatment window. While this test fails to show parallel trends over the 12 months leading up to the introduction of STEP Ukraine, where we find a small, significant difference between the treatment and control groups, it does show parallel trends over the whole 20-month period leading up to STEP Ukraine. Therefore, we advise some caution on over-interpretation of these results due to the lack of evidence of parallel trends in both the stricter and weaker tests in the 12 months leading up to STEP Ukraine.

Table 4.11 Pre-treatment trend analysis for monthly wages (real pay in logs)

Employment probability pre-treatment trend test: Chi^2	P-value
20 months before treatment	0.000
12 months before treatment	0.000

Employment probability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	-0.008	0.004
12 months before treatment	-0.008***	0.002

Note: Estimates of the Wald Chi^2 statistic provided in the top panel test the null hypothesis that all pre-intervention ATTs are equal to zero. A significant Chi^2 statistic indicates that there was a significant difference between the treatment and control groups in at least one pre-intervention period. Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

4.3 Impacts in Scotland, Wales, and Northern Ireland

Characteristics of STEP participants in devolved administrations

The characteristics of STEP Ukraine participants included in this analysis in Scotland, Wales, and Northern Ireland are presented in Table 4.12. Across all three nations, the samples were predominantly comprised of women, as at the UK level. The share of men among STEP participants was, however, marginally higher in Scotland (23.2% and lower in Wales (17.6% than the UK average (19.1%. STEP participants across the devolved administrations broadly followed the UK-level age distribution. In July 2023, the highest share of STEP participants across Scotland, Wales, and Northern Ireland was in the age bracket 35-44. This trend was particularly pronounced in Wales, where 48.8% of the sample fell in the 35-44 age range, compared with 45.2% at the UK level. In addition, there was a difference in the distribution of age in the Northern Ireland sample, where 7.5% of individuals were aged 55-65; this is compared with the UK level (5.0%.

Arrival time in the UK also differed for the sample of STEP participants from Northern Ireland, among whom 17.7% arrived in the UK after July 2023 (see Table 4.12. This compares to 10.4% across the UK as a whole, 8.9% in Scotland and 7.5% in Wales. Furthermore, the sample of STEP participants in Northern Ireland who were in the UK by July 2023 had typically spent less time in the UK than those in other devolved administrations – in Northern Ireland, 15.5% of the sample had been in the UK 0-6 months by July 2023, whereas in Scotland and Wales only 7.3% and 7.7% of the sample fell into this bracket. This is also reflected in the sample who had been in the UK 13-18 months by July 2023, which represents 27.8% of the Northern Irish sample, but 40.7% of the Scottish sample, 53.0% of the overall UK sample, and 57.0% of the Welsh sample.

Table 4.12 Characteristics of STEP Ukraine participants in devolved administrations by sex, age group and time since arrival in the UK

Descriptive characteristics	Category	Northern Ireland sample	Northern Ireland %	Scotland sample	Scotland %	Wales sample	Wales %
Sex	Men	37	19.8	267	23.2	73	17.6
Sex	Women	150	80.2	886	76.8	343	82.5
Age group	24 or under	55*	29.4*	96	8.3	32	7.7
Age group	25-34	55*	29.4*	253	21.9	73	17.6
Age group	35-44	87	46.5	498	43.2	203	48.8
Age group	45-54	31	16.6	253	21.9	89	21.4
Age group	55-65	14	7.5	53	4.6	19	4.6
Time since arrival in the UK (July 2023)	Arrived after July 2023	33	17.7	103	8.9	31	7.5
Time since arrival in the UK (July 2023)	0-6 months	29	15.5	84	7.3	32	7.7
Time since arrival in the UK (July 2023)	7-12 months	52	27.8	444	38.5	105	25.2
Time since arrival in the UK (July 2023)	13-18 months	56	30.0	469	40.7	237	57.0
Time since arrival in the UK (July 2023)	Over 19 months	17	9.1	53	4.6	11	2.6

*Note: Time since arrival in the UK is determined as of July 2023. * denotes where disclosive statistics are pooled together in the case of very small sample sizes.*

Table 4.13 presents descriptive employment statistics of the sample of STEP Ukraine participants at the baseline of July 2023, i.e., just before the first month of the programme occurred in August 2023. We note that the treated sample size is small in Wales and especially so in Northern Ireland. Among the devolved administrations, employment rates were highest for STEP Ukraine participants in Wales prior to starting the programme, with a mean of 0.560. A much smaller share of participants was employed in Scotland prior to the programme's start, at 0.426. For both Scotland and Wales, the employment rates among men and women participating in STEP Ukraine were relatively consistent. However, in Northern Ireland, women participating in STEP Ukraine were much less engaged in

employment compared to men just before the programme started, with employment rates of 0.492 and 0.607, respectively.

Average monthly earnings among participants were highest in Northern Ireland, though these again reflect a distinct inequality between men and women participating in the programme. After Northern Ireland, average monthly earnings were next highest in Wales, followed by Scotland. Scotland's participants in STEP Ukraine similarly had a significant gendered distinction in earnings, with an average of £1,041.31 among male participants compared to £564.48 among female participants. In Wales, employment rates and monthly earnings were more consistent across both men and women participating in the STEP Ukraine programme. The descriptive statistics show a very different employment picture across the three devolved administrations just prior to the start of STEP Ukraine for participants.

Table 4.13 Baseline employment and monthly earnings of STEP Ukraine participants in devolved administrations

Devolved admini-stration	Treated sample size	Total employ-ment*	Men employ-ment*	Women employ-ment*	Total monthly earnings**	Men monthly earnings**	Women monthly earnings**
Northern Ireland	187	0.514	0.607	0.492	828.13	1208.29	737.93
Scotland	1,153	0.426	0.458	0.417	663.65	1041.31	564.48
Wales	416	0.560	0.550	0.562	763.89	918.91	735.01

*Note: The baseline employment statistics only include participants of STEP Ukraine who were in the UK and aged 18 or above in July 2023. * denotes that the figures presented are mean employment, where 1 = employed. ** denotes that the figures presented are mean monthly earnings in £.*

Employment impacts in devolved administrations

We also examined the question of whether impacts varied across the devolved administrations. We were unfortunately able to address this only partially – in particular, for Scotland.

The devolved administration analysis was conducted using the treated sample from the devolved administration in question and the control group from the whole of the UK. We did not conduct the analysis using a control group from the specific devolved administration. This is because of the nature of the dataset, whereby HMRC PAYE records containing region/country data were not available for individuals in the control group who were never employed during the analysis window. To limit the control group to individuals within the specific devolved administration, and therefore exclude individuals without region/country data, would be to exclude individuals who have never been in employment from the control group. This would bias the control group by only including individuals who have employment at some point in the panel.

Impacts in Scotland

The estimated impacts of STEP Ukraine on employment are broadly the same in Scotland as in the UK as a whole. We note that the confidence intervals for Scotland – the largest of the devolved nations – are substantially larger than for the UK, indicating that there is less certainty about the accuracy of the impact estimates. Despite this, the estimated impact on employment in Scotland is statistically significant and follows a similar path to the impact that we estimate in the UK, as shown in Figure 4.6. Table 4.14 summarises the average post-treatment effect of the programme in Scotland. We estimate that STEP Ukraine increased employment probabilities among participants in Scotland by 6.0 percentage points on average over the 20 months after starting.

Figure 4.6 Average effect of STEP Ukraine on employment probability in Scotland

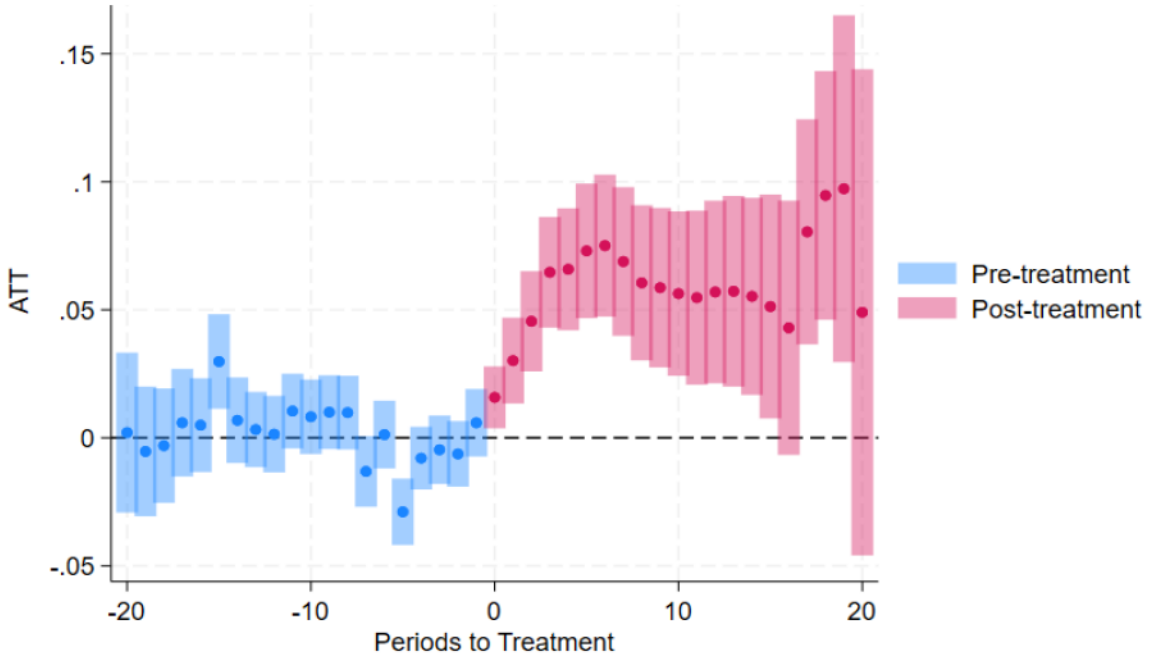


Table 4.14 Post-treatment average effects on employment probability in Scotland

Average effect on employment probability	Coefficient	Standard error
Post-treatment average effect (0 to 20 months)	0.060***	0.014

Note: Statistical significance is denoted as follows: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

The assessment of common trends is also similar to the UK-level analysis, with the strictest test rejected but the weaker test passing, as shown in Table 4.15. This indicates that although there were deviations in some individual time-periods before the intervention began (tested by the Chi^2 statistic), overall, the two groups followed similar trends in the lead up to STEP Ukraine.

Table 4.15 Pre-treatment trend analysis for employment probability in Scotland

Employment probability pre-treatment trend test: Chi²	P-value
20 months before treatment	0.000
12 months before treatment	0.000

Employment probability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	0.002	0.002
12 months before treatment	-0.001	0.001

Note: Estimates of the Wald Chi² statistic provided in the top panel test the null hypothesis that all pre-intervention ATTs are equal to zero. A significant Chi² statistic indicates that there was a significant difference between the treatment and control groups in at least one pre-intervention period. Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

For employment sustainability in Scotland, similar trends are again observed to the UK-level effects in Figures 4.7, 4.8, and 4.9.

Figure 4.7 Average effect of STEP Ukraine on employment sustainability (3 months) of participants in Scotland

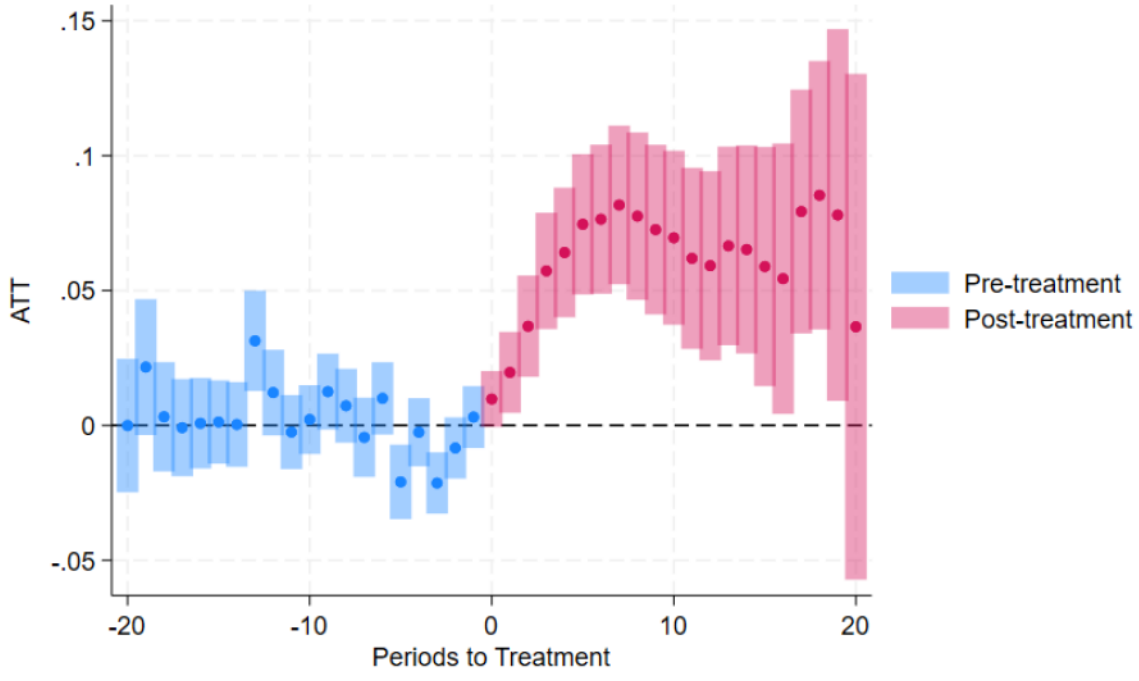


Figure 4.8 Average effect of STEP Ukraine on employment sustainability (6 months) of participants in Scotland

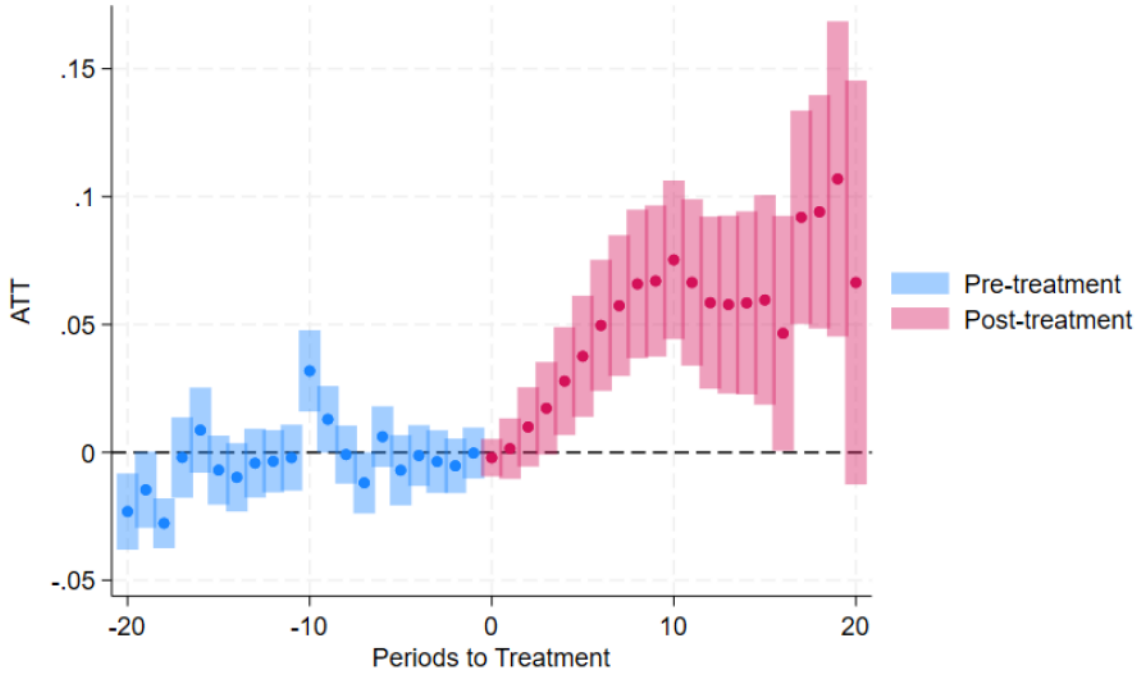


Figure 4.9 Average effect of STEP Ukraine on employment sustainability (12 months) of participants in Scotland

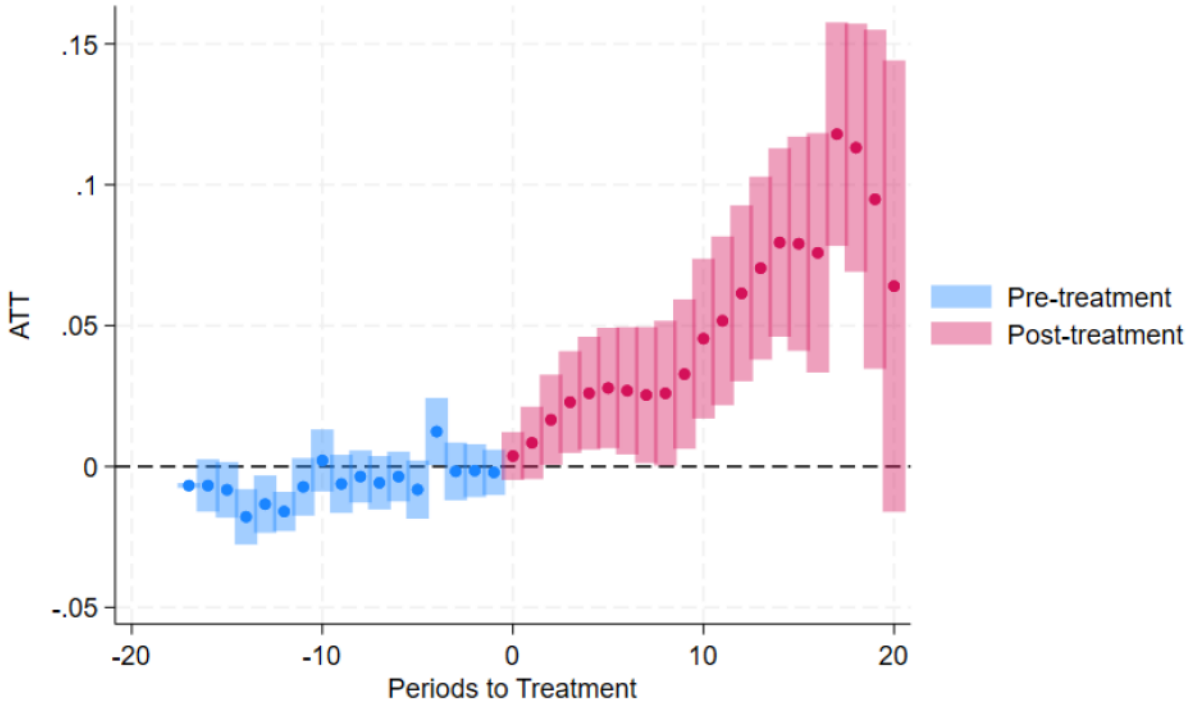


Table 4.16 summarises the positive and statistically significant effects of STEP Ukraine on the probability of participants in Scotland being in sustained employment for 3, 6, and 12 months or more. We find evidence of a positive and statistically significant effect of STEP Ukraine on continuous employment for 3, 6, and 12 months; however, we advise caution about over-interpretation of the results for 12 months as explained below.

Table 4.17 summarises the formal pre-treatment trends analysis. As with the UK analysis, the stricter test fails to establish parallel trends for all employment sustainability outcomes in Scotland. For 3-month and 6-month employment sustainability, the weaker tests display no statistically significant differences pre-treatment between the treated and control groups. However, for the 12-month employment sustainability analysis, the weaker test also fails to establish parallel trends, meaning the 12-month results should be interpreted much more cautiously.

Table 4.16 Average effect on employment sustainability in Scotland

Average effect on employment sustainability	Coefficient	Standard error
3-month employment sustainability (0 to 20 months)	0.061***	0.014
6-month employment sustainability (0 to 20 months)	0.053***	0.129
12-month employment sustainability (0 to 20 months)	0.051***	0.012

Note: Statistical significance is denoted as follows: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Table 4.17 Pre-treatment trend analysis for employment sustainability in Scotland

3-month sustainability pre-treatment trend test: Chi²		**P-value**
20 months before treatment		0.000
12 months before treatment		0.000
3-month sustainability pre-treatment trend test: Average	**Coefficient**	**Standard error**
20 months before treatment	0.002	0.001
12 months before treatment	-0.001	0.478
6-month sustainability pre-treatment trend test: Chi²		**P-value**
20 months before treatment		0.000
12 months before treatment		0.000
6-month sustainability pre-treatment trend test: Average	**Coefficient**	**Standard error**
20 months before treatment	-0.003**	0.001
12 months before treatment	0.001	0.001
12-month sustainability pre-treatment trend test: Chi²		**P-value**
20 months before treatment		0.000
12 months before treatment		0.000
12-month sustainability pre-treatment trend test: Average	**Coefficient**	**Standard error**
20 months before treatment	-0.006***	0.001
12 months before treatment	-0.003**	0.001

Note: Estimates of the Wald Chi² statistic provided test the null hypothesis that all pre-intervention ATTs are equal to zero. A significant Chi² statistic indicates that there was a significant difference between the treatment and control groups in at least one pre-intervention period. Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

The analysis of the impact of STEP Ukraine on monthly wages in Scotland did not produce a robust result. Estimated effects were not statistically significant, but parallel pre-treatment trends could not be established through either the stricter or weaker statistical tests. Those tests are included in Appendix 4.

Overall, we found significant evidence of a positive impact of STEP Ukraine on employment probability (around 6 percentage points higher for STEP Ukraine participants) and on continuous employment over a 3-, 6-, and 12-month period after participation in Scotland. However, we advise caution over interpretation of the 12-

month continuous employment outcome, as we were not able to full establish parallel trends pre-intervention. We were also not able to establish parallel trends for the monthly wages outcome for Scotland and therefore do not present results for this outcome.

Impacts in Northern Ireland and Wales

It unfortunately proved difficult to reach robust conclusions about the devolved administrations beyond Scotland, due to low sample sizes and more difficulty establishing robust control groups than in the UK-level analysis. In Wales, common pre-trends could not be established on either the more stringent or the weaker basis for most outcomes. The lower sample size limited the options available for addressing this (e.g., excluding those who never participate in STEP at any point, who may be more different from participants than those who are simply yet to participate, from the control group. The statistical tests for common trends for these analyses are included in Appendix 4. In Northern Ireland, the sample size was simply too low to draw meaningful conclusions (confidence intervals spanned the impacts that we estimate for the UK, but also significant negative values.

4.4 Conclusion

Impact of STEP Ukraine on employment outcomes

STEP Ukraine seemingly had a positive impact on employment probability. Over the 20 months following the start of the programme, the likelihood of being employed increased by 6.6 percentage points for STEP Ukraine participants compared to if they had not participated in the programme. STEP Ukraine also had a positive impact on employment sustainability. STEP Ukraine participants were 7.7 percentage points more likely to be in 3 months or more of sustained employment, 6.5 percentage points more likely to be in 6 months or more of sustained employment during the 20 months following the start of the programme. Not only did STEP Ukraine help people find employment, but it also helped them to stay employed for longer. STEP Ukraine also increased participants' monthly wages. The analysis looked at monthly pay in logs and found that the programme had a significant positive effect on earnings, which increased by 11% over the 20-month period.

The analysis showed that these employment impacts took time to take effect. For employment probability, the effect increased steadily over the first 7 months after participants started the 3-month STEP Ukraine programme before stabilising. It is worth noting that STEP participants receive a 12-month action plan upon completion of the programme, and that they have access to the online portal for 6 of these 12 months. Therefore, participants are likely to follow these action plans and take around 7 months to experience the maximum impact of participating in STEP Ukraine.

For monthly wages, the effect continued to increase over the 20-month period of analysis after starting the programme. This suggests that our analysis may not capture the full extent of the impact of STEP Ukraine on earnings. This is likely because it takes time for language and employability skills to develop and be put into practice.

Similar effects for employment probability and employment sustainability to the UK-wide effects are found for Scotland. In Scotland, STEP Ukraine resulted in a 6.0 percentage point increase in the likelihood of being employed for participants, compared to the counterfactual. However, while the analysis of monthly wages for STEP Ukraine participants in Scotland also estimates a positive effect, the common trends do not hold, so the robustness of the result is limited.

Primarily due to the limited sample size of the treated group in Wales and Northern Ireland, it was not possible to produce robust and meaningful analysis at the level of devolved administrations. For future evidence development of the impact of language and employment support across devolved administrations of the UK, it is important that the delivery of future programmes is scaled up to meaningfully reach populations across Scotland, Wales, and Northern Ireland.

Limitations of the analysis

Accompanying the considerable evidence presented of the positive impacts of STEP Ukraine on employment outcomes of participants, there are some caveats to the employment impact analysis that we need to consider:

- None of the analyses for the three outcomes satisfied the parallel trends assumption using the stricter test. In some cases, we find significant deviations just before the programme starts, indicating there may be a small anticipatory effect of participating in STEP Ukraine and a risk of selection bias.
- For the employment sustainability analysis, we do not know if people stayed in the same employment during their spell of sustained employment. This is due to the nature of the HMRC PAYE data we received, with all identifiable information removed, so the analysis was based on gross monthly earnings. It could be that individuals in 'sustained' periods of employment over consecutive months are moving between many jobs, but have some income in each month regardless.
- The DA analysis includes the never-employed cohort in the control group, but this means that it also includes non-participants who live in other parts of the UK (i.e., outside the DA being analysed).
- The analysis window spans 20 months after the start of the programme. As a result, it is not possible to examine the lasting impact of the programme. This is especially the case for the monthly wages and 12-month employment sustainability outcomes, where the event studies show a positive effect that continues to increase until the end of the 20-month analysis window.
- We are not able to rule out that STEP Ukraine participants took part in other interventions that aimed to provide employability or English language support. However, we find significant positive effects of STEP Ukraine overall across the group of STEP participants and believe it is unlikely that all participants received structured external support.

Overall, the employment impact evaluation suggests that STEP Ukraine had a positive effect on participants' ability to find employment, stay in sustained employment, and earn higher wages.

5 The Economic Evaluation

5.1 Methods

The economic evaluation examined the Value for Money (VfM) of the STEP Ukraine programme by comparing the costs against the benefits of the intervention. Benefits were quantified both from the narrower perspective of the Exchequer (i.e., what was the effect on the public finances?) and from the wider perspective of society as a whole (i.e., what was the value of the economic benefits generated?).

The modelling examined the benefits of two principal labour market outcomes of the programme: the movement of unemployed participants into employment, and the uplift in wages for those participants already in employment. Sensitivity analysis was also implemented to reflect uncertainty in some parameters.

The key findings and takeaways of the employment impact evaluation are summarised in Box 5.1.

Box 5.1 Key findings of the economic evaluation

- **From the Exchequer's perspective, every £1 invested in the programme generates a gross saving of £2.04** (a net saving of £1.04 per £1 spent), according to our central scenario. Sensitivity analysis based on a plausible range for key assumptions indicates that the benefit-cost ratio (BCR) could vary between 0.88 and 3.27.
- **From a social perspective, every £1 spent on the programme delivers a gross benefit of £3.67** (a net benefit of £2.67 per £1 invested). Sensitivity analysis places this BCR between 1.59 and 5.84.

Model structure

The key steps in the economic modelling were as follows:

1. Use estimates from the impact evaluation on increases in employment probability to determine the number of STEP participants who moved into employment.
2. Use impact estimates of increased wages from participation in STEP (of those who were already employed as of entry into the programme) to calculate the average wage uplift.
3. Utilise DWP analysis to estimate per participant benefits for individuals unemployed at baseline who moved into employment.
4. Utilise DWP analysis to estimate per participant benefits for individuals employed at baseline who experienced a wage increase.
5. Aggregate the total benefits of movements into employment and uplifts in wages.

6. Compare the total benefits with the total programme costs from the perspectives of the Exchequer and of society at large.

7. Apply sensitivity analysis to key parameters to obtain central, low, and high estimates of programme Value for Money (VfM).

Employment probability

The model estimated the per participant benefits of the movement of STEP Ukraine participants into employment, drawing on DWP analysis.

From the perspective of the Exchequer, higher employment or earnings tends to mean higher tax revenue from taxes on earnings, and lower benefits spending due to reductions in entitlements to means-tested benefits. These represent gains for the public finances, to be weighed against the initial outlay on the programme.

From the perspective of society at large, there are benefits from the increase in productive economic activity caused by higher employment. Hence, employment impacts are translated into impacts on economic output, and the value of that additional output is counted as a society-wide benefit of the programme. Impacts of increased travel are regarded as an additional social cost, since individuals entering employment will travel more to reach their workplaces. These are compared to the costs of programme implementation. From the point of view of society as a whole, changes in taxation and benefit spending are not counted because they are effectively just transfers between different individuals (i.e., the payers of tax and the beneficiaries of public spending).

STEP participants who entered employment were broken down into three equally sized subgroups (“tertiles”) based on their post-treatment earnings. DWP analysis was then used to compute the net benefit from entering employment for an average participant in each subgroup.

The ONS survey data also implied that only 81% of unemployed STEP participants received benefits, with the remaining 19% not claiming any additional support. The DWP analysis was therefore adapted by removing all elements related to Universal Credit, enabling the modelling of impacts for STEP participants who were not claiming UC. Net benefits per participant were then again calculated for each of the three subgroups, excluding UC-related components.

Estimates of increased employment probability due to the STEP Ukraine programme came from the impact evaluation. The impact analysis found an increased probability of employment due to the STEP programme of 6.6 percentage points. This estimate of increased employment probability was used to derive the overall number of participants who moved from unemployment into employment within each subgroup. The average of net per participant benefits was then aggregated by the number of participants moved into employment to find the total benefits due to increased employment for the programme as a whole.

Increased wages

In addition to the benefits of the programme in moving unemployed participants into employment, the VfM analysis also explored the benefits of wage increases for those STEP participants already in employment at the start of the programme.

The impact analysis estimated an increase in log wages of 0.1066, which translates to an increase of 11.25%. This was then applied to average baseline pay among those in employment, to determine the average uplift in gross pay per month for STEP participants who were already in employment at the start of the programme.

Three equally sized subgroups were again formed based on baseline earnings. Within each group, average baseline earnings were calculated. The proportional uplift in earnings from our impact estimates was applied to that baseline. Finally, DWP analysis was used to determine the net benefits to the Exchequer of this change in earnings. The Exchequer benefits from higher tax revenues arising from increased income tax, National Insurance contributions, and indirect taxes, along with reduced UC claims resulting from the wage uplift. For society, the primary benefits include increased output, reflecting a projected rise in productivity aligned with the wage increase.

Average per participant net benefits were determined across the three subgroups, among both recipients and non-recipients of UC. This was then multiplied by the total number of participants who were employed at the start of the STEP programme (and remained employed to determine overall benefits due to the uplift in earnings.

Assumptions and sensitivities

To account for uncertainty in key model parameters, we conducted a sensitivity analysis to assess how variations in these assumptions affected our overall results. Specifically, we examined four elements of the model: (i the persistence of the treatment effect, (ii the estimate of the increase in probability of employment, (iii the estimate of the magnitude of wage uplift, and (iv the inclusion of a substitution effect. These parameters were varied to generate three scenarios: a central case, a low case, and a high case. Further details on the parameter adjustments and additional assumptions are provided below and in Appendix 5.

Persistence refers to the length of time for which the treatment group had better outcomes than they would have had without treatment. In some contexts, one might assume that the impacts of an intervention last forever, in which case one can generate very large BCRs. However, in a setting such as this, it is typical to be more conservative and assume a finite time horizon for the benefits.[158] The rationale is that, eventually, non-participants would reach the same employability levels as participants, e.g., because with more time spent in the UK, their English language

[158] For example, Scottish Government (2022) Fair Start Scotland: economic evaluation. Available at: https://www.gov.scot/publications/fair-start-scotland-economic-evaluation/.

skills would continue to improve. Therefore, the programme brings forward these improvements, but non-participants would still get to the same level eventually.

The estimate of persistence used is an important consideration as this parameter has a particularly large impact on the overall benefits of the programme and hence on its VfM. In the central estimate, persistence is assumed to be 20 months. This central estimate was derived from the impact analysis, which showed a sustained effect of the programme throughout the observation period up to 20 months. This parameter is subject to uncertainty because participants were not observed beyond 20 months post-treatment. We therefore lack evidence on how long the effect persists after this point. However, because it is unlikely that the treatment effect would have declined to zero immediately after the observation period, this central estimate may be considered conservative. In sensitivity analyses, we varied the persistence parameter by six months above and below the central estimate, yielding a lower bound of 14 months and an upper bound of 26 months.

Substitution effects refer to the fact that some participants may move into employment or receive higher wages by displacing existing workers. In economic terms, this happens if labour demand is not perfectly elastic. To account for this, we applied a 20% substitution effect, as suggested in DWP guidance for supply-side employment policies. In line with the guidance, this adjustment is used only in our sensitivity analysis and reduces overall benefits by 20% in the low-sensitivity scenario.

Our impact estimates on employment probability and wage uplift are statistically significant. However, as in any empirical analysis, the central point estimates are subject to uncertainty. We therefore vary these parameters in the sensitivity analysis. Specifically, we use the bounds of the 95% confidence interval, which contains the true coefficient with 95% confidence. In the low-sensitivity scenario, we apply the lower bounds of the confidence interval for both employment probability and wage uplift effects. In the high-sensitivity scenario, we apply the corresponding upper bounds.

Discounting and adjusting prices of costs and benefits was undertaken to determine their real present value. The prices of all flows of benefits and costs were adjusted to real prices using September 2025 prices as a baseline. These flows were adjusted using past Consumer Price Index (CPIH indices and future CPI forecasts. Most benefits and costs occurred in the past and therefore did not require discounting. However, the final programme payment, scheduled for May 2026, was discounted at a real rate of 3.5%, in accordance with HM Treasury's Green Book guidance.

Benefit-cost ratio

Total programme benefits were divided by total costs to derive BCRs representing the value accruing to each perspective per £1 of programme expenditure. This was done separately for the two different perspectives to yield both an Exchequer BCR and an overall social BCR. Tables 5.1 and 5.2 below display the components of benefits and costs from the perspectives of the Exchequer and of society, respectively.

Table 5.1 Exchequer benefits and costs

Benefits	Costs
Increased income tax	Programme costs
Increased employees' National Insurance contributions	-
Increased employers' National Insurance contributions	-
Increased indirect taxes	-
Reduction in Universal Credit	-
Reduction in operational costs	-

Table 5.2 Social benefits and costs

Benefits	Costs
Increased Output	Programme costs
Reduction in operational costs	Impacts from increased travel

5.2 Results

This section outlines and discusses the results from our VfM analysis of the STEP Ukraine programme. This will examine the results separately from the perspective of the Exchequer first, and later from the perspective of society.

Perspective of the Exchequer

The results from the perspective of the Exchequer are positive, with our central estimates suggesting that every £1 spend on the programme leads to a gross savings of £2.04 (net savings of £1.04 per £1 spent) to the Exchequer (see Table 5.3). This translates to total net savings of £8.9 million. Sensitivity analysis suggests that under a low-case scenario, the BCR could fall to 0.88, while in a high-case scenario, it could rise as high as 3.27.

Table 5.3 Results from perspective of the Exchequer by scenario

Exchequer	Central	Low	High
Total present value of benefits	£17.5m	£7.6m	£28.1m
Net present value of programme costs	£8.6m	£8.6m	£8.6m
Net present value	£8.9m	-£1.0m	£19.5m
BCR	2.04	0.88	3.27

Perspective of society

The results clearly indicate that the STEP Ukraine programme was a net positive to society, with every £1 spent on the programme translating to a £3.67 gross benefit (net benefit of £2.67 per £1 spent to society (see Table 5.4. This translates to an overall social net benefit of £22.9 million. Sensitivity analyses indicate that society experiences a net benefit across all scenarios, with the BCR ranging from a conservative estimate of 1.59 to a high estimate of 5.84.

Table 5.4 Results from perspective of society by scenario

Society	Central	Low	High
Total present value of benefits[159]	£31.5m	13.7m	£50.2m
Total present value of programme costs	£8.6m	£8.6m	£8.6m
Net present value	£22.9m	£5.1m	£41.6m
BCR	3.67	1.59	5.84

Limitations

It is important to note some limitations of the analysis:

- Assumptions were required regarding the impacts of the STEP Ukraine programme beyond those that could be directly estimated from our quasi-experimental impact evaluation. Key examples include the persistence of impacts and the magnitude of any substitution effects. Where guidance was available within DWP analysis, we applied the relevant standard assumptions. Where uncertainty remained, such as for our parameter of persistence, we made informed assumptions based on available evidence and the results of the impact analysis. We also conducted sensitivity analysis, varying these parameters to assess how their uncertainty affected our findings.

- While the model aimed to capture the principal costs and benefits arising from the STEP Ukraine programme, we cannot be certain that it accounted for all potential impact channels. There are likely additional costs and benefits that were not quantified and therefore not included in the analysis. For example, beyond increased employment probability, higher wages, and the associated rise in output, the programme may also have generated other important effects, such as improvements in health or wellbeing, which would represent additional benefits to society.

[159] Note that costs from increased travel are accounted for within the benefits (i.e., they reduce the net benefit), as in the DWP analysis framework – the "costs" that are relevant for the BCR are the costs of the programme itself.

- The DWP analysis used within the evaluation models the impact on the Exchequer of changes to UC and taxation. However, the model does not include devolved benefits nor non-UC payments administered by DWP. While this represents a limitation since not all benefit types are captured, UC is the UK's primary income-related benefit. As such, it captures a large share of the fiscal effects associated with changes in employment and wages.

5.3 Conclusions

We assessed the VfM of the STEP Ukraine programme by comparing its streams of benefits with the associated costs. This analysis was conducted from two perspectives: (i) the Exchequer (i.e., public finances); and (ii) society as a whole. We used impact evaluation estimates to determine two key benefit channels generated by the programme: an increased probability of employment among STEP participants and higher wages for those already employed at baseline. These estimates were used to estimate both the number of participants moved into employment and the wage uplift for those already in work.

We utilised DWP analysis to monetise the impacts of transitions from unemployment into employment and uplifts in earnings. From the Exchequer's perspective, benefits include increased revenues from income tax, employer and employee National Insurance contributions, and indirect taxes, as well as reductions in UC payments. These are compared against the costs of implementing the programme. From the social perspective, the main benefit stream is increased economic output, which is weighed against programme costs and any additional travel costs associated with employment.

Our headline results indicate that, from the Exchequer's perspective, every £1 invested in the programme generates gross savings of £2.04 (a net saving of £1.04 per £1 spent). Sensitivity analysis suggests the BCR could range from 0.88 to 3.27. From the social perspective, every £1 spent delivers gross benefits of £3.67 (a net benefit of £2.67 per £1 invested), with sensitivity analysis indicating a possible BCR range of 1.59 to 5.84.

Overall, the economic evaluation provides strong evidence of net benefits. Positive net benefits are observed from both the Exchequer and social perspectives in the central estimate, and they persist for society even under the most pessimistic scenario.

6 Conclusion

6.1 The Process Evaluation

Throughout Chapter 2, the process evaluation explored how STEP Ukraine was implemented and experienced by participants and delivery partners. The findings of the process evaluation highlighted the dual-focus model of STEP Ukraine, which combines English language training with personalised employment support; in turn, this means that STEP Ukraine is well-placed to address both immediate integration needs and longer-term career aspirations. Furthermore, this model proved adaptive to needs and was feasible to deliver at scale while being mobilised quickly in response to emerging needs through online provision and adaptive scheduling. Ukrainian-speaking Employment Advisers and partnerships with organisations such as Business in the Community (BITC and The Entrepreneurial Refugee Network (TERN further strengthened delivery and supported culturally sensitive, tailored employment support. Finally, iterative adaptations to recruitment, onboarding and data systems, including strengthened eligibility screening and the move to a CRM, improved feasibility, reduced early dropout, and enhanced the quality of operational management over time.

However, the process evaluation also identified constraints, which are likely to limit STEP Ukraine from achieving the full impacts that it is capable of achieving. Employment Advisers who provided STEP Ukraine often had very high caseloads, which led to limited integration between employment and English language support, and also resulted in more variation in adviser practice than could be advisable under ideal provision of the programme. In addition, the intensity of the English language support and the regular scheduling of English language and employment support at the same time limited the effectiveness of these courses by creating barriers that STEP Ukraine participants had to overcome. This was particularly true for participants who had a lower proficiency in English at the start of the programme and those who had childcare responsibilities, which limited their ability to attend sessions.

Overall, the process findings suggest that the STEP Ukraine model is operating effectively and is valued by participants. However, learnings from the process evaluation indicate that future programmes would benefit from more integration between delivery tracks, consideration of all types of support being provided, and lower caseloads for employment advisers.

6.2 The Impact Evaluation

Language impact evaluation

In Chapter 3, the language impact evaluation examined the impact of STEP Ukraine on the English proficiency of participants using pre-/post-STEP Ukraine test scores. Unfortunately, a pre-/post-STEP Ukraine analysis does not provide causal estimates. Therefore, the language impact evaluation also conducted a counterfactual analysis, which estimated what would have happened in the absence of the programme using the relationship between entry test scores and time spent in the UK.

The language impact evaluation found that English test scores showed significant improvement between when participants entered the programme and when they exited the programme. Overall, there were improvements of between 20.8% and 29.2% in English Language scores in before-after comparisons of average scores. In addition, after estimating what would have happened in the absence of STEP Ukraine, we found that the English language skills of STEP Ukraine participants improved by between 17.6% and 25.9% when compared to the counterfactual. These effects were statistically significant and consistent across cohorts, providing strong evidence that the intensive, employability-focused English course resulted in significant improvements in English language skills, which are greater than what would have occurred in the absence of the programme.

The language evaluation also showed that English ability varies across different groups. In particular, younger participants, women, and those with higher levels of education tended to have stronger English skills when they entered the STEP Ukraine programme; there were also some regional variations in English language ability across the UK. These findings highlight the fact that some participant groups, namely older individuals, men, those with lower educational attainment, and people with childcare needs, are likely to begin support programmes like STEP Ukraine with a lower level of English language proficiency and may therefore face more significant barriers to entry.

Taken together, the language impact findings indicate that the English course as implemented is an effective model for improving proficiency within a relatively short time period. In addition, future programmes that aim to maximise improvements in the English language could benefit from targeted outreach and support towards groups with weaker language skills on programme entry.

Employment and wage impact evaluation

The employment impact evaluation aimed to estimate the impact of STEP Ukraine on participants' probability of being in employment, whether they were in continuous employment for a period of 3 or 6 consecutive months, and how much their employment paid in terms of monthly wages. The analytical approach compares the outcomes of STEP Ukraine participants to a counterfactual group of Ukrainian nationals who did not participate in STEP Ukraine, but who share similar observed characteristics.

The employment evaluation revealed that STEP Ukraine increased the probability of being in employment by an average of 6.6 percentage points over the 20 months following programme entry. In addition, we found that while it took a few months for the impact to take full effect, the sustained effect appears to stabilise with an 8-percentage point difference between the treatment and control group from 7 months after the treatment group began STEP Ukraine. Importantly, participants receive a 12-month action plan and access to an online portal for 6 months following completion of the programme. While by no means evident, it is plausible that participants are making use of these resources after completing the intervention,

which eventually translates to the full effect of the programme being realised over the course of several months.

Furthermore, there were significant and positive effects of STEP Ukraine on sustained employment, as defined by being in employment for more than 3 or 6 consecutive months. This suggests that STEP Ukraine not only helps participants find employment, but that this employment is sustained into the future, indicating long-lasting effects of the programme.

For participants who were already employed prior to the start of STEP Ukraine, the employment evaluation found that there was a positive effect on monthly earnings. Our evidence suggests that these monthly earnings of STEP participants increased by an average of around 11% over the 20 months after the intervention began. It is important to note that this includes the impacts of STEP Ukraine both in terms of increased hours of work and of increased hourly wages, both of which would result in higher wages; unfortunately, it is not possible for this evaluation to differentiate between the two.

Our findings are in line with some similar findings from impact evaluations of employability programmes aimed at supporting refugees, albeit in a different country context. For example, prior evidence indicates that job market coaching in Sweden can result in an increase of around 6 percentage points in the refugee employment rate.[160] However, other evaluations have found stronger impacts on refugee employment resulting from language support training, as Lochmann et al. (2019 indicate an increased likelihood of between 15-27 percentage points for labour force participation of French refugees.[161] Unfortunately, the impact evaluation of the previous programme, which was most similar to STEP Ukraine (i.e., the Refugee Employability Programme, is not yet available to draw comparisons to.

The employment evaluation also aimed to evaluate the impact of STEP Ukraine within the devolved administrations of the UK. Unfortunately, limited sample sizes and control group constraints prevented us from conducting a robust evaluation of the impact of STEP Ukraine in Wales or Northern Ireland, we were able to produce impact estimates for Scotland. We found similar impacts in Scotland to those across the whole UK, outlined above. These findings indicate that STEP Ukraine was effective at improving employment outcomes for participants in Scotland, but prevent us from drawing conclusions about the effectiveness of STEP Ukraine in Northern Ireland or Wales.

In sum, the employment impact findings indicate that STEP Ukraine led to statistically significant improvements in the likelihood of finding employment, of staying in employment for 3 or 6 months continuously overall for participants, and of improving monthly wages for those who were already employed when they began the programme.

---

[160] Joona, P; Nekby, L. (2012). Intensive coaching of new immigrants: An evaluation based on random program assignment. The Scandinavian Journal of Economics, Vol. 114, No. 2, Pp 575-600.

[161] Lochmann, A; Rapport, H; Speciale, B. (2019). The effect of language training on immigrants' economic integration: empirical evidence from France. European Economic Review, Vol. 113, Pp 265-296.

6.3 The Economic Evaluation

The economic evaluation compared the estimated costs and benefits of the programme from both the Exchequer perspective and that of society as a whole in order to assess the VfM of the STEP Ukraine Programme. Using DWP analysis, we estimated the benefits arising from the programme. This considered benefits driven by increased employment probabilities among STEP Ukraine participants and wage gains for those already in employment.

Our central scenario indicates that the programme represents VfM from both the Exchequer and social perspectives. We estimate that for every £1 invested, the programme generates £2.04 in gross savings to the Exchequer and £3.67 in gross benefits to society overall. Sensitivity analysis, undertaken to reflect uncertainty in key model parameters, suggests that these BCRs range from 0.88 to 3.27 from the Exchequer perspective and from 1.59 to 5.84 from the social perspective.

In summary, the results show that net benefits outweigh costs from both the Exchequer and social perspectives under our central scenario. Moreover, positive net benefits persist from the social perspective even under the most pessimistic scenario, indicating that the STEP Ukraine programme represents an effective use of public funds.

6.4 Conclusion and Recommendations

The evidence from across the evaluation indicates that STEP Ukraine delivered an effective model, which was able to address both the linguistic and employment-related barriers faced by Ukrainian guests.

To achieve this, the model brought together English language training and personalised employment support. The impact evaluation suggests that these two elements resulted in measurable improvements in English language skills, employment probability, employment sustainability, and, for those who were already employed when they entered STEP Ukraine, improvements in monthly pay.

The positive impacts estimated in the impact evaluation are complemented by the findings from the process evaluation. The process evaluation indicates that positive outcomes for participants were achieved primarily as a result of the dual-model approach, which promoted learning complementary English language and employment skills. The process evaluation also provides indicators about how stronger impacts could have been achieved, as there were clear constraints around operationalisation of STEP Ukraine, such as high caseloads and the varied needs and requirements of the large participant pool, indicating that the dual model approach was not integrated as effectively as it could have been. If these constraints and barriers to entry for participants had been mitigated sooner or considered further during the rapid deployment phase, then the impacts on employment and English language skills could have been greater in magnitude.

Taken together, the findings of the process evaluation suggest the programme is operationally viable, and that the provision of English language support alongside employment support provides the basis for a model that, based on the results of the impact and economic evaluation, improves the ability of participants to find and remain in employment. In particular, the positive effects on the English language identified in Chapter 3 are evidenced by findings from the process evaluation, which highlighted that participants felt more confident in their language skills and had a better ability to communicate in English after participating in the STEP Ukraine programme.

This evaluation suggests that not only does personalised employment support help participants apply language skills learned from English language support, but also that participants are able to utilise the skills learned across STEP Ukraine in practice. Findings from the process evaluation, which highlight the importance that English language support played in enabling participants to find employment, are evidenced by the findings from the impact evaluations that English language skills for participants improved, and that participants were more likely to find and remain in continuous employment following the programme.

While overall the evidence suggests that the STEP Ukraine model produces meaningful improvements in outcomes, there are a series of important considerations. Despite positive impacts, findings from the process evaluation indicate that true impacts were not fully realised due to limitations in programme implementation, including limited coordination in the delivery teams, and scheduling concerns caused by intensive, short-term courses.

A set of recommendations can be drawn from the findings across the process, impact, and economic evaluations:

- Differentiating between participants who join the programme with different levels of English language ability would enable future programmes similar to STEP Ukraine to provide more equitable support to all participants. Findings from the process evaluation demonstrate that the English language courses provided were sometimes too difficult for those starting the programme with a low level of baseline proficiency. Therefore, providing courses that can be more easily tailored by tutors to specific levels of prior English proficiency would likely result in stronger positive effects on the English language outcomes estimated in Chapter 3. Further flexibility in providing courses at different times of day would help participants who expressed difficulties in attending courses due to existing commitments, such as childcare or work. If this flexibility were provided, then participants may be more likely to attend courses and employment support sessions, which in turn, may result in larger improvements to outcomes.

- Stronger communication and connection between providers of English language support and employment support would enable better structuring and sequencing of the participant journey through STEP Ukraine, mitigating some of the concerns about timetabling clashes and difficulties. While this study did find positive impacts on employment and English language outcomes, the process evaluation revealed that English tutors and EAs struggled to communicate effectively, resulting in concurrent timetabling of courses/sessions, and a confusing journey through the programme for some participants. Improving communication between the two strands of support would lead to a more cohesive journey, ultimately resulting in stronger impacts on both employment and English outcomes for participants of future programmes.

- Refinement of outcome monitoring and data collection processes would enable future impact evaluations of similar programmes to estimate the impact of programmes on confidence, motivation, and other non-employment outcomes. Tutors and EAs had some concerns about being under-trained to provide mental health support in the process evaluation, but it was not possible to estimate the impact of STEP Ukraine on mental health outcomes during the quantitative evaluation, as this data was not routinely collected. It was not possible to estimate impacts for sub-groups (e.g., based on visa status), which could influence the extent to which participants experience impact. If this data was collected, future evaluations could conduct informative sub-group analysis to estimate whether there are different impacts between individuals.

- Enhanced post-programme support, i.e., in addition to the 12-month action plan that is part of the existing model, would ensure that skills learned during the programme are utilised effectively, but could also provide sector-specific guidance and advice after employment is found by programme participants. This support should aim to supplement the existing support and therefore should take the form of activities such as drop-in sessions during which participants can get one-to-one guidance on their updated situation, or alumni groups, which would enable peer-to-peer support and learning to enhance the existing impacts found in this study.

Appendix 1: Process evaluation

Table A1.1 Evaluation questions grid

Evaluation question (EQ)	Desk research	Process mapping	Interviews (Programme & Team Managers)	Online survey (EAs & English tutors)	Interviews (EAs)	Interviews (English tutors)	Ukrainian guests (participants)	Ukrainian guests (non-participants)	Ukrainian guests (focus groups)
EQ1: To what extent was the programme delivered as intended? • What changes were made and why were they made? • How did these changes affect outcomes?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
EQ2: Who did the programme reach? • Who did the programme not reach, and why?	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes
EQ3: What worked well, and less well? • What were the key facilitators and challenges to implementation? • What can be learned from STEP Ukraine to better support Ukrainian guests into the British labour market?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes

Full list of resources included in the desk research:

- British Council. ‘English Online. STEP Ukraine Research Proposal’.
- Business in the Community and World Jewish Relief. ‘STEP Ukraine: Behind the Scenes’ Brochure.
- MHCLG. ‘Cohort timeline Table - updated 17July2024'.
- MHCLG. ‘Theory of change (ToC)’
- MHCLG, World Jewish Relief, and the British Council. ‘Issues log’
- Pre- and post-programme beneficiary surveys, and12 months post-programme beneficiary surveys.
- ‘Monthly Reporting summary’. Sept 23.
- ‘Monthly Reporting’: multiple reports including qualitative data.
- ‘Monthly Reporting’: multiple reports including quantitative monitoring data.
- ‘Quarterly reports’: quantitative and narrative data.
- World Jewish Relief. ‘Data Flow Chart, v.4'.
- World Jewish Relief. ‘STEP Ukraine Process Map’.
- World Jewish Relief, British Council, and UK Government. 2024. ‘STEP Ukraine: Commission paper (slides)’. August 2024.
- World Jewish Relief, British Council, and UK Government. 2024. ‘STEP UKRAINE Phase 1 Final Narrative Report’. 20th September 2024.

Table A1.2 Adaptations log

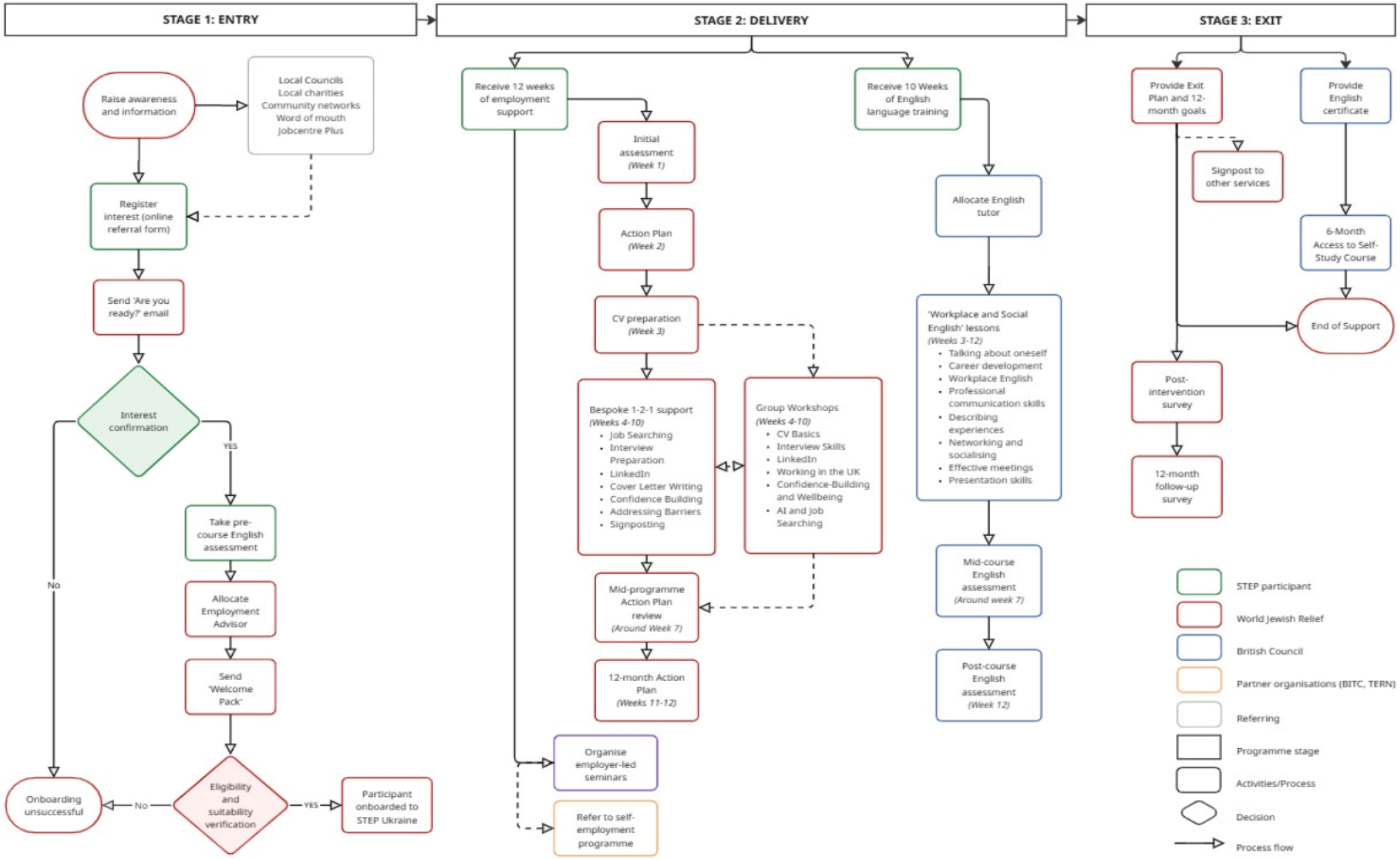
-	What was modified	Who made decision	When did it occur	Where was the adaptation made	Why was it made	Category of adaptation	Impact (expected and/or unexpected)	Source
1	Recruiting more EAs	WJR + MHCLG	30/01/2024	Delivery stage: While high numbers registered for the programme the uptake was ca. 40%	To improve fidelity: to improve adherence, dose, or quality of intervention use	Systematic and reactive	Increase the number of EAs with a proven track record.	Issues log, PM3
2	Introducing group work curriculum to accommodate the necessary increase in caseload for EAs	WJR + MHCLG	30/01/2024	Delivery stage: While high numbers registered for the programme the uptake was ca. 40%	To improve fidelity: to improve adherence, dose, or quality of intervention use	Systematic and reactive	EA response appears to have been generally positive, viewing this change as an effort by WJR to decrease their workload. However, some EAs (e.g., EA7) felt uncomfortable delivering the workshops, preferring to work one-on-one.	Issues log, TM2, EA6, EA7, EA8
3	Introduction of the "Are you ready" email	WJR + MHCLG	Early in the programme	Entry stage: While high numbers registered for the programme the uptake was ca. 40%.	To improve adherence (client engagement) and reduce dropouts.	Systematic and reactive	Increased client engagement and reduced dropouts (this change helped filter out unengaged clients early in the process)	PM1
4	Timing and process of the British Council English test.	WJR+BC	After initial implementation issues	Entry stage: While high numbers registered for the programme the uptake was ca. 40% [moved from delivery stage]	To improve adherence (client engagement), streamline the process and reduce pressure on staff.	Systematic and reactive	More efficient onboarding and better preparation for classes (this change allowed for better planning and reduced last-minute pressures)	PM1, PM2, TM2

-	What was modified	Who made decision	When did it occur	Where was the adaptation made	Why was it made	Category of adaptation	Impact (expected and/or unexpected)	Source
5	Centralisation of communication and hiring of a marketing manager	WJR	After initial implementation issues	Entry stage: While high numbers registered for the programme the uptake was ca. 40% [also throughout the programme: to increase attendance and survey uptake]	To effectively disseminate information (improve communication strategy, use social media and Telegram channels) and improve message delivery (e.g., reduce emails going to spam).	Systematic and reactive	Improved communication efficiency (consistency and clarity) and client response rates.	PM1
6	Timing of client selection for English classes	WJR+BC	After initial implementation issues	Entry stage: While high numbers registered for the programme the uptake was ca. 40%	To reduce last-minute changes and ensure class availability.	Systematic and reactive	Smoother class allocation (improved planning and resource allocation for classes) and improved client satisfaction.	PM1, PM2
7	Stricter eligibility and suitability checks	WJR	Later phases of the programme	Entry stage: While high numbers registered for the programme the uptake was ca. 40%	To ensure participants genuinely meet the programme requirements (including the suitability criteria, such as having sufficient free time to enable genuine commitment to the programme).	Systematic and reactive	Reduced dropout rates.	PM3, EA8
8	Improved data management system (CRM) and team management dashboards	WJR	After initial implementation issues (EA3 implies the CRM system was introduced in autumn 2023)	Delivery stage	To improve data management and access compared to using spreadsheets	Systematic and reactive	Easier and more efficient tracking the percentage of people turning up for appointments	PM3, EA1, EA2, EA3

-	What was modified	Who made decision	When did it occur	Where was the adaptation made	Why was it made	Category of adaptation	Impact (expected and/or unexpected)	Source
9	Introduction of team structures and communication channels for staff.	WJR	After initial implementation issues	Delivery stage	To improve internal communication and contribute to team building	Systematic and reactive	EAs were placed in teams with a Team Manager (TM), increasing the quality of support and guidance	PM1, TM1, TM2, EA1, EA3
10	Refinement of allocation process (allocation of clients to EAs) based on language proficiency and location	WJR	Early in the programme	Entry stage	To improve adherence and streamline internal processes	Systematic and reactive	Decrease in the number of errors and the time required to allocate participants. Perceptions of negative impacts for EAs who only allocated participants with very low English levels were only allocated participants with very low English levels were only allocated participants with very low English levels were only allocated participants with very low English levels were	PM1, EA4, EA8
11	Introduction of eligibility calls.	WJR	Early in the programme	Entry stage: While recruitment was effective, modifications were made to help reduce dropouts	To improve adherence and streamline internal processes	Systematic and reactive	Helped filter out unengaged or unsuitable participants early in the onboarding process, thereby reducing dropouts	PM1, PM3, TM1, TM2
12	Introduction of information webinars for potential participants	WJR	Later phases of the programme	Entry stage: While recruitment was effective, WJR introduced modifications to reduce dropouts	To re-engage those who had applied but not yet started due to waiting lists.	Systematic and reactive	High turnout, revealed misunderstandings about programme details (e.g., intensity)	PM1

-	What was modified	Who made decision	When did it occur	Where was the adaptation made	Why was it made	Category of adaptation	Impact (expected and/or unexpected)	Source
13	English assessment conducted during the last class	BC	From cohort G	Delivery to exit stage	To increase completion rates	Systematic and reactive	Improved completion rates	PM2, ET7
14	Streamline operational processes between BC and WJR, with WJR programme assistants assuming responsibility for handling timetabling issues with the British Council	WJR + BC	After initial implementation issues	All stages	To facilitate communication and integration between delivery partners	Systematic and reactive	Fostered communication and coordination between delivery partners, smother overall programme delivery	PM2, PM3
15	Direct provision of laptops	WJR	Later in the programme (extension cohorts?)	Around initial screening [late entry/early delivery stage]	To overcome the barrier of lack of access to technology, and the slow and inconsistent support provided by local councils' hardship fund.	Systematic and reactive	Direct and improved programme technology support route	TM1
16	Development of TL/TMs for specific areas of expertise (higher skills, well-being)	WJR	Early in the programme	Delivery stage	To provide more holistic and specialist support in specific areas relevant to employment	Systematic and proactive	Enhanced quality and breadth of support available to clients by having designated experts in key areas	TM1
17	WJR programme assistants assuming responsibility for initial contact with client and eligibility checks	WJR	Early 2024 (from cohort E)	Entry stage	Not explicitly stated, but presumably to reduce EAs' workload	Systematic and reactive	Improved efficiency and client communication, and decreased EAs workload	EA5, EA7

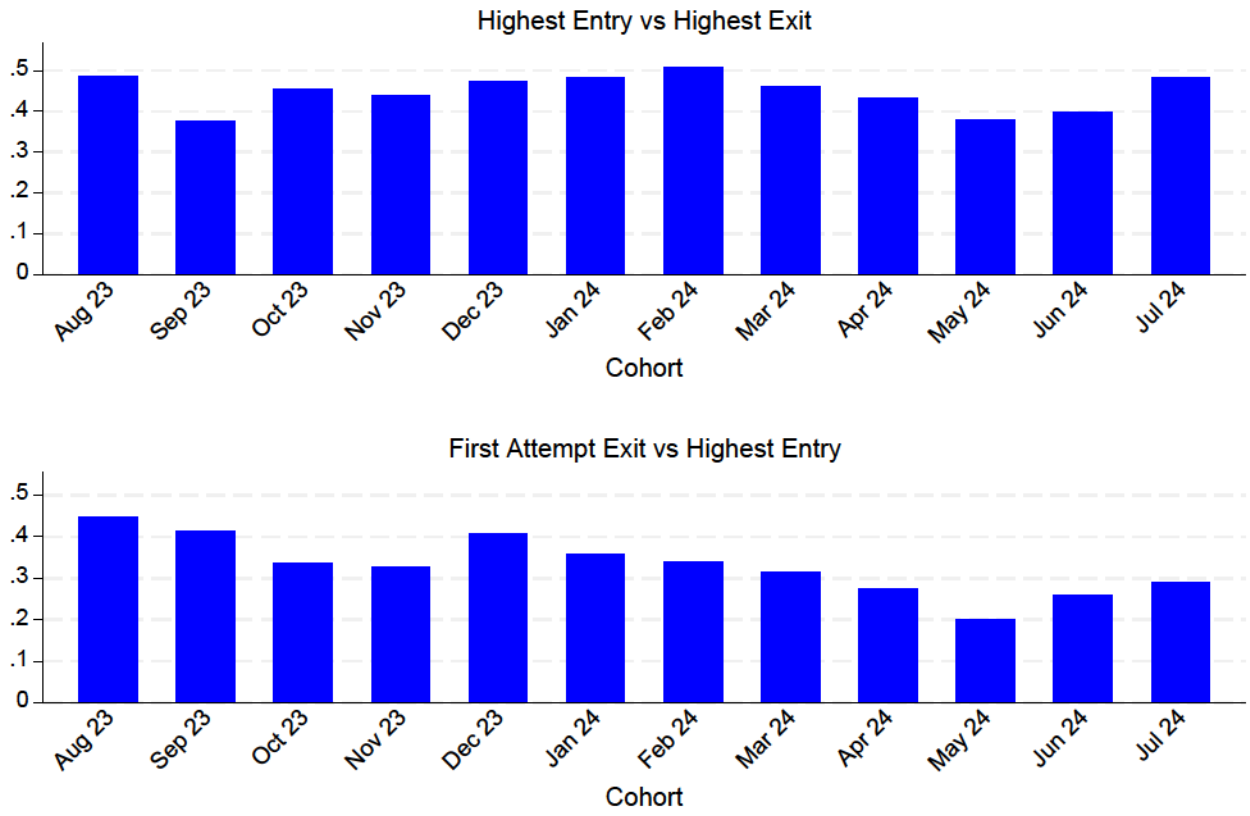
Figure A1.1 Flowchart of STEP Ukraine's process map



Appendix 2: Impact on English skills

Figure A2.1 shows the change in STEP participants' average English test scores from programme entry to exit in standard deviations, while distinguishing between cohorts.

Figure A2.1 Standard deviations change in average entry vs exit test score by cohort



Tables A2.1 and A2.2 below present the results of t-tests to determine whether the observed pre-post English test score differences are statistically different from zero. We find that the average score difference for each cohort is statistically significant at the highest conventional level (1%). This holds regardless of whether the exit score used is the first attempt or the highest of multiple attempts.

Table A2.1 t-tests on average test score difference (highest exit vs highest entry) by cohort

Cohort	Aug 23	Sep 23	Oct 23	Nov 23	Dec 23	Jan 24	Feb 24	Mar 24	Apr 24	May 24	Jun 24	Jul 24	All
Average entry test score	294.0	353.6	318.2	316.3	295.1	284.4	274.0	281.3	273.3	300.3	271.1	267.1	289.0
Average exit test score	352.8	398.9	373.2	369.3	352.5	342.8	335.6	337.0	325.8	346.3	319.3	325.5	343.4
Average test score difference	58.7	45.4	55.0	53.0	57.4	58.3	61.7	55.7	52.5	46.0	48.1	58.4	54.4
t-statistic	-6.97	-5.88	-24.6	-19.3	-29.3	-19.9	-28.4	-27.8	-16.6	-20.5	-22.0	-18.7	-72.5
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table A2.2 t-tests on average test score difference (first attempt exit vs highest entry) by cohort

Cohort	Aug 23	Sep 23	Oct 23	Nov 23	Dec 23	Jan 24	Feb 24	Mar 24	Apr 24	May 24	Jun 24	Jul 24	All
Average entry test score	293.1	356.3	319.8	321.8	294.1	287.1	275.8	288.7	275.8	305.2	275.2	270.6	291.9
Average exit test score	347.7	406.5	360.7	362.0	343.7	330.6	317.1	326.9	309.1	329.5	306.3	305.8	330.0
Average test score difference	54.6	50.2	40.9	40.3	49.6	43.5	41.3	38.2	33.3	24.3	31.1	35.2	38.0
t-statistic	-5.82	-5.62	-14.3	-12.9	-19.7	-13.4	-17.9	-16.9	-8.75	-10.2	-11.3	-9.64	-43.3
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table A2.3 presents the output from regressing participants' highest entry test average score on the number of months they have spent in the UK at the time of enrolling in STEP. The following controls are included: (i) STEP cohort, (ii) age, (iii) gender (whether they are male), (iv) their region/country of residence, (v) whether they are unemployed, and (vi) whether they have childcare needs. Dummy variables for gender and region/country are relative to Female and North (England), respectively, which are therefore omitted.

Table A2.3 Output from regressing entry test average score (highest) on months in UK

-	(1)	(2)	(3)
VARIABLES	**Entry test average score**	**Entry test average score**	**Entry test average score**
Months since arrival in UK	2.747***	2.519***	2.512***
	(0.193)	(0.194)	(0.204)
Cohort	N/A	-3.449***	-3.397***
		(0.446)	(0.438)
Age	N/A	N/A	-2.677***
			(0.125)
Male	N/A	N/A	-11.46***
			(3.231)
Region = 2, Midlands (England)	N/A	N/A	17.26***
			(4.513)
Region = 3, South and East (England)	N/A	N/A	13.23***
			(3.290)
Region = 4, Northern Ireland	N/A	N/A	-21.11**
			(8.595)
Region = 5, Scotland	N/A	N/A	1.094
			(4.196)
Region = 6, Wales	N/A	N/A	6.606
			(6.150)
Unemployed	N/A	N/A	-3.499
			(2.570)
Childcare needs	N/A	N/A	-14.65***
			(2.723)
Constant	202.0***	234.5***	344.2***
	(6.107)	(7.359)	(9.628)
Observations	9,484	9,484	9,444
R-squared	0.021	0.027	0.081

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Tables A2.4 and A2.5 present the results of t-tests to show that the pre-post English test score differences after applying a counterfactual adjustment of 7.5 points over the duration of the programme remain statistically different from zero at the 1% significance level.

Table A2.4 t-tests on average test score difference (adjusted highest exit vs highest entry) by cohort

Cohort	Aug 23	Sep 23	Oct 23	Nov 23	Dec 23	Jan 24	Feb 24	Mar 24	Apr 24	May 24	Jun 24	Jul 24	All
Average entry test score	294.0	353.6	318.2	316.3	295.1	284.4	274.0	281.3	273.3	300.3	271.1	267.1	289.0
Average exit test score	345.2	391.4	365.7	361.8	345.0	335.2	328.1	329.5	318.3	338.8	311.8	318.0	335.9
Average test score difference	51.2	37.8	47.5	45.5	49.9	50.8	54.2	48.2	45.0	38.5	40.6	50.9	46.9
t-statistic	-6.07	-4.91	-21.2	-16.6	-25.4	-17.3	-24.9	-24.0	-14.2	-17.1	-18.5	-16.3	-62.4
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table A2.5 t-tests on average test score difference (adjusted first attempt exit vs highest entry) by cohort

Cohort	Aug 23	Sep 23	Oct 23	Nov 23	Dec 23	Jan 24	Feb 24	Mar 24	Apr 24	May 24	Jun 24	Jul 24	All
Average entry test score	293.1	356.3	319.8	321.8	294.1	287.1	275.8	288.7	275.8	305.2	275.2	270.6	291.9
Average exit test score	340.2	399.0	353.1	354.5	336.1	323.1	309.6	319.4	301.6	322.0	298.8	298.3	322.4
Average test score difference	47.1	42.6	33.3	32.8	42.1	36.0	33.8	30.7	25.8	16.7	23.6	27.7	30.5
t-statistic	-5.02	-4.78	-11.7	-10.5	-16.7	-11.1	-14.6	-13.6	-6.77	-7.01	-8.53	-7.58	-34.7
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table A2.6 presents the output from regressing ONS survey respondents' self-assessed English level (the lowest among their self-assessed speaking, reading and writing levels) on the number of months they have spent in the UK at the time of taking the survey.

The following controls were included: (i) whether the respondent is 30 or older, (ii) gender (whether the they are male), (iii) whether they have children aged 0-17, (iv) the level of their highest qualification, (v) their current living arrangements, (vi) whether they have been experiencing physical or mental ill health lasting 12 months or more, and (vi) their region/country of residence.

Dummy variables for gender, highest qualification, current living arrangements and region/country are relative to being female, degree level or above, living with a sponsor, and the North (England), respectively. As such, these categories are omitted from the analysis. Also note that the sample used for this analysis was restricted to non-participants of STEP Ukraine in the ONS survey data who were found to exhibit comparable

characteristics on the basis of the included controls to STEP participants in the ONS data. This was determined using propensity score matching.

Table A2.6 Output from regressing self-assessed English level (minimum) on months in UK

-	(1)	(2)
VARIABLES	**Minimum level of English (speaking reading or writing)**	**Minimum level of English (speaking reading or writing)**
Months since arrival in UK	0.0203***	0.0126**
	(0.00566)	(0.00554)
Aged 30 or older	N/A	-0.442***
		(0.0460)
Male	N/A	-0.132***
		(0.0376)
Has children aged 0-17	N/A	-0.0632*
		(0.0375)
Highest qualification = 2, Vocational or job-specific qualification	N/A	-0.774***
		(0.0526)
Highest qualification = 3, Below degree level (e.g., school qualification)	N/A	-0.355***
		(0.124)
Highest qualification = 4, No qualification	N/A	-0.936***
		(0.114)
Current living arrangements = 2, Living with family/friends	N/A	-0.0660
		(0.0958)
Current living arrangements = 3, Renting	N/A	0.143***
		(0.0383)
Current living arrangements = 4, Temporary or homeless	N/A	-0.00573
		(0.0843)
Physical or mental ill health lasting 12 months or more	N/A	-0.0861*
		(0.0485)
Region = 2, Midlands (England)	N/A	-0.125*
		(0.0636)
Region = 3, South and East (England)	N/A	0.0327
		(0.0475)
Region = 4, Northern Ireland	N/A	-0.0800
		(0.126)
Region = 5, Scotland	N/A	-0.0647
		(0.0579)
Region = 6, Wales	N/A	-0.212**
		(0.0873)
Constant	3.095***	3.815***
	(0.171)	(0.174)
Observations	3,534	3,534
R-squared	0.004	0.128

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix 3: Phase 1 Employment Impact Evaluation

This section presents our Phase 1 findings of the programme's impact on the following employment outcomes: (i the likelihood (probability of being employed; (ii monthly take-home pay among those in work; and (iii self-assessed job-skill match, referring to the degree to which the participant believes their employment matches their skills. These outcomes were measured using the pre-, immediately post- and 12 months post-STEP beneficiary surveys administered by WJR.

After matching between the pre- and 12 months post-STEP datasets, we were left with a sample of 45 individuals who were identified across the two datasets, representing only 0.6% of the eligible population. We also present results from comparing the pre- and post-STEP datasets, for which we drew on a matched sample of 375 individuals, representing 4.8% of the eligible population.

Given the small and potentially self-selected nature of these samples, the following results should be viewed with caution. A further caveat is that while we did apply a counterfactual adjustment to our take-home pay results, participants may also have experienced some benefit to their employment probability and job-skill match in the absence of the programme, which is not accounted for in these estimates.

Employment probability

We found the average employment rate across cohorts rose by about 5 percentage points from before to immediately after the programme. This difference was found to be statistically significant at the 5 percent level. It is worth noting, however, that the absolute change was modest and masks substantial variation by cohort (differences range from -6 to +17 percentage points.

Looking instead at the average change from pre-STEP to 12 months post-STEP, we found a more considerable increase of 8 percentage points in employment probability, from a pre-STEP employment probability of 68% to a 12 months post-STEP probability of 76%. However, this difference was not statistically significant at conventional levels. This was likely due to the limited sample size of the matched dataset (45 individuals, which may have been inadequate to distinguish the relatively modest change of 8 percentage points from a scenario of no change.

Monthly take-home pay

The difference in average monthly pay immediately post-STEP relative to pre-STEP was small and not statistically significant across cohorts. This likely reflects the insufficient time elapsed since participating in the programme for participants to effectively transition to higher-paid employment.

The 12 months post-STEP survey results pointed to a relatively substantial increase in monthly pay compared to pre-STEP levels of £236, which was statistically significant at the 1% level. Note, however, that this amount does not account for wage growth that

participants may have experienced anyways. To adjust for this, we used the average nominal wage growth over the 12 months between the initial and follow-up surveys. During this period, UK average weekly regular pay rose from £623 to £660 according to ONS figures, an increase of 5.9%. Applying the 5.9% figure to the 12-month results finds that the net gain attributable to the programme was £181.11 per month, a 19.3% uplift accounting for counterfactual wage growth.

Back-calculating from net figures (i.e. after tax to gross (i.e. before tax, we found that participants of STEP received on average £213.80 extra gross pay each month above counterfactual wage growth, or £2,566 more gross income per year 12 months after starting the programme, corresponding to a 22.8% increase over gross baseline pay.

Job-skill match

In addition to employment status and wages, the beneficiary surveys asked employed participants how well their job matches their existing skills on a scale of 1 to 5. Overall, the average rating was 2.35 before STEP, which increased to 2.78 12 months after STEP. While this suggests an improvement in perceived job-skill match of around 8.6 percentage points, the increase was not statistically significant. However, mirroring our results for employment probability, we did find a statistically significant, but smaller, increase in self-assessed job-skill match in the period immediately after the programme. Once again, the loss of statistical significance in the 12-month results may have been due to the small sample size of the matched 12 months post-STEP dataset.

Propensity score matching using ONS data

We also conducted propensity score matching (PSM using survey data of Ukrainian visa holders in the UK collected by the ONS in April 2024. PSM is a statistical matching technique that estimates the probability ("propensity score" of an individual participating in the programme based on their characteristics and then pairs participants with non-participants who have similar propensity scores. Comparing the average outcomes of these matched groups then provides an impact estimate that accounts for observable differences between the two groups that can be attributed to the programme. We applied two different matching methods (nearest neighbour and kernel to obtain estimates of the impact on the likelihood of being employed, monthly take-home pay, and the extent to which the respondent agrees that their job matches their skills. However, these results were unsuccessful in eliminating the effect of selection into the programme of Ukrainians with worse employment outcomes and were therefore unable to yield viable impact estimates.

Appendix 4: Employment Impact Evaluation

Employment Impact Evaluation Methodology

Overview of staggered difference-in-differences approach

The evaluation of STEP Ukraine's impact on participant's employment outcomes employed a quasi-experimental design (QED, comparing outcomes of STEP Ukraine participants to Ukrainian guests in the UK who did not participate in the programme. As with English language level, Ukrainian guests' employment outcomes are likely to change with more time spent in the UK, as language skills, network and experience all increase with time. By constructing a plausible control group of people whose outcomes would be expected to evolve similarly to participants in the absence of STEP Ukraine, we can attempt to control not only for the impact of time in the UK but also for any other factors influencing trends in the labour market outcomes of Ukrainian migrants. We do this by comparing trends in the outcomes of participants and the control group around the time that participants are treated. This style of approach is known as a Difference-in-Differences (DiD analysis.

A staggered design is applicable because the programme was sequentially implemented for different cohorts at different points in time. Hence, even within the participant group there were moments at which some had been treated, and some had not. More generally, different cohorts differed in the number of months before or since they started STEP. In recent years the literature has uncovered, and found solutions for, pitfalls in traditional two-way fixed effects DiD estimation in settings with staggered treatment timing.[162] Our approach follows the Callaway and Sant'Anna framework (CSDID.[163] We define:

- **Treatment groups:** The cohort to which each treated individual belongs, corresponding to the month of their programme start date.
- **Time periods:** Calendar months, normalised so that t = 0 corresponds to each group's treatment date.
- **Control groups:** These are defined separately each month, so that they comprise not only those who were never exposed to STEP Ukraine but also participants who were yet to receive the intervention.

For each cohort and each time period, the estimator computes the average treatment effect on the treated (ATT) by comparing changes in outcomes for that cohort to corresponding changes for that month's control group at the same time. The ATTs were aggregated into a single average treatment effect across all cohorts (weighted by cohort size), all post-treatment periods, and all pre-treatment periods, which acts as a check of the parallel trends assumption (see below).

---

[162] Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of econometrics*, 225(2), 254-277. https://doi.org/10.1016/j.jeconom.2021.03.014

[163] Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2), 200-230. https://doi.org/10.1016/j.jeconom.2020.12.001

We combine the Callaway and Sant'Anna DiD regression framework described above with a weighting procedure. This combination is known as "doubly robust" DiD, proposed by Sant'Anna and Zhou.[164] The weights are "inverse probability weights" designed to make the treatment and counterfactual groups (post-weighting similar with respect to observed characteristics (more precisely, to make them similar in terms of the likelihood one would have assigned to them participating in STEP, given those characteristics. The weights are produced in a prior step to the DiD regression itself, based on a propensity score regression. Weights were estimated and applied for each month in the panel separately.

Only limited, non-identifying information was available from HMRC records to the evaluation team due to the data sharing constraints in place. Therefore, the variables used as controls and therefore included within that first stage throughout the evaluation include:

- Age (continuous);
- Sex (a binary representation);
- Date of arrival in the UK (operationalised as the number of days in the UK since arrival).

Parallel trends assumption

The validity of DiD rests on the parallel trend assumption: in the absence of treatment, the outcomes of participants and non-participants would have evolved similarly. While this assumption is never directly testable (as we never observe treated units in the untreated state post-treatment), we assessed its plausibility by examining pre-treatment trends among the comparison group compared to treatment cohorts.

In addition to visual plots, we assess parallel trends using statistical tests. These take two forms. The first is a test of whether parallel trends between treatment and control groups holds exactly in every individual period prior to treatment (we conduct this test using all 20 pre-treatment months and separately using the 12 months immediately prior to treatment). This is the benchmark that one would expect to hit with perfectly random assignment of a treatment (e.g., with an experimental control group). The second is a test of whether parallel trends between treatment and control groups holds on average when we take all pre-treatment periods together.

These tests are complementary and, together, provide a useful assessment of the robustness of the results. The first test is much stricter than the second. As an example of why this is the case, this test rejects parallel trends even if it holds in 19 months out of 20 pre-treatment periods, if in 1 month there is sufficient evidence of a divergence in trends. The second test looks at whether there is a systematic divergence in trends between treatment and controls prior to treatment on average. For example, if parallel trends hold in all but two pre-treatment months but is violated in one direction in one month and violated by roughly the same magnitude in the other direction in another, parallel trends would likely be rejected by the first test but not by the second. The ideal scenario is that both tests fail to reject parallel trends, and this is what would be expected with a purely

[164] Sant'Anna, P. H., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of econometrics*, 219(1), 101-122. https://doi.org/10.1016/j.jeconom.2020.06.003

randomised control group. However, particularly when any deviation from parallel trends is modest in magnitude relative to estimated treatment effects, it is reasonable to have a good level of confidence in the results if the second parallel trends test is not rejected - even if the strictest test is rejected. As we shall show, this is generally the scenario that we are in. In addition, it is important to note that since many tests of differences in pre-treatment periods are undertaken, we might expect a small number of results to be significant by chance (i.e., at the 5% level, we might expect 1 in 20 results to be significant by chance. Despite this, the analysis presented in this report provides more statistically significant results than we might expect by chance.

Outcomes and covariates

We examined three primary employment outcomes, each constructed from monthly PAYE earnings data:

- **Employment probability:** A binary indicator equal to 1 if the individual had positive earnings in a given month, and 0 otherwise.
- **Employment sustainability:** Binary indicators for sustained employment over specified durations (e.g., employed for at least 3, 6 or 12 consecutive months). These outcomes capture whether STEP Ukraine helps participants not just find employment, but maintain a stable, continuous employment relationship.
- **Monthly wages (conditional on employment):** The log of gross monthly inflation-adjusted earnings, conditional on employment, for those with positive earnings at programme entry. We use the natural logarithms so that estimates can be interpreted approximately as percentage effects on wages.

The covariates used for the inverse probability weights are demographic characteristics that: (i) predict programme participation, and (ii) are plausibly related to employment outcomes. These include age (continuous), sex (binary), and days since arrival in the UK (continuous). Geographic location is another potential covariate. The limitation is that as HMRC PAYE records only exist for those individuals who receive earnings, and as this is the only source of information on the location of non-participants, we did not have geographic region/country data for those in the control group who were never employed during our panel. To maintain the comparability between the treatment and control groups, geographic region/country was dropped as a covariate in the main analysis so both groups could encompass individuals who were employed at one point during the panel and those who were never employed during the panel. However, models including region/country as a covariate, and consequently restricting the control group by excluding those who were never employed during the panel, were used for a sensitivity analysis and this made little difference to the results.

Impacts in Scotland

Table A4.1 Post-treatment average effects on monthly wages in Scotland (real pay in logs)

Average effect on employment probability	Coefficient	Standard error
Post-treatment average effect (0 to 20 months)	0.039	0.050

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Table A4.2 Pre-trend analysis for monthly wages in Scotland (real pay in logs)

Employment probability pre-treatment trend test: Chi^2	P-value
20 months before treatment	0.000
12 months before treatment	0.000

Employment probability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	-0.061	0.060
12 months before treatment	-0.020***	0.006

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Impacts in Wales

Table A4.3 Post-treatment average effects on employment probability in Wales

Average effect on employment probability	Coefficient	Standard error
Post-treatment average effect (0 to 20 months)	-0.150**	0.513

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Table A4.4 Pre-trend analysis for employment probability in Wales

Employment probability pre-treatment trend test: Chi^2	P-value
20 months before treatment	0.000
12 months before treatment	0.000

Employment probability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	-0.015**	0.005
12 months before treatment	-0.016***	0.005

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Table A4.5 Post-treatment average effects on employment sustainability in Wales

Average effect on employment sustainability	Coefficient	Standard error
3-month employment sustainability (0 to 20 months)	-0.148***	0.052
6-month employment sustainability (0 to 20 months)	-0.160**	0.051
12-month employment sustainability (0 to 20 months)	-0.140**	0.047

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Table A4.6 Pre-treatment trend analysis for employment sustainability in Wales

3-month sustainability pre-treatment trend test: Chi^2	P-value
20 months before treatment	0.000
12 months before treatment	0.000

3-month sustainability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	-0.015***	0.004
12 months before treatment	-0.015***	0.004

6-month sustainability pre-treatment trend test: Chi^2	P-value
20 months before treatment	0.000
12 months before treatment	0.000

6-month sustainability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	-0.009**	0.003
12 months before treatment	-0.012**	0.004

12-month sustainability pre-treatment trend test: Chi^2	P-value
20 months before treatment	0.000
12 months before treatment	0.000

12-month sustainability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	-0.003	0.001
12 months before treatment	0.000	0.002

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Table A4.7 Post-treatment average effects on monthly wages in Wales (real pay in logs)

Average effect on employment probability	Coefficient	Standard error
Post-treatment average effect (0 to 20 months)	0.306*	0.147

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Table A4.8 Pre-trend analysis for monthly wages in Wales (real pay in logs)

Employment probability pre-treatment trend test: Chi^2	P-value
20 months before treatment	0.000
12 months before treatment	0.000

Employment probability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	-0.068	0.051
12 months before treatment	-0.029	0.018

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Impacts in Northern Ireland

Table A4.9 Post-treatment average effects on employment probability in Northern Ireland

Average effect on employment probability	Coefficient	Standard error
Post-treatment average effect (0 to 20 months)	-0.017	0.047

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Table A4.10 Pre-trend analysis for employment probability in Northern Ireland

Employment probability pre-treatment trend test: Chi^2	P-value
20 months before treatment	0.000
12 months before treatment	0.000

Employment probability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	-0.004	0.007
12 months before treatment	-0.017	0.005

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Table A4.11 Post-treatment average effects on employment sustainability in Northern Ireland

Average effect on employment sustainability	Coefficient	Standard error
3-month employment sustainability (0 to 20 months)	-0.020	0.048
6-month employment sustainability (0 to 20 months)	-0.030	0.048
12-month employment sustainability (0 to 20 months)	-0.075	0.042

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Table A4.12 Pre-treatment trend analysis for employment sustainability in Northern Ireland

3-month sustainability pre-treatment trend test: Chi^2	P-value
20 months before treatment	0.000
12 months before treatment	0.000

3-month sustainability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	-0.014*	0.006
12 months before treatment	-0.022***	0.006

6-month sustainability pre-treatment trend test: Chi^2	P-value
20 months before treatment	0.000
12 months before treatment	0.000

6-month sustainability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	-0.013*	0.005
12 months before treatment	-0.025***	0.006

12-month sustainability pre-treatment trend test: Chi^2	P-value
20 months before treatment	0.000
12 months before treatment	0.000

12-month sustainability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	-0.002	0.005
12 months before treatment	-0.075	0.042

Note: Statistical significance is denoted as follows: p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

Table A4.13 Post-treatment average effects on monthly wages in Northern Ireland (real pay in logs)

Average effect on employment probability	Coefficient	Standard error
Post-treatment average effect (0 to 20 months)	-0.139	0.073

Note: Statistical significance is denoted as follows: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Table A4.14 Pre-trend analysis for monthly wages in Northern Ireland (real pay in logs)

Employment probability pre-treatment trend test: Chi^2	P-value
20 months before treatment	0.000
12 months before treatment	0.000

Employment probability pre-treatment trend test: Average	Coefficient	Standard error
20 months before treatment	0.000	0.142
12 months before treatment	-0.001	0.014

Note: Statistical significance is denoted as follows: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Appendix 5: Economic evaluation

The following section outlines the economic evaluation in more detail. Section 1 explains how the per-person benefits were calculated from increased employment. Section 2 describes the calculations to derive the benefits from higher earnings. Section 3 explains how the per participant benefits were aggregated. Section 4 outlines the sensitivity analyses, Section 5 summaries the calculations of programme costs, and Section 6 briefly describes how the final figures for the benefit-cost ratios (BCRs were found.

1. Benefits of increased employment probability

Estimates of increased employment probability

The key parameter for understanding the benefits due to increased employment among participants was the change in employment probability associated with the programme. The estimate for the increase in employment probability comes the impact analysis which found that the probability of STEP participants being employed increased by 6.6 percentage points as a result of the programme.

Subgroup formation

The benefits per participant of moving from unemployment to employment were calculated using DWP analysis. This allowed us to estimate the net benefits of moving unemployed individuals into employment through considering the benefits and costs from multiple perspectives.

We created separate subgroups to reflect the fact that the impact on tax revenues and Universal Credit payments from entering employment depends on the level of earnings. Using a single earnings rate could result in all or none of the participants crossing key Universal Credit or tax thresholds and distort results. To address this, we divided those entering employment into three equally sized subgroups based on their post-treatment earnings. For each subgroup, we estimated average earnings and calculated the corresponding benefits per participant.

The DWP analysis also requires participant characteristics as inputs. These include factors such as age, number of children, and rental costs, which influence the level of Universal Credit received by unemployed individuals and, consequently, the fiscal benefits of moving them into employment. Using data on programme participants, we were able to determine the characteristics of a typical STEP Ukraine participant.

Given the limited availability of data, assumptions were required for many inputs. In most cases where data were lacking, the guidance provided by DWP was followed to make the relevant assumptions.

It is important to note that from a social perspective, benefits and costs are unaffected by these characteristics. Both benefits (increased output and reduced operational costs and costs (programme expenses and increased travel impacts are independent of individual participant characteristics. Therefore, claimant characteristics influence BCRs only from the perspective of the Exchequer and do not affect the BCR from a social perspective.

Our limited available data allowed for the estimation of some typical characteristics of STEP participants. Data on the number of children was limited as the only variable was on whether the participant had children, rather than the exact number. It was found that most participants had children, and it was therefore assumed that all participants with children had only one child and that the child was under the age of 14.

The available data allowed us to determine the average age of STEP participants. The only relevant input for the DWP analysis is whether participants are over 25 years of age. Since the average participant was older than 25, we assumed that all participants were over 25.

For the purposes of the analysis, it was assumed that participants did not have health problems limiting their ability to work, nor were they responsible for caring for disabled individuals or children. This assumption was based on data indicating that only a small number of participants fell into these categories. As an additional check, the DWP analysis was rerun assuming that 30% of STEP participants had limited capability for work and work-related activity (LCWRA, which could alter Universal Credit impacts and costs due to increased travel. The results showed that including this subgroup had only a negligible effect on the results.

Benefit unit capital, non-wage income and partner income were all assumed to be zero. This follows DWP recommendations for when data is lacking on these variables.

Data was not available on the type of claimant in terms of whether they were single, in a couple or a lone parent. Given the nature of the participants as predominately refugees, it was assumed that all participants were lone parents.

Rental costs were calculated using DWP rental guidance which gave the average rental costs for social and private depending on claimant type, number of children and region/country of the UK. This data was combined with data on the region/country of the UK in which STEP participants lived in order to find a weighted average cost of rental for participants by subgroup. This value-weighted average, once adjusted to September 2025 prices, came out to be £585 per month.

DWP analysis outputs

After incorporating the above inputs into the DWP analysis, the following benefits were derived as outputs from the different perspectives:

The Exchequer:

- Benefits included increases in income tax, indirect taxes, employee and employer National Insurance contributions and reductions in Universal Credit and operational costs.

Society:

- The benefits included the increase in output and the reduction in operational costs.
- The costs included the social impacts from increased travel.

Social benefits from increases in output are assumed to be equivalent to the combined rise in wages and employer National Insurance contributions, consistent with standard economic theory that a worker's marginal product equals the marginal cost of employing them. However, it should be recognised that estimates of increased output as a social benefit are likely to be subject to greater uncertainty than fiscal benefits to the Exchequer, such as higher tax revenues and reduced Universal Credit expenditure.

Through the above outputs, average net benefit (before considering programme costs was found per participant for each of the subgroups. This average was then weighted by the proportion of participants in each subgroup.

The proportion of unemployed participants that do not claim universal credit was also factored into the weighted average. The ONS data suggested that 81% of unemployed STEP participants received benefits, with the remaining 19% not claiming any additional support. The benefits from the perspective of the Exchequer of those not claiming universal credit differ from those who were claiming universal credit. This is because the Exchequer does not benefit from reduced benefit transfers when the participants entered employment. The net benefits to the Exchequer were therefore lower for individuals who were not claiming Universal Credit. To estimate benefits for participants not claiming Universal Credit, we adapted the DWP analysis by removing components applicable only to Universal Credit claimants. This adjusted model was then used to calculate the average net benefit per participant for each subgroup among those not claiming Universal Credit.

2. Benefits of increased earnings

Impact estimate

The estimated increase in gross monthly pay was taken from the impact analysis. The results showed that STEP participants that were already in employment prior to the programme saw an average treatment effect of a rise of 0.1066 in log real wages (roughly equivalent to an increase of 11.25%.

Benefits from increased earnings

Three subgroups were again created, this time based on baseline earnings for participants who were already employed at the start of the programme and subsequently experienced higher earnings. The participants were divided into three tertiles, and the average baseline pay for each group was calculated. A 0.1066 increase in log wages was then applied to each tertile to determine the earnings uplift for the three subgroups.

DWP analysis was again used to estimate the Exchequer and social benefits from the wage uplift experienced by STEP participants. Calculations were again performed separately for UC recipients and non-recipients. The same categories of benefits and costs are captured as the analysis of employment probability increases. Specifically, from the Exchequer perspective, these include increased tax revenues and reduced UC payments. From the social perspective, they include increased economic output and additional transportation costs.

3. Aggregation of benefits

Below, we outline how the per participant benefits were aggregated to find the total benefits from the STEP Ukraine programme as a whole. This covers important factors that were considered such as persistence, substitution effects, and discounting.

Average per participant annual benefits were summed up by the relevant number of STEP participants to get aggregate figures for each perspective. Per participant benefits due to increased employment was aggregated by the estimated number of participants pushed into employment by the programme. This was found by multiplying the total number of STEP participants by the percentage point increase in employment probability. Yearly benefits from higher wages were also aggregated through the number of STEP participants who were already employed at baseline. Since not all STEP participants who were employed at baseline remained employed after the programme, this figure was adjusted by the proportion of post-treatment periods in which baseline-employed participants retained employment, which was 86%.

Substitution effects

As STEP participants secure new employment and increased wages, there is the possibility that there is some level of displacement of existing workers. This will reduce the potential benefits from the perspective of the Exchequer and society. Guidance from DWP suggests undertaking a sensitivity analysis using a substitution effect of 20% for supply side programmes and hence the Exchequer and society yearly benefits were reduced by this factor under the low scenario sensitivity analysis.

Discounting

Given that the majority of flows of both the benefits and costs of the STEP programme occurred prior to this evaluation, these flows were not discounted, as in line with HM Treasury's Green Book guidance. The final payment for the programme has, however, yet to occur and is scheduled to be paid in May 2026. We therefore applied a real discount rate of 3.5% to derive the present value of this cost flow, using September 2025 as the base period. This is in line with HM Treasury's Green Book recommendation when estimating the future costs and benefits of a programme and represents social time preferences for such flows.

Price adjustment

All prices in the model, including earnings and programme costs, were adjusted to September 2025 values to allow assessment of real, rather than nominal, flows of costs and benefits. Prices prior to September 2025 were adjusted using the Office for National Statistics (ONS CPIH indices. For prices beyond September 2025, adjustments were made using either ONS price indices or future CPI forecasts from the Office for Budget Responsibility (OBR.

Persistence

Persistence refers to the length of time for which the benefits of the programme continue for. This is a key parameter in the model and has a significant impact on the final results.

Unfortunately, STEP participants were only observed up to 20 months post the programme's end and hence we were unable to derive exact estimates of the persistence of the employment and wage increases of the programme.

Whilst some literature on the persistence of employment programme effects suggests long lasting benefits (e.g., Card et al., 2017, employment effects from STEP Ukraine may not persist as long due to non-participants catching up over time in language acquisition and its associated implications for labour market integration. That is, benefits relative to non-participants are likely to fall over time as non-participants develop language skills themselves through immersion in the UK and other means of language acquisition.

For our central scenario analysis, we set the persistence period to the duration over which participants were observed in the programme (20 months. Our impact estimates indicate a sustained effect on both employment probability and wage growth for at least 20 months following the start of treatment. We selected 20 months as the central estimate because we lack data beyond this period and therefore cannot make reliable assumptions about the impact's duration after the observation window. It is likely that the effects persist beyond 20 months, as it is improbable that they would suddenly disappear immediately after the observation period. Consequently, this scenario can be regarded as relatively conservative.

As there remains uncertainty in the length of persistence, sensitivity analysis was also done on this parameter with a lower estimate of 14 months and an upper estimate of 26 months employed in additional modelling. This is also broadly in line with the persistence ranges used in Alma Economics' previous economic evaluation of the Fair Start Scotland programme, which used the DWP model to calculate the benefits and costs of the Fair Start employment support programme.

4. Sensitivity analysis

Given the uncertainty in certain parameter estimates, specifically our impact estimates, the value of persistence, and the substitution effect, we conducted a sensitivity analysis of these components within the VfM model. This approach allowed us to assess how variations in key parameters influence the final results. In addition to our central scenario, we analysed two alternative sensitivity scenarios (low and high by adjusting the relevant parameters. For each scenario, BCRs were calculated from both the Exchequer and social perspectives.

The changes in the parameters in the sensitivity checks are outlined in Table A5.1.

Table A5.1 Sensitivity analyses conducted of VfM model

-	Persistence	Employment impact estimate	Log earnings uplift impact estimate	Substitution effect
Central scenario	20 months	0.0661pp	0.1066	0%
Low scenario	14 months	0.0541pp	0.0772	20%
High scenario	26 months	0.0781pp	0.1361	0%

5. Programme costs

Data on the costs of the programme were provided by MHCLG, including the spread of these costs over time. These costs included upfront payments, test performance payments, KPI payments and contract variation payments, with the payments beginning in 2023 and the final payment due in 2026.

Using ONS past CPI inflation indices and OBR future CPI inflation forecasts, the flows of costs were put into real September 2025 prices. Once real costs prices had been obtained, future flows were discounted at a rate of 3.5% to obtain present values for the costings.

Nominal costs of the programme were £8,178,288. After discounting and adjusting to real September 2025 prices, the total cost of the programme amounts to £8,601,869.

6. Comparison of costs and benefits

Once total costs and benefits of the programme had been calculated from the Exchequer and society, the BCR could be found by simply dividing the total benefits by the total costs. This gave the figure for the gross benefits per £1 spent on the programme for each group.