The Employability Skills Assessment Tool (ESAT)

Phase 3 Final Report

September 2025

Blueprint

Table of contents

Acknowledgements

About the Future Skills Centre

The Future Skills Centre is a forward-thinking centre for research and collaboration dedicated to driving innovation in skills development so that everyone in Canada can be prepared for the future of work. We partner with policymakers, researchers, practitioners, employers and labour, and post-secondary institutions to solve pressing labour market challenges and ensure that everyone can benefit from relevant lifelong learning opportunities. We are founded by a consortium whose members are Toronto Metropolitan University, Blueprint, and The Conference Board of Canada, and are funded by the Government of Canada's Future Skills Program.

About Blueprint

Blueprint is a nonprofit that helps leaders use data and evidence to tackle complex public policy challenges across Canada.

We partner with government, community, philanthropic, and industry leaders to strengthen public systems and deliver better outcomes. Our team brings together policy analysts, evaluators, economists, data scientists, and implementation experts—people who know how to turn insight into action. Our work is grounded in deep subject-matter expertise, rigorous methods, and a real-world understanding of how systems operate and evolve. More than just an advisor, we're also partners in change. We provide key support at every stage of the policy and program lifecycle: from early strategy and design to implementation, evaluation, and continuous improvement.

As a consortium partner of the Future Skills Centre, Blueprint works with partners and stakeholders to collaboratively generate and use evidence to help solve pressing future skills challenges.

About Futureworx

Futureworx is a social purpose organization that develops and delivers innovative employment tools and training and provides outstanding services that support people on their work and personal journeys. Futureworx Society is a charitable organization based in Nova Scotia, whose purpose, since its inception in 1984, has been to help individuals facing employment barriers overcome their obstacles and reach their full potential, both in their community and in the workplace. Today, Futureworx is a leader in developing and delivering innovative employment tools and training, always with a wholistic view that incorporates social and emotional skills or 'soft' skills development. As a high-growth and innovative organization, Futureworx delivers services in every county of Nova Scotia and every Canadian province and works with partners across Canada and around the globe.

The ESAT Phase 3 *Final Report* is funded by the Government of Canada's Future Skills Program.

The opinions and interpretations in this publication are those of the author(s) and do not necessarily reflect those of the Government of Canada.

About this report

This report shares findings from Blueprint's evaluation of the **Employability Skills Assessment Tool (ESAT)**. ESAT is an online training tool developed by **Futureworx** that helps individuals assess and strengthen their social and emotional skills (SES), including their self-awareness and motivation to do so, leading to better outcomes in work, education, and life. It also offers employers and service delivery practitioners (SDPs) insights into their clients' SES, readiness for work, and training needs.

Blueprint's Phase 2 *Final Report* (April 2025) provided results from an evaluation of ESAT's effectiveness, adaptability, reliability, and validity conducted between September 2021 and May 2024. This Phase 3 *Final Report* builds on those findings by establishing an ESAT theory of change and discussing the steps needed to evaluate ESAT's impact rigorously through a randomized controlled trial (RCT).

The report is organized into five sections:

1. **Introduction (p. 5)** provides a brief background on the importance of SES in the labour market, ESAT's history, and its scaling journey as part of the FSC portfolio.
2. **About ESAT (pp. 6–8)** details the ESAT tool, including its purpose, structure, core skills, and participant journey.
3. **Theory of change (pp. 9–11)** describes how ESAT is designed to help people strengthen the SES they need to succeed in work and life. Our theory of change shows how this happens—step by step—starting with training for staff and ending with better job outcomes for participants.
4. **RCT design options (pp. 12–21)** identifies options for designing and implementing an ESAT RCT. We consider strategies for implementation, randomization, measurement, and sample size.
5. **Recommendations (p. 22)** summarizes our recommendations for an RCT.

1. Introduction

1.1. Background and context

Social and emotional skills (SES) are increasingly critical to employability—and workforce development programs are investing in tools that can assess and build these capabilities. Futureworx's Employability Skills Assessment Tool (ESAT) is one such innovation: a digital platform designed to help individuals and service providers understand, assess, and foster SES growth. Since its launch in 2014, ESAT has been implemented in a range of contexts across Canada, including public employment services, essential skills training, and college programs.

In 2020, Futureworx partnered with the Future Skills Centre and Blueprint to conduct a preliminary evaluation of ESAT. That report offered insights for improving and expanding ESAT's use. In 2021, Futureworx received additional funding from FSC for a second phase: to work with Blueprint to generate evidence on ESAT's effectiveness, adaptability, validity, and scaling potential as part of the Scaling Up Skills Development Portfolio. Our Phase 2 *Final Report* (April 2025) explored how ESAT is used in different contexts and how it can fill the gap in adult SES assessment in Canada, using data collected from September 2021 to May 2024.

1.2. Measuring ESAT's effectiveness

Phase 1 and 2 provided insights into ESAT's implementation, including participant satisfaction, staff perceptions, and employment outcomes. Building on these findings, Futureworx received funding to work with Blueprint in 2024 to conduct a feasibility study for a randomized controlled trial (RCT) to rigorously assess ESAT's causal impact. This Phase 3 *Final Report* offers two supports in that evidence-building journey:

a) **Theory of change.** The ESAT theory of change lays out how people using the tool—both staff and participants—move through a series of actions and experiences that lead to meaningful growth. It connects what happens in programs (like training and feedback conversations) to the changes we expect to see, like better communication, greater confidence, and improved job outcomes.

b) **RCT options.** We then outline how ESAT can be tested rigorously despite the complex nature of its implementation. We consider design options for an RCT, assess the feasibility of these approaches, and offer recommendations and a roadmap to prepare for large-scale testing.

2. About ESAT

2.1. What is ESAT?

The Employability Skills Assessment Tool (ESAT) is a web-based platform designed to help individuals assess, understand, and develop their SES. It is used primarily by publicly funded workforce development programs to help jobseekers and learners develop their employability skills. Unlike traditional assessments that provide a final score only, ESAT is an interactive, developmentally focused tool that fosters self-awareness. It provides structured feedback and progress tracking through self-assessments, staff evaluations, and coaching discussions.

ESAT is often used in the context of either a job-readiness program, an essential skills program, or a technical training program. These programs are typically delivered as part of a community-based employment program or in a college setting. Programs using ESAT vary in length from three to four weeks to up to 12 months or more.

2.2. ESAT core features

ESAT provides a structured developmental approach with five key features:

1. **Clear skill definitions.** ESAT focuses on nine core SES, each with defined behavioural indicators (see **Table 1** on the next page).
2. **Multi-perspective assessments.** Participants complete self-assessments while staff (job coaches, trainers, or employers) observe and provide their own evaluations.
3. **Structured feedback system.** ESAT generates visual reports to help participants and staff discuss skill strengths and areas for improvement (see **Figure 1** on page 8 for examples).
4. **Flexible application.** Organizations can tailor ESAT to different training programs, selecting which SES to assess.
5. **Progress tracking.** ESAT allows for multiple assessments over time, enabling participants to track their growth.

> "ESAT is designed to help people understand, develop, and apply social and emotional skills in real-world work and life contexts."

Table 1 | ESAT's nine social and emotional skills

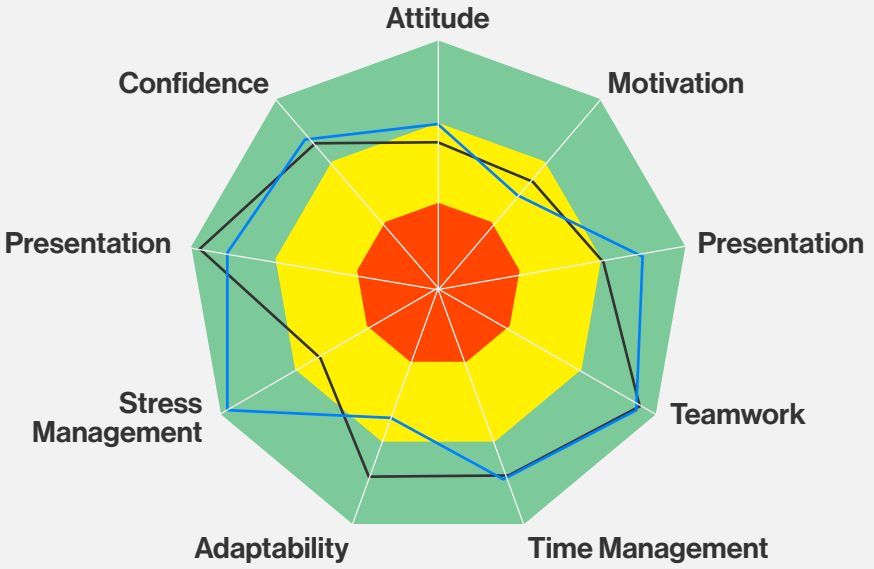
Skill	Description
Accountability	Willingness to admit mistakes, accept responsibility for and learn from them, accept feedback constructively, monitor quality of work when unsupervised, and display an honest and ethical approach to work and others in the workplace.
Adaptability	Ability to react constructively to both anticipated and unanticipated changes in the workplace, take responsibility for learning needed to adapt to change, and adjust interactions with others based on previous experiences or the formality of the situation.
Attitude	Ways to show or express feelings about a person, work activity, event, or idea, shown verbally or behaviourally. The focus is not on internally held attitudes, but rather on how they are expressed; thus 'attitude' can be thought of as a tone imparted to an interaction.
Confidence	Belief in one's competence and the ability to express one's perspective to others. It includes being able to function in uncertain situations, being appropriately assertive, and taking reasonable chances.
Motivation	Desire to set and achieve high standards on the job. Fundamentally, motivation speaks to desire to be an excellent employee; it involves showing a willingness to expend the effort needed to excel. This can include striving for specific work standards as well as high personal standards associated with setting goals, showing initiative, and doing one's best with and without supervision.
Presentation	How a person appears to others in terms of their dress and adornments, hygiene, etiquette, manners, and language.
Stress Management	Ability to experience workplace stress without it impacting performance or coworkers. It also includes one's ability to keep personal stress out of the workplace. It requires that an employee seek help for stress in a timely manner.
Teamwork	Ability to set and follow priorities, follow schedules for arrival, breaks, etc., and stay on task in an efficient manner, which includes meeting deadlines and communicating efficiently with others.
Time Management	Ability to work cooperatively with others, handle conflict and anger appropriately, communicate effectively, be empathetic, and respect differences (e.g., cultural, religious, gender, etc.).

2.3. How ESAT is used

The ESAT process has six broad stages:

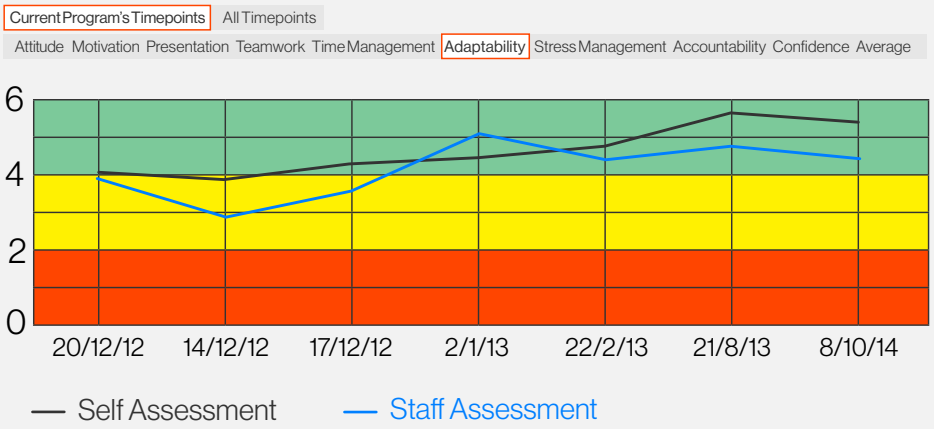
1. **Program staff introduce ESAT** and **explain how it works.**
2. **Participants complete an online self-assessment** that uses a common language for SES; participants build their self-awareness of SES strengths and areas for improvement.
3. **Staff conduct structured observations** and assess participants based on behavioural indicators.
4. **Staff meet to review scores** and reach a consensus, ensuring consistent evaluations.
5. **Participants and staff discuss results with participants** using ESAT-generated reports (e.g., radar charts) to visualize strengths and opportunities for growth.
6. **Repeat assessments (optional)**. Programs using ESAT over many weeks can conduct re-assessments to track skill development over time.

Figure 1 | Sample radar plot produced by ESAT for participant debriefs



Note. Staff and participant ratings are plotted against three colours. Red (1–1.9) indicates that behaviour is not at an appropriate level for the workplace, yellow (2–3.9) that skills are progressing, and green (4–6) that skill levels are appropriate for the workplace.

Sample "distance travelled" plot to illustrate participant progress



Participant view of ESAT question mapped onto colour-coded scoring system

Question I am open to learning new ways of doing things.					
Correspnding Skill Adaptability					
Learner Response					
Never	Rarely	Sometimes	Usually	Almost always	Always
Corresponding Scores and Radar Map Colours					
1 to 1.9	2 to 2.9	3 to 3.9	4 to 4.9	5 to 5.9	6

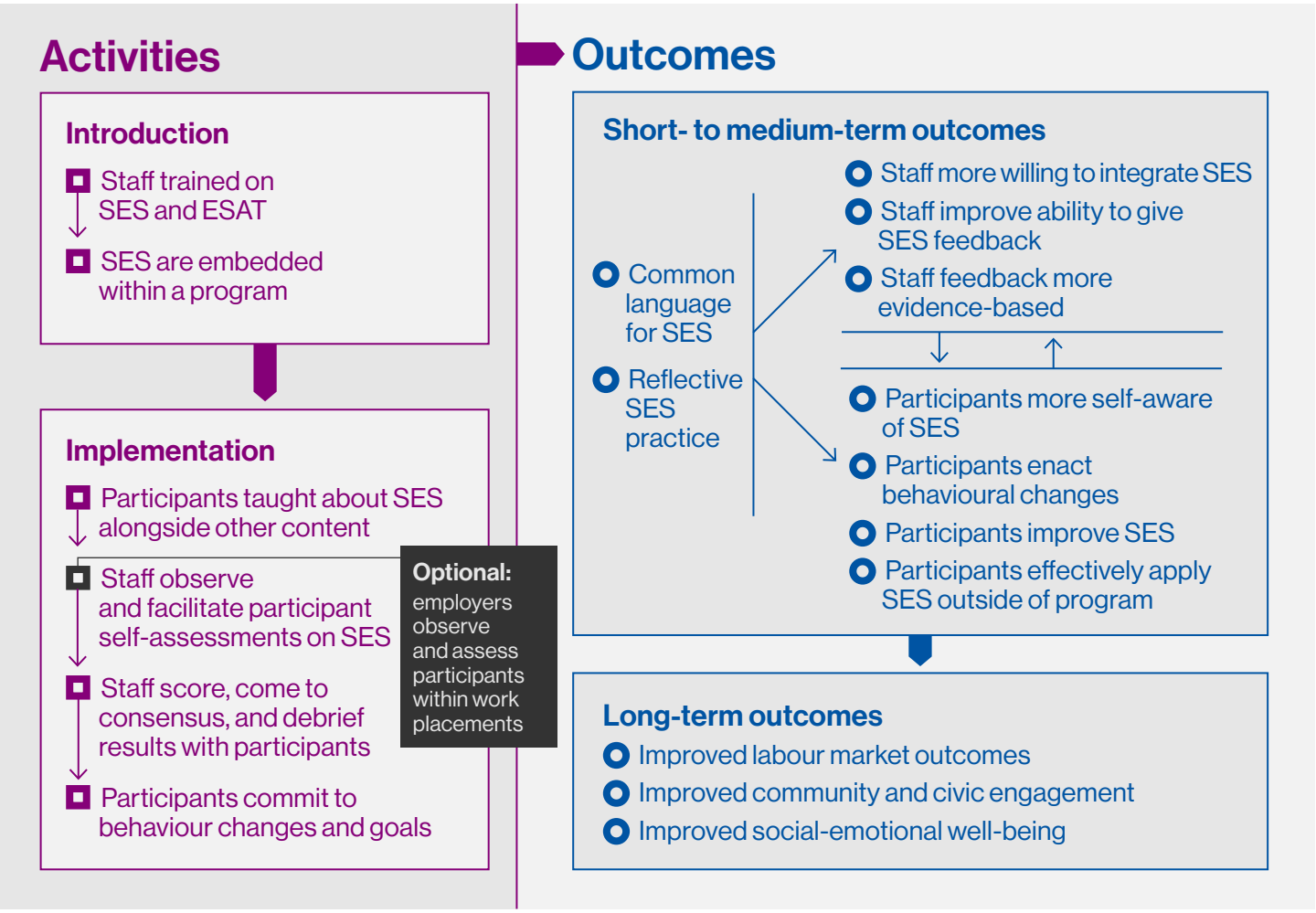
3. Theory of change

To build our theory of change (ToC), Blueprint and Futureworx collaborated across three project phases. We listened to staff, reviewed the results of our evaluations, and discussed how people actually experience and use ESAT. The result is a clear picture of how ESAT can help people grow their skills—and how we can test if it is working as intended.

3.1. ToC diagram

Figure 2 offers a visual representation of the key activities needed to implement ESAT and the outcomes we predict based on those activities.

Figure 2 | ESAT ToC



Note. Boxes represent activities required to implement ESAT, separated into preparation and implementation phases. Circles represent outcomes, categorized into short-term and long-term. Large arrows represent the expected sequence from preparation to implementation to outcomes. Thinner arrows represent the sequence of activities within each category and predicted causal relationships between groups of outcomes.

3.2. Theory of change narrative

3.2.1. Preparation

ESAT is delivered by **community-based employment organizations** and **colleges**. In either setting, ESAT is delivered in a similar way. First, ESAT service delivery partner staff receive training in social and emotional skills (SES). This helps them understand what SES are, how to support participants in developing them, and how to use ESAT in program activities.

Once trained, staff weave SES content into their programming type—either in **employment-readiness programs, essential skills programs**, or **technical training programs**—and find ways to practice and reflect on these skills with participants.

3.2.2. Implementation

Next, participants begin learning about SES in a hands-on way (described above in **section 2. About ESAT**). They explore how these skills manifest in their daily lives and training. Using ESAT, they reflect on where they are strong and where they can grow. With support from staff, they set goals, explore new behaviours, and track their progress over time.

3.2.3. Short-term outcomes

In the short term, we expect ESAT will help **both participants and staff**:

- speak a shared language about SES so it is easier to give and receive helpful feedback; and
- spend more time reflecting on how to build and apply SES in real life.

For **staff**, this means they:

- feel more confident teaching and supporting SES development; and
- give clearer, more meaningful feedback using common terms and participant input.

For **participants**, this means they:

- better understand their own SES—what they are good at and where they can improve;
- try out new behaviours that help them grow; and
- strengthen their SES and start using those skills beyond the program, including in job settings.

3.2.4. Long-term outcomes

As **participants** grow their SES over the **longer term**, we expect them to have:

- better experiences at work, including finding and keeping jobs and earning more;
- a stronger sense of wellbeing and belonging; and
- greater involvement in their communities.

3.3. Testing the ToC in practice

To test the ToC rigorously, we need to identify the degree to which ESAT directly causes each of these outcomes and whether the predicted relationships between outcomes occur. For example, does ESAT help staff and participants speak a shared language about SES? Do improvements in SES confidence result in higher rates of employment? Does developing SES lead to better jobs and better lives, and can ESAT help people get there faster?

Conducting such a test requires an experimental design in which we compare outcomes of two groups: a) people who participate in programs *using* ESAT; and b) people with similar characteristics, participating in similar programs, but who do not use ESAT. This allows us to assess the added value that ESAT brings to employment and training programs. The following section outlines a design for a **randomized controlled trial (RCT)**.

"A shared language and more time for reflection are the foundation of ESAT's short-term impact—for both participants and staff."

4. RCT design options

An RCT can measure ESAT's impact—separate from other elements of employment and training programs—by comparing ESAT users (the **program group**) with similar people who participate in programs that do not use ESAT (the **control group**). In other words, the RCT should compare two programs: one with ESAT embedded in it, and one without. Below, we recommend the most feasible approach to an RCT—the conditions needed to conduct this type of research effectively.

4.1. RCT components

An ESAT RCT needs to consider five elements. For each, we provide an analysis of options and recommendations for a preferred approach.

- **Tool development strategy.** Can ESAT be optimized before beginning an RCT?
- **Implementation strategy.** What implementation guidelines should be put in place for ESAT? What program contexts would best fit an RCT?
- **Randomization strategy.** How should participants be randomized into program and control groups?
- **Measurement strategy.** What measures should be considered to estimate impact? What data sources should be used?
- **Sample size.** How many participants should be recruited to detect ESAT's effects reliably?

4.1.1. Tool development strategy

Recommendation 1. Refine ESAT's reliability and validity before pursuing an RCT.

ESAT's theory of change suggests that more accurate measures should lead to greater effectiveness and improve our ability to measure effects in an RCT context. Therefore, it is important to ensure ESAT is as accurate a measure of SES as possible prior to exploring its impact in an RCT.

As part of our Phase 2 analysis, Blueprint shared a range of findings on ESAT's reliability and validity. Our analysis indicated that ESAT had a strong foundation for validation, but that there were opportunities to refine the tool further.

By adjusting ESAT content and re-measuring reliability and validity in an iterative study, Futureworx can improve the degree to which the tool validly assesses the skills it aims to measure. This will improve its effectiveness in generating participant skills and employment outcomes. Refining reliability and validity should happen before the RCT.

4.1.2. Implementation strategy

Recommendation 2. Service delivery partners should deliver ESAT according to enhanced implementation guidelines.

Using ESAT as Futureworx intended will enhance participant experience and increase the probability of positive outcomes. This will create the best possible conditions for a fair test of ESAT's impact.

In our Phase 2 *Final Report*, we found that fidelity to Futureworx's implementation guidelines mattered for participant experiences. Participants were more likely to be satisfied with ESAT in cohorts with greater fidelity to the guidelines. High-fidelity programs had rates of satisfaction **19 percentage points** higher than those with low fidelity. Every **one-point** increase in fidelity above **3/6** correlated with a **12%** higher likelihood of satisfaction.

We also found that most guidelines were realistic to implement. On average, programs followed **4.75/6**, and **75–100%** adopted **5/6**. All participating organizations met at least **3/6** criteria. However, **15/16 (94%)** were unable to meet all criteria due to a range of logistical and practical constraints. Due to time constraints in shorter programs, the most challenging guideline to achieve was having three separate engagement points for staff and participants.

In response, Futureworx adapted criteria to ensure they were easier to implement for staff while still ensuring the most satisfying experience for participants. Futureworx adapted **two** of the **six** criteria and added new, **seventh** criterion. **Table 2**, on the following page, provides a description of the original criteria, organizational fidelity to them, and modifications made by Futureworx in response to our report.

Table 2 | Implementation guidelines, fidelity, and modifications

Recommended implementation guidelines	**Fidelity** (staff end-of-cohort survey)	Modification from original
1. Staff explain the purpose and functionality of ESAT to all participants.	**100%** (16)	No modifications
2. A program staff member hosts a debrief session at the end of each assessment cycle for participants to review and interpret their ESAT scores.	**88%** (14)	No modifications
3. Observer assessments begin at least two weeks after the program begins to ensure a reasonable period to observe participant behaviour.	**75%** (12)	Observer assessments begin after a **baseline period** (three days to two weeks). *Flexibility added to the baseline period based on program length.*
4. A minimum of two people (program delivery staff, case counsellors/managers, or employers) provide feedback and document observations within the assessment.	**88%** (14)	No modifications
5. Observers meet and compare notes and scores at the end of each cycle to determine consensus scores for each skill and determine a debriefing strategy for participant strengths and opportunities for improvement.	**94%** (15)	No modifications
6. Observers and participants use ESAT a minimum of three times, meaning there is at least three assessment timepoints.	**31%** (5)	There is **more than one round of assessments and debriefs** during a program, where possible. Three rounds are encouraged. *Relaxed requirement from a minimum of three to two rounds of assessments and debriefs.*
	N/A	7. Participants are observed across a variety of program contexts **and activities**. *Changed to include program activities, not just delivery formats. Organizations can meet this criterion via a range of activities if programming does not allow for a range of formats or contexts.*

Recommendation 3. An RCT can be implemented in any program type or setting, but once one is chosen, the RCT should be carried out within that single setting and type only.

ESAT is a flexible tool, and our Phase 2 *Final Report* showed high satisfaction and effectiveness in both community-based employment organizations and colleges, as well as in essential skills and technical training programs. Differences in satisfaction across settings were small and likely due to individual partner variation, not the setting or program type itself.

This indicates that there is no obvious, preferable category of program type or delivery context to pursue for an RCT. However, to increase the likelihood of detecting ESAT's impact, the RCT should take place in just one setting and program type once chosen. This limits variability caused by unrelated factors and makes it easier to attribute outcomes to ESAT.

4.1.3. Randomization strategy

Recommendation 4. An RCT should implement an instructor-level randomization approach, in which participation in ESAT is randomized at the level of individual instructors. If this is not possible, a cohort-level randomization approach is a second-best option, in which each program cohort is randomly assigned to implement ESAT.

To measure ESAT's impact, the only difference between program and comparison groups should be whether participants use ESAT. Randomizing participants into these groups helps prevent pre-existing differences (such as education levels or work experience) from affecting results and ensures participants have an equal chance of being assigned to either.

ESAT is often delivered as an instructor-led program delivered in cohort-based or group settings. Given this delivery approach, we need to choose a randomization approach that makes sense without disrupting how programs function.

Table 3, on the following page, describes the three ways we could approach randomization and their pros and cons.

Table 3 | Randomization approaches

Randomization type	Description	Pros	Cons
Individual randomization	Each participant is randomly assigned to either use ESAT or not within the same 'class' or 'cohort.'	This method is easiest to implement statistically and requires the fewest participants overall.	It is difficult to keep ESAT separate within the same classroom. Instructors would need to treat participants differently in the same session. Additionally, instructor improvements (like better feedback) might spill over to all participants, including those in the control group. Individual randomization works best if ESAT is a short-term add-on rather than something that changes how the entire program runs.
Cohort-level randomization	Entire groups of participants (i.e., 'cohorts') are randomly assigned to the program or control group. In other words, some cohorts receive training with ESAT and others without.	Instructors can integrate ESAT into the program for one group only. This keeps the learning experience consistent within each cohort.	If the same instructor teaches both cohorts, their improved teaching methods (from using ESAT) could benefit both groups. This method requires more cohorts to ensure differences between groups are not due to luck. Cohort-level randomization is logistically more difficult than individual randomization.
Instructor-level randomization	Each instructor at a participating organization (community-based employment service or college) is randomly assigned to use ESAT in their teaching or not. All their learners either receive ESAT training or do not.	This avoids spillover between groups—each instructor uses one approach consistently. This is the best way to measure the full impact of ESAT, including staff-level changes.	Instructor-level randomization is the most difficult to organize: it needs many instructors and a large number of participants to make it work.

We recommend instructor-level randomization. It provides the cleanest test of ESAT's effects by separating the entire teaching approach. This means we can better measure how ESAT changes both staff behaviour and participant outcomes. If instructor-level randomization is not possible, **cohort-level randomization** is a solid second-best approach. Randomizing ESAT by cohort still allows instructors to integrate ESAT fully without needing to juggle two versions of the program at once.

We do not recommend **individual-level randomization** because it would be too difficult to keep the program and control groups separate in practice.

Recommendation 5. When using instructor- or cohort-level randomization, collect participant information (such as age, gender, and employment status) before randomization. These data can be used to match cohorts or instructors with similar participant profiles into pairs. For each pair, one is randomly assigned to the ESAT group and one to the control group.

When using cohort- or instructor-level randomization, we can reduce the risk of having very different types of participants in each group by using a matching process.

Before we randomly assign who receives ESAT and who does not, we collect basic sociodemographic information about participants, like age, gender, and employment history. Then we group similar cohorts or instructors together into pairs based on these characteristics. For example, if two cohorts are starting at the same time and both have a similar average age and employment background, we can treat them as a pair. Then we randomly assign one cohort to use ESAT (the program group) and the other to the comparison group.

Matching helps ensure both groups are as similar as possible before we test ESAT, which makes the results of our study fairer and more reliable.

4.1.4. Measurement strategy

Recommendation 6. An RCT should adopt a measurement strategy that relies on survey data for short-term indicators and linked administrative data for long-term indicators, such as labour market outcomes.

A measurement strategy establishes the right **evaluation questions** and **data sources** to determine if the short- and long-term outcomes—identified in our theory of change—are actually occurring due to ESAT.

We propose using a measurement strategy based on Blueprint's common outcomes framework used across the Scaling Up Skills Development portfolio along with specific indicators related to this study. To identify these study-specific indicators, we consider four evaluation questions:

1) **What is the causal effect of ESAT on short-term outcomes?** To determine if participants are achieving short-term outcomes (i.e., they get better at understanding their own SES; try out new behaviours that help them grow; and strengthen their SES and use those skills beyond the program, including in job settings), we should investigate ESAT platform data and participant surveys (baseline, exit, and follow-up).[1] ESAT platform data contains both self- and instructor-assessments; this is important because instructors play a vital role in contributing to participant success by using a shared SES language.
2) **What is the causal effect of ESAT on longer-term outcomes?** To determine if participants are achieving longer-term outcomes (i.e., they have better experiences at work, including finding and keeping jobs and earning more; a stronger sense of wellbeing and belonging; and greater involvement in their communities), we should investigate participant surveys and linked administrative data (incorporating labour market

1 Note that our Phase 2 *Final Report* provided a longitudinal analysis of skill changes using ESAT platform data and staff- and participant-assessed scores.

outcomes).[2] Linking with Statistics Canada would allow for the collection of reliable data without burdening participants with additional surveys. Linkage is particularly critical in an RCT context since it ensures minimal follow-up attrition and preserves sample sizes needed to detect effects.

3) **Who is being reached by the RCT?** To understand the generalizability of the RCT, we need to understand participant characteristics, including age, gender, race, disability, immigrant status, and labour market history.[3]

4) **What is ESAT's effect on instructor outcomes?** To determine if instructors are achieving intended outcomes (i.e., they feel more confident teaching and supporting SES development and give clearer, more meaningful feedback, using common terms and participant input), the RCT should include instructor self-assessment surveys (at baseline and exit) and participant exit surveys.

Table 4 aligns each evaluation question with outcomes, indicators, and data sources.

Table 4 | Randomization approaches

Evaluation questions	Outcomes/indicators	Data sources
What is the causal effect of ESAT on participant short-term outcomes?	• Employment readiness[4] • Participant self-perceived: ◦ SES; ◦ SES self-awareness; ◦ behavioural changes; and ◦ application of SES in new contexts after completing the program • Instructor perceptions of participant SES.	• ESAT scores • Participant surveys (baseline, exit, and follow-up)
How do ESAT effects vary by participant demographics?	• All indicators from above • Participant socio-demographics • Delivery context information	• Participant surveys • Administrative data • Instructor survey (baseline)
What is ESAT's effect on instructor outcomes?	• Instructor perceptions of: ◦ development of common SES language with participants; ◦ development of instructors' reflective SES practice; ◦ willingness to integrate SES into programs; ◦ ability to give effective SES feedback; and ◦ ability to give evidence-based SES feedback.	• Instructor survey (baseline and exit) • Participant exit survey

2 Our previous study also reported on participant employment and earnings outcomes (from baseline to nine months post-program) using participant surveys. These offered correlational evidence of effects, whereas an RCT can establish evidence of causation.

3 While our Phase 2 *Final Report* collected sociodemographic information on participants—including gender, age, immigration status, and highest level of education—it did not analyze participants' experiences or outcomes based on these demographic differences.

4 We suggest measuring employment readiness using the Employment Readiness Scale (ERS: Employment Readiness Model), which is effective at predicting job success within 12 weeks. This tool has been validated with diverse populations, including social assistance recipients, making it particularly relevant for ESAT.

4.1.5. Sample size

Recommendation 7. Assuming instructor-level randomization, the RCT should recruit a minimum of 4,882 participants, split evenly between a program and control group (2,441 in each). This assumes that 40 instructors participate (with 20 in each group).

We must determine the minimum number of participants required for an RCT to have enough statistical power to detect ESAT's effects. To do so, we need to estimate how large an impact ESAT might have on key outcomes. Sample size depends on the magnitude of the effect we expect to detect. To ensure we can confidently detect even modest effects, we need a large sample. This helps reduce the chance that observed differences are simply due to random variation.

We base our predictions on previous research on ESAT, comparator data from other Blueprint-run RCTs,[5] and broader literature on the effectiveness of labour market programs. We used the following process to arrive at the numbers:

Estimating ESAT's effect size

- **Program group employment rates.** First, we estimate ESAT's effect on employment rates at three months, using the three-month employment rates of ESAT participants from our Phase 2 *Final Report*: **41% (49/119)**.
- **Control group employment rates.** In comparator RCTs, average employment rates of control group participants increased by **12 percentage points** (ppt) from baseline to three months. We add this rate of change to the percentage of ESAT participants who were employed at baseline (from our Phase 2 *Final Report*). In this case, **20%** of participants were employed at baseline; thus **32%** of control group participants will be employed at three months.
- **Comparing employment rates.** Comparing these figures suggests a **nine ppt** effect of ESAT on employment (**41%** minus **32%**).
- **Variations in total effect.** Since we face considerable uncertainty in estimating the total potential effect of ESAT plus associated programming, we can consider a range of possibilities for this effect size. For purposes of effect estimation, we can consider a **small** total effect (**4.5 ppt**), a **medium** total effect (**9 ppt**), and a **large** total effect (**13.5 ppt**).
- **Attribution of ESAT.** In an ESAT RCT, the program group would receive essential or technical training *as well* as ESAT training, creating the **nine ppt** difference between groups. Therefore, we need to determine what percentage of the effect of training plus ESAT could be attributed to ESAT alone. Assuming that ESAT represents half (**50%**) of this effect results in a **4.5 ppt** effect of ESAT on employment three months after exit. While the control group would remain at **32%**, the program group would have a **36.5%** employment rate (i.e., **32% + 4.5%**). This is a small-to-moderate effect size in terms of employment RCTs.
- **Variations in attribution of ESAT.** As with the total effect of ESAT plus associated programming on employment, we face uncertainty regarding the proportion of the effect that can be attributed to ESAT. For the

5 At the time of this writing, Blueprint is conducting two RCTs of other training programs with similar population groups to ESAT as part of the Scaling Up Skills Development portfolio. These projects involve NPower Canada and In Motion & Momentum+.

purposes of effect estimation, we can also consider a small attribution (**25%**), medium attribution (**50%**), and large attribution (**75%**).

Using effect size to estimate sample size

- Taking the **4.5 ppt** effect associated with a **medium** total effect and **medium** ESAT attribution, we used a standard statistical method called a power calculation to determine how many participants we need to detect this difference. We assume statistical power of **0.8** (we want to be **80%** confident that our study will detect a real effect if it exists). We also assume a **5%** chance of a false positive, or of detecting a difference when none exists (this is a standard significance level of **$p < 0.05$**).
- Following this equation, we find that we need a total of **3,844** participants. However, we calculate our sample size assuming random assignment of instructors, rather than participants. When we randomize via instructors instead of participants, we randomize a smaller number of groups, and when smaller numbers of groups are randomized, there is a greater chance that the groups are different post-randomization simply due to random chance. Thus, we need to increase our total sample size. This increases the required sample size by approximately **27%** due to variations introduced by differences between instructors in each group.
- This calculation results in a target sample size of **4,882**, or **2,441** in each of the program and control groups. These numbers are required to detect an effect of this size in an RCT context where randomization is carried out at the instructor level. This assumes that **40** instructors participate (with **20** in each of the program and control groups).

Table 5 , on the following page, indicates the range of potential sample sizes needed to detect an effect (the total effect of ESAT plus associated programming on employment outcomes and the proportion of this effect that can be attributed to ESAT). In lower-impact scenarios, an RCT would require an unfeasible number of participants; in higher-impact scenarios, it would require a large but plausible number of participants.

Table 5 | Sample sizes based on effect estimates

Total effect size (ESAT + intervention)	% of effect attributed to ESAT (and net ESAT effect)	Individual randomization		Instructor randomization	
		Number of participants	Number of instructors*	Number of participants	Number of instructors*
4.5 ppt	**25%** (1.125 ppt)	60,668	505	77,074	642
4.5 ppt	**50%** (2.250 ppt)	15,118	125	19,200	159
4.5 ppt	**75%** (3.375 ppt)	6,694	55	8,502	70
9.0 ppt	**25%** (2.250 ppt)	15,432	128	19,599	163
9.0 ppt	**50%** (4.500 ppt)	3,844	32	4,882	40
9.0 ppt	**75%** (6.750 ppt)	1,700	14	2,159	17
13.5 ppt	**25%** (3.375 ppt)	6,888	57	8,748	72
13.5 ppt	**50%** (6.750 ppt)	1,720	14	2,185	18
13.5 ppt	**75%** (10.125ppt)	762	6	968	8

*Assumes 120 participants per instructor, per year.

"Instructor-level randomization offers the cleanest test of ESAT's impact and best reflects how programs are delivered."

Recommendation 8. Given these sample size and instructor requirements, the RCT should be implemented across multiple service delivery organizations, delivering a standardized version of the same programming that integrates ESAT.

In addition to the large number of participants required for the sample sizes noted above, many instructors need to participate to facilitate the training of this many individuals and to support instructor-level randomization. For example, in the base scenario of a **nine ppt** total impact and an attribution of **50%** of that impact to ESAT, we expect that **40** total instructors would be required—each delivering training to approximately **six** cohorts of **20** participants per cohort over a period of **two** years.

Table 6 | Potential instructor and cohort numbers in medium effect, medium attribution scenario

Required sample size	Number of instructors	Participants per cohort	Cohorts per instructor (over two years)
4,882	40	20	6

This poses a significant logistical challenge for an RCT in that it is very unlikely that a single service provider organization would employ this many instructors. Therefore, the RCT would likely need to be conducted across multiple organizations, all delivering the same model. To minimize service provider-level variation in the intervention, service offerings and implementation approach should be as standardized as possible between these service providers before launching the RCT.

5. Summary of recommendations

While an RCT of ESAT poses design and implementation challenges, it is feasible as a way to test both the tool's effectiveness in driving a range of outcomes and its ToC. To design and implement this RCT, we offer the following recommendations:

1. Refine ESAT using Phase 2 reliability and validity findings before pursuing an RCT.
2. Service delivery partners should deliver ESAT according to the modified, seven-part version of Futureworx's implementation guidelines.
3. An RCT can be implemented in any appropriate program context (i.e., any program type or setting), but the RCT should take place in the same setting and type once chosen.
4. An RCT should implement an instructor-level randomization approach, in which participation in ESAT is randomized at the level of individual instructors. If this is impossible, a cohort-level randomization approach is a second-best option, in which each cohort of a program is randomly assigned to implement ESAT.
5. When using instructor- or cohort-level randomization, collect basic participant information (such as age, gender, and employment status) before randomization. These can be used to match cohorts or instructors with similar participant profiles into pairs. Within each pair, one is randomly assigned to the ESAT group and one to the control group.
6. An RCT should adopt a four-part measurement strategy that relies on survey data for short-term indicators and linked administrative data for long-term indicators, particularly labour market outcomes.
7. Assuming instructor-level randomization, the RCT should recruit a minimum of **4,882** participants, split evenly between a program and control group (or **2,441** in each). This assumes that **40** instructors participate (with **20** in each group).
8. The RCT should be implemented across multiple service providers delivering the same program to make these target participant and instructor numbers feasible.

"A small-scale, proof-of-concept RCT could build capacity and de-risk a larger implementation."

Future Skills Centre | Centre des Compétences futures

Blueprint

Canada