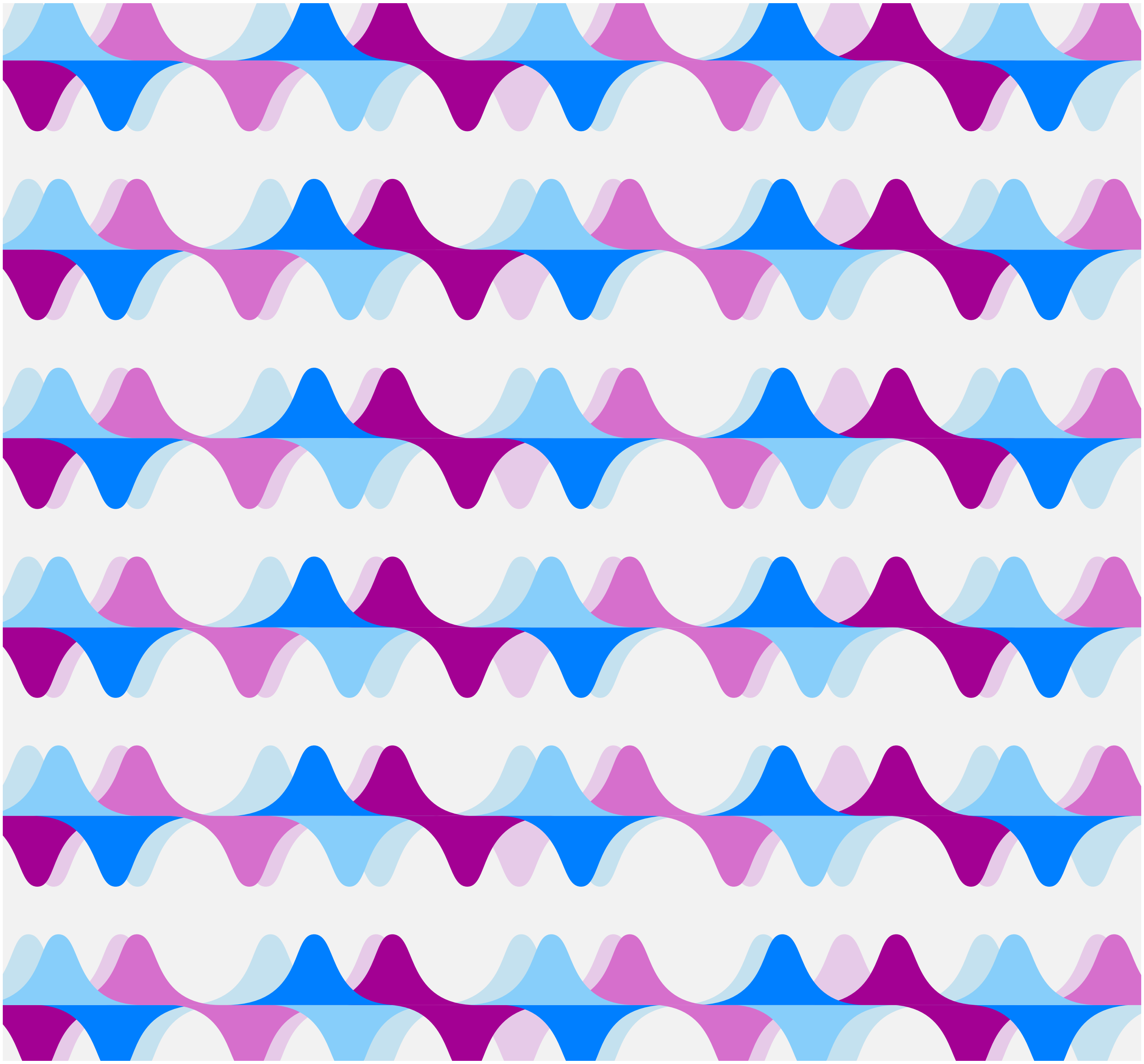


Leveraging Government Data Initiative: Benchmarking Model

Design Report

June 2024

Blueprint

This report was produced as part of a project funded by the Future Skills Centre (FSC), with financial support from the Government of Canada's Future Skills Program.

FSC is a forward-thinking centre for research and collaboration dedicated to preparing Canadians for employment success. We believe Canadians should feel confident about the skills they have to succeed in a changing workforce. As a pan-Canadian community, we are collaborating to rigorously identify, test, measure, and share innovative approaches to assessing and developing the skills Canadians need to thrive in the days and years ahead. The Future Skills Centre was founded by a consortium whose members are Toronto Metropolitan University, Blueprint ADE, and The Conference Board of Canada

Future Skills Centre | Centre des Compétences futures

Canada

Table of Contents

About the Future Skills Centre

The Future Skills Centre (FSC) 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 was founded on the simple idea that evidence is a powerful tool for change. We work with policymakers and practitioners to create and use evidence to solve complex policy and program challenges. Our vision is a social policy ecosystem where evidence is used to improve lives, build better systems and policies and drive social change.

Our team brings together a multidisciplinary group of professionals with diverse capabilities in policy research, data analysis, design, evaluation, implementation and knowledge mobilization.

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.

Building data capacity of the skills ecosystem

Unlocking the power of data to better meet the changing needs of Canadians.

Unlocking the power of data is foundational to a future-state skills ecosystem where public services are navigable, supportive, targeted, integrated and transparent. Through our experience with frontline practitioners, community leaders and government, we've identified three challenges that slow our progress:

1. **Community data capacity gaps.** Service providers collect data to meet reporting requirements, but these data rarely enable rapid learning, continuous improvement and impact measurement. Service providers struggle to know if they are maximizing impact.
2. **Friction within data-linking.** Opportunities to combine administrative data (e.g., tax records) with program data—like Statistics Canada's Social Data Linking Environment—could provide seamless ways to track long-term outcomes—but they are under-utilized. Long-term outcomes from skills training remain difficult to track.
3. **Under-leveraged datasets.** Governments collect large volumes of data that are not used due to constraints in time, technology and skills. The ecosystem is missing insights about service effectiveness.

Building data capacity

In response, we launched the Building Data Capacity portfolio to show how we can put data to work for Canadians. We are doing so via two workstreams:

Leveraging Government Data (LGD)

We are testing the use of government data holdings to understand needs and service demand, support continuous learning and improve outcomes.

- Using StatCan data to measure long-term outcomes of Canadians participating in our **Scaling Up Skills Development** programs; and
- Developing a **benchmarking** model that leverages the Labour Market

Program Data Platform (LMPDP) to identify outcomes benchmarks for employment and training programs.

Practitioner Data Initiative (PDI)

We are providing funding, tools, advice and technical assistance to nonprofit organizations to help them better collect, manage and use data.

- Providing funding, advice, technical assistance and tools to 15 non-profits to help build their capacity to collect, manage and use data; and
- Producing evidence on initiative outcomes to support scaling and build a sustainable path to help more organizations develop data capacity.

1. Introduction

About the Leveraging Government Data: Benchmarking Model

To understand the impact of workforce development programs, interventions need to be measured using rigorous methods. A randomized controlled trial (RCT) is considered the 'gold standard' of doing so, but it requires time, resources and expertise, and cannot be feasibly utilized on an ongoing basis. Benchmarking outcomes offers an alternative approach. By systematically comparing the performance of programs against a 'benchmark,' organizations can monitor the effectiveness of their models without needing complex research designs.

Blueprint is developing a **novel benchmarking model** to help funders and service delivery organizations forecast the performance of new interventions and enhance them over time. Our model uses existing microdata—describing the characteristics and outcomes of historical labour market interventions—and Bayesian Multilevel Modeling (BMM), an adaptable, highly accurate and scalable approach to understanding program outcomes. Through BMM, we can balance the trade-off between capturing context-specific variation and providing statistically precise, generalizable predictions—in other words, **we can account for differences in the design and context of past interventions while making predictions that are generalizable to new programs**. We can also provide rigorous counterfactuals without the substantial resource and technical costs currently required for experimental research designs.

Currently, our model uses data from projects in the FSC-funded **Scaling Up Skills Development** portfolio. Our aim is to complement this with data from ESDC's Labour Market Program Data Platform (LMPDP), which will greatly enhance the model's relevance and utility. Ultimately, the model is designed to offer guidance to practitioners, tailored to their specific target audiences, program features and contexts. In so doing, it supports both effective design and delivery.

About this report

This report describes Blueprint's preliminary designs and objectives for our benchmarking model. Our work is divided into five sections:

1. **Context** (pgs. 6–8) notes the benefits and drawbacks of RCTs and quasi-experimental methods; some features of existing benchmarking models; and how our model improves on these designs.
2. **Learning Agenda** (pgs. 9–10) presents our research questions, aligned to the innovation cycle and developed to ensure continuous improvement and learning throughout the initiative.
3. **The Benchmarking Model** (pgs. 11–18) explains our model: its analytical features, objectives, structure, steps toward development, components, and current and future data sources.
4. **Example Case Study** (pgs. 19–21) provides an example of a practical application of the model in developing and enhancing a hypothetical workforce development intervention.
5. **What's Next?** (pg. 22) offers a look at upcoming goals and phases of reporting, noting that outcomes and learnings derived from initial model performance will follow in a future report.

Context

The most reliable way to measure the effectiveness of a workforce development program is through an RCT or a quasi-experimental method. While credible, these approaches are expensive, require a high level of expertise and are not always feasible for community organizations or governments to conduct routinely. Benchmarking is another cost-efficient way to supplement existing efforts and significantly increase the supply of credible evidence about what works, for whom and under what conditions. This process involves comparing a primary model against a 'benchmark' to optimize inputs, outputs and overall effectiveness for decision-making.

In designing our model, Blueprint reviewed several approaches with similar goals to identify design considerations and constraints. Below, we describe representative benchmarking strategies, outlining the advantages and limitations of each.

Government target setting

One example of a government-administered benchmarking system is the Australian Star Rating System, used by the Department of Employment and Workplace Relations to rate its employment service providers publicly. The model calculates expected and actual performance ratings for each site based on numerous indicators using a regression model and controlling for various participant information (e.g., age, gender, unemployment duration, etc.) and labour market variables (e.g., labour market strength, urban/rural categorization, etc.). The Department assigns providers a rating of one to five stars depending on their ratio of actual-to-expected performance and how it compares to other providers.

This system provides data-driven expectations of what different interventions should achieve and establishes a record of public documentation. Nevertheless, like many other funder-driven, target-setting strategies, it leaves it up to individual providers to estimate how their program design affects outcomes. It does not benchmark the impact of varied program designs on outcomes—and thus misses a crucial opportunity to improve all programming.

Academic inquiry

Academic studies also aim to identify the drivers of effectiveness within workforce development models—a line of analysis which is a key element of benchmarking. As a case study, we look to "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations"[1] for its careful consideration of over 200 studies of active labour market programs. As a 'meta-analysis,' the study provides a broad overview of the impacts of programs across different demographic groups and economic conditions. Card et al. generate findings specific to the effectiveness of labour market interventions for different population groups, economic contexts and time scales.

The authors stratify findings by program type, participant characteristics and labour market conditions at the time of program operation, among other variables. Due to high variation in the programs captured in the meta-analysis, this stratification lacks fine detail but is a necessity to achieve the authors' desired statistical rigour. While it provides specific and rigorous conclusions about labour market programming, its findings cannot be used to create targeted benchmarks or inform program design beyond a very high level.

Targeted benchmarking

Another approach involves targeted efforts to generate replicable benchmarks for workforce development interventions. As a case study, we consider the US-based Workforce Benchmarking Network (WBN). In 2011, the WBN launched a data collection program attempting to create a national dataset with the goal of offering credible benchmarks that could be used by nonprofit organizations to understand the effectiveness of their workforce development programming. To build this dataset, WBN reached out to community-based organizations to report their job placement and retention outcomes, as well as features of their clients and delivery approaches. They used Analysis of Variance (ANOVA) to identify program characteristics correlated with better outcomes and created benchmarks using the median and 75th percentile outcomes for different combinations of program components.

WBN's approach was limited by the nature of the data—aggregated reports of outcomes from organizations, as opposed to client-level microdata—which prevents a more nuanced understanding of how client characteristics can relate to outcomes. Moreover, using the ANOVA technique for identifying effective features provides only a binary indication of whether a feature is associated with better outcomes, and not the magnitude of this association. More granular data and a more sophisticated statistical model could generate more precise and valuable predictions.

1 Card, D., Kluve, J., & Weber, A. (2018). What works? A meta-analysis of recent active labor market program evaluations. Journal of the European Economic Association, 16(3), 894–931. https://doi.org/10.1093/jeea/jvx028

Implications for our model

The preceding examples illustrate efforts to produce program-level benchmarks. They also highlight several limitations that our model can improve upon:

- **Limited consideration of intervention features and interactions:** Previous models vary in how well they consider intervention features and their interaction with participant characteristics and contexts. In accounting for intervention features and allowing the importance of these features to vary across contexts, our approach can provide **more contextually specific benchmarks**.
- **Model designs do not support prediction:** Much of the work undertaken to date either does not directly support prediction (such as Card et al.) or integrates prediction only as part of a relatively opaque system (such as the Australian Star Rating System). Our approach focuses primarily on **program-level predictions to support practical decision-making; it creates intuitive outputs to ensure accessibility**.
- **Limitations in data and output richness:** Some existing approaches are not able to fully make use of the richness of workforce development administrative data—such as the reliance on standardized results of other studies for Card et al. or aggregate reporting for WBN. Our approach can focus specifically on participant-level microdata to **better capture the interactions between context and outcomes**. It can also provide **more detailed distributions of potential outcomes as part of its prediction process**.

2. Learning agenda

Blueprint developed a learning agenda to guide the design of our benchmarking tool.

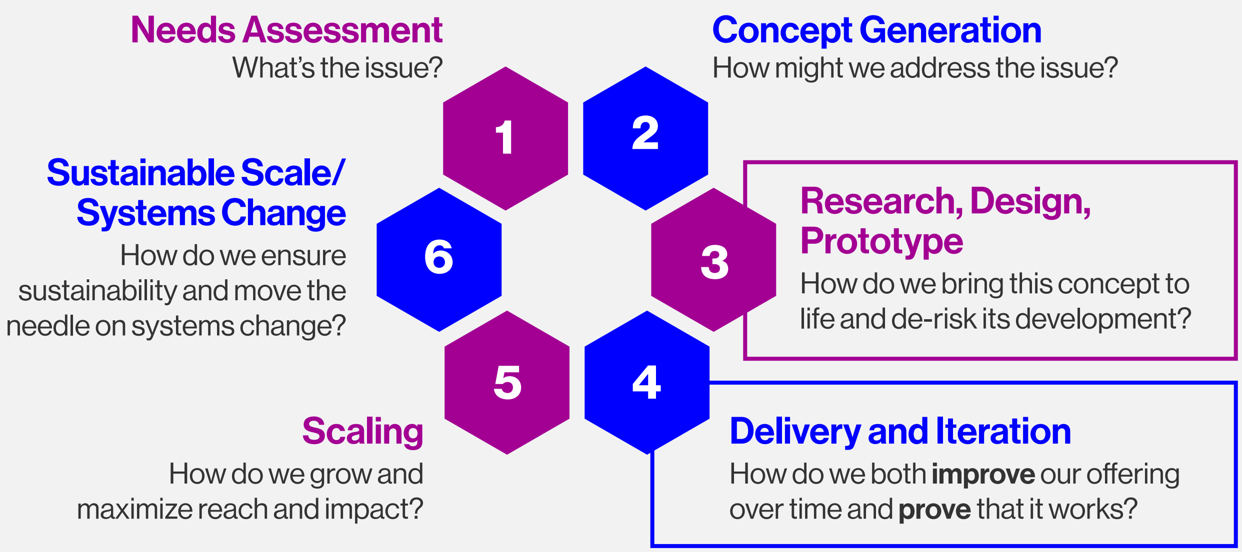
2.1 Aligning questions with the innovation cycle

Across all our portfolios of work, Blueprint aims to move the innovations we support through a six-stage cycle using our evidence generation toolkit (see **Figure 1**). In a well-functioning innovation ecosystem, innovations start with a needs assessment, move to conceptualization and design and then to delivery, testing and iteration. For the interventions that are proven to work, the goal is to expand to meet the need at scale and create system changes to institutionalize the innovation.

Knowing where an intervention is in the innovation cycle allows us to ask the right questions and generate the right evidence to move the project forward. Fostering early-stage innovation requires understanding and assessing complex issues, generating new and untested concepts and exploring the feasibility and desirability of these concepts with stakeholders. Projects that have moved into the delivery, testing and iteration stage are typically ready for evaluation.

Our current focus is on generating evidence that advances our benchmarking model through Stage 1: Needs Assessment; Stage 2: Concept Generation; and Stage 3: Research, Design, Prototype. As we continue to iterate with the model, we will explore methods for Delivery and Iteration (Stage 4) and Scaling (Stage 5) with the ultimate goal of Sustainable Scale and Systems Change (Stage 6).

Figure 2 | **The Six-stage Innovation Cycle**



2.2 Learning questions

We will explore each question below in a Final Report.

1. **Needs Assessment**
   a. What are the major barriers government and nonprofit stakeholders face in setting data-driven expectations for their programs?
   b. What approaches currently exist for setting benchmarks? What are their strengths and gaps?

2. **Concept Generation**
   a. What model designs would best support these addressed needs?
   b. What data sources exist to support the development and refinement of the model?

3. **Research, Design, Prototype** (User Testing)
   a. Given the data sources available, in what situations does the model perform better or worse?
   b. What use cases may have the highest impact—and how can the model be effectively deployed in these use cases?

3. Findings

3.1 Model objectives and features

Our benchmarking model can assist stakeholders with the design of workforce development programs by providing guidance on expected outcomes. It does this by:

- analyzing data on the characteristics, participants and outcomes of past labour market interventions;
- accounting for differences in the models, participants and contexts of these past interventions; and
- using these data to make generalizable predictions about how well new interventions will work for future participants.

To make fair comparisons, the process of generalization must be sensitive to differences between programs: their participants, goals, models and delivery contexts. For example, a training program focused on employability skills will likely yield different outcomes compared to one designed to upskill workers for high-skilled roles. Similarly, even comparable models will result in different outcomes when delivered in diverse economic contexts or to different population groups.

However, the need to be sensitive to the unique context of each intervention involves a trade-off between specificity and generalizability. Separately analyzing data from different programs and participants to capture how they differ may yield highly specific data, but it may also result in an unmanageable number of unique models, each with small samples and low statistical precision, making it difficult to arrive at reliable generalizations. Conversely, pooling data from across programs and participants results in large samples and statistically precise generalizations but may not speak to the contextual differences of greatest interest to program designers and evaluators.

Therefore, when trying to predict the outcomes of new interventions, the optimal approach is to balance these trade-offs; it should attempt to capture variations in outcomes associated with divergent contextual factors while still providing precise, generalizable predictions. Our model navigates this trade-off using an advanced machine learning technique known as **Bayesian Multilevel Modeling (BMM)**. BMM accounts for program and population differences without sacrificing statistical precision, allowing us to learn what works well, for whom and in what contexts. The results are more reliable estimates of effectiveness, even for underrepresented groups, in less common delivery contexts, and in new combinations of program, context and target populations.

BMM has five key benefits:

1. **Adaptability.** BMM is highly adaptable to diverse program designs and demographics. By analyzing past data at the program and participant levels simultaneously, it can adjust to accommodate the specifics of different scenarios, ensuring that the modeling approach is always suited to the data it analyzes. Similarly, the model is 'parametric,' meaning that the features used while fitting the model can be customized based on the background knowledge of program designers and evaluators.

2. **Accuracy.** BMM consistently maintains higher predictive accuracy for outcomes than other statistical approaches because it partially pools information across participants and contexts. This partial pooling allows the model to resist overgeneralizing in contexts for which little data is available, borrowing information from other contexts to inform its predictions in a way that has proven, on average, to increase predictive accuracy. This robustness is essential for delivering reliable insights, especially when program-related decisions are based on these predictions.

3. **Utility.** BMM uses an advanced algorithm to compute the uncertainty of its predictions in a more comprehensive way than other statistical models typically used in this domain. This ensures that users have an accurate understanding of the most likely outcomes of a program for a certain participant and the full range of other likely outcomes for that participant. This more accurate picture of uncertainty provides stronger guidance around expected vs. surprising outcomes and identifies areas where past programs do not give definitive answers.

4. **Replicability.** A key advantage of BMM is that it can seamlessly incorporate new data as they become available. While other methods can result in interpretations of the evidence that change drastically with the addition or exclusion of certain data points, BMM can use new data to continuously update past results.

5. **Capacity to generate insights.** BMM is designed to be transparent and interpretable, which contrasts with more opaque 'black box' machine-learning methods. For example, while many machine learning methods output a prediction without indicating why it was made, BMM also returns parameter estimates that correspond to specific program and participant characteristics, allowing users to see clearly which factors most influenced a given prediction. By analyzing these parameter estimates, users can identify important interactions between variables dictating the success of interventions—such as the interplay between participant demographics and program specifics. This granularity of analysis stands in contrast to purely predictive black box approaches, fostering a more nuanced understanding of what works (and what doesn't) to facilitate continuous improvement in program design and implementation.

3.2 Model structure

Below, we outline the steps used in developing the model and provide a definition of its inputs, outputs and the relationships between them. We also detail the variables included in the model, setting the stage for a more in-depth technical discussion available as a separate document . For a visual representation of the model, see **Figure 2** on **pg. 15**.

Development steps

While the four steps used in constructing the model are presented sequentially, in practice, development happens iteratively. For example, we may repeat steps one to three to make improvements while step four is carried out as needed to generate new predictions for new interventions.

1. **Collect data.** As a first step, we collect data on workforce development interventions through periodic, comprehensive assessments of overall performance or monitoring (i.e., ongoing processes to track progress and adjust accordingly during implementation). As outlined in Section 3.4, for our initial design, we are leveraging data collected from interventions in the Scaling Up Skills Development portfolio. These data include the individual characteristics and outcomes of participants served by the interventions, contextualized by information on each model design and local economic conditions during delivery. For example, for a scaling initiative focused on improving technical skills in the manufacturing sector, we would collect data on participant demographics, prior education and employment status, as well as post-training employment outcomes, wages and job retention rates. Additionally, we would gather contextual data on the local economic conditions during the intervention's implementation.

2. **Aggregate and standardize data.** To integrate these data into our model, we aggregate them across programs and standardize them where needed to ensure they align with definitions outlined in our Common Outcomes Framework (see **Box 1**). Continuing with the manufacturing sector intervention example, we would aggregate data from multiple cohorts over several years, standardizing participant characteristics such as age, education level and employment status. We would also standardize outcome measures, such as job retention rates and wage increases, to align with our Common Outcomes Framework.

Box 1 | Common outcomes framework

Our measurement approach includes indicators that are specific to an intervention as well as a set of common indicators that are measured for every intervention in the **Scaling Up Skills Development** portfolio.

These common indicators are drawn from Blueprint's Common Outcomes Framework, which was developed in consultation with our partners and was informed by review of employment-related outcomes frameworks and measurement approaches both within Canada and internationally.
They include:

- Intermediate outcomes that reflect 'in-program' participant experiences and gains (e.g., program satisfaction and skills development).
- Long-term outcomes such as employment and educational attainment.

Using a consistent approach to measuring outcomes is part of our commitment to understanding how each intervention in the Portfolio is reaching people across Canada and allows us to measure long-term outcomes using Statistics Canada's Social Data Linking Environment.

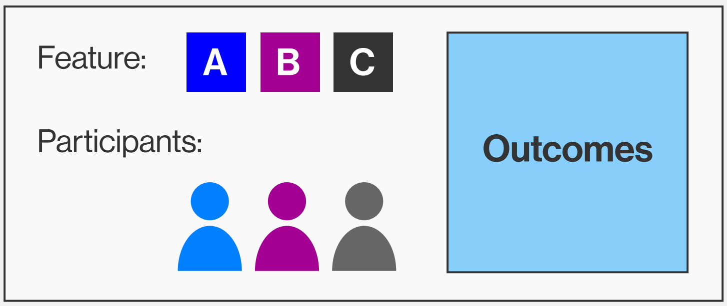
For more information on Blueprint's Common Outcomes Framework, see **Appendix A**.

3. **Model relationships.** Using this integrated dataset, we apply the BMM approach to model the relationships between participant characteristics, program features, context and outcomes. Using a multilevel modeling approach allows these relationships to vary. For example, participant characteristics can have different associations with outcomes depending on intervention design features. Continuing with our manufacturing sector example, we would model how different levels of prior education and local economic conditions influence post-training employment outcomes. This would help us understand how the effectiveness of the intervention varies for different participant groups and contexts.

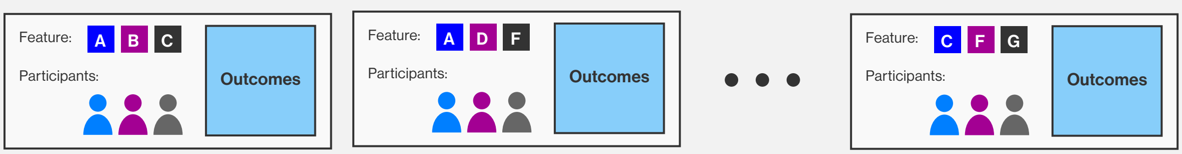
4. **Predict outcomes.** Once the model has analyzed the input data, it can be used to generate predictions for new interventions. By **inputting** the design features of the intervention and expected (or actual) participant characteristics, the model will **output** the expected distribution of employment and earnings outcomes for participants, providing users—both funders and service providers—with a benchmark of what they should expect from the intervention. In our manufacturing sector example, we would predict the employment outcomes for participants in a new manufacturing skills program in a region with similar economic conditions and participant demographics to previous cohorts. This would allow us to tailor interventions to specific contexts and improve overall effectiveness.

Figure 2 | Steps in training and using the model

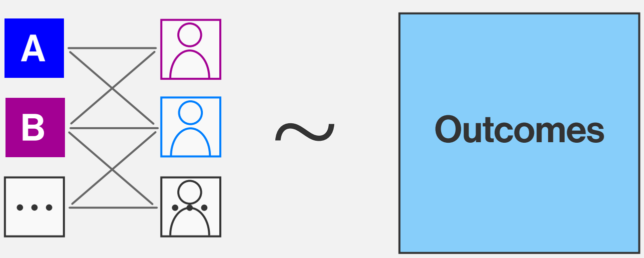
Step 1
Collect data on participant characteristics, program features and outcomes from programs



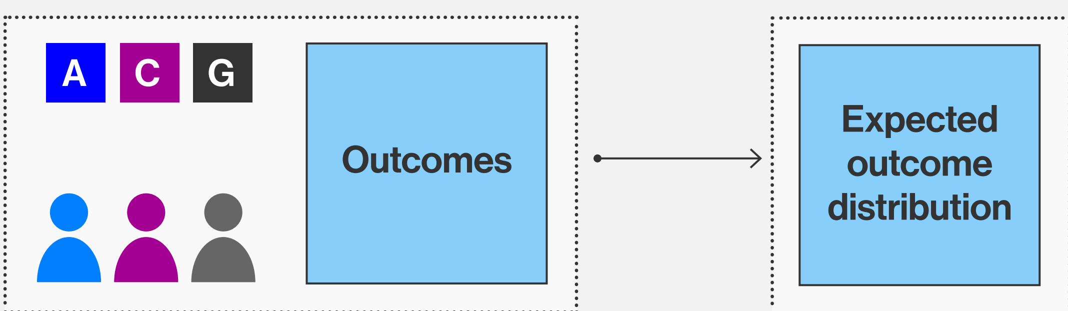
Step 2
Aggregate a large data set of programs



Step 3
Model relationships between design features, participant characteristics and programe outcomes



Step 4
Use model ro predict range of outcomes for new programs



3.3 Model components

In the steps outlined above, we discussed three core components of the model:

Inputs	**Relationships**	**Outputs**
or the data to be analyzed	or the parameters the model generates between participant characteristics, program design features, context and outcomes	or the predictions the model creates about new interventions based on their characteristics

These three components are discussed in greater detail below.

Inputs

Data inputs can include a range of sources from workforce development interventions, such as information on participant demographics, details about the training received and employment outcomes. Incorporating diverse datasets allows our model to capture a comprehensive picture of the program landscape and provide a robust base from which to draw predictions. Table 1 describes the categories of variables we currently include in the model, sorted by participant characteristics, intervention features, intervention contexts and outcomes.

Table 1 | Variable categories currently used in the model

Participant characteristics	Intervention features	Intervention contexts	Outcomes
• Age • Gender (and transgender status) • Educational attainment (and highest education attained in Canada) • Employment status • Marital status • Parental status • Identifies as having a disability • Identifies as racialized • Identifies as Indigenous • Newcomer • English as a second language (ESL)	• Duration • Delivery channel (e.g., in-person, virtual, hybrid, etc.) • Delivery schedule (e.g., scheduled classes, flexible, etc.) • Skill focus (e.g., technical skills, employability skills, etc.) • Wraparound supports • Job placement component	• Geographical location • Local employment rates • Local median earnings	• Employment status • Salary • Overall program satisfaction • Would recommend the program to others

While these variables represent our current focus, this set will grow as we further standardize the dataset used to support the model, develop a comprehensive typology of intervention features and integrate new sources. In other words, the model can incorporate new data types on participant and intervention characteristics as well as new outcome predictions, such as skill gains or enrolment in further training, as they become available.

Moreover, the model can integrate data generated from different sources, such as surveys of program participants or administrative data directly collected by service delivery organizations or funders. Current and future data sources are described in greater detail in **Section 3.4.**

Relationships

At its core, the model operates by identifying patterns and **relationships** within data and making generalized predictions about them: it works by linking **inputs** (data on program features, participation characteristics, etc.) to **outputs** (predictions about participant outcomes). The model is structured to recognize and adjust for variations across different programs and participant demographics, thereby enhancing the accuracy of its predictions.

In practice, these relationships are estimated at the program and individual level. The multilevel nature of the model means that the strength of these relationships can vary with different intervention designs or participant characteristics. For example, the model might identify that technical training programs are particularly effective for participants within a certain age range or educational background. Beyond mere prediction, the strength of these inferred relationships can provide insights into the overall match between stakeholder needs and choice of intervention.

Model outputs

The model then predicts intervention outcomes as its outputs—the expected distributions of employment and earnings outcomes for a given workforce development intervention. These outputs are designed to be actionable and easily interpretable to assist stakeholders in making informed decisions. Because we use a Bayesian model, predictions are presented as a distribution instead of a single number, recognizing that while there may be one outcome that is most likely, there are others that are less likely but still plausible. This format helps users understand whether the observed outcomes are expected, surprising or somewhere in between.

These outputs can also be used to predict intervention impacts against historical comparators and 'no program' scenarios. For example, for a program that aims to upskill unemployed participants, the model might output expected employment rates and earnings post-training based on other comparable interventions. The actual results of the program could be compared against these to see how close the outcomes were to analogous cases.

3.4 Data sources

Current data sources: FSC projects

We based early iterations of the model on project-specific data collected from initiatives in the **Scaling Up Skills Development Portfolio**. These projects provide key foundations due to their comprehensive inclusion of participant, intervention and outcome data. They also cover a wide spectrum of skill development programs designed to meet the evolving demands of the labour market and have direct relevance to the workforce development context.

Data from the Scaling projects were harmonized using our Common Outcomes Framework, ensuring standardized definitions. While data quality considerations (such as sample size, response rates and retention on follow-up) vary across projects, our aggregation process ensures standardization. The BMM framework also allows for analytical approaches that can reduce the impact of missing data on analysis results.

The primary shortcoming of the FSC portfolio projects is their small sample sizes. As the sample currently encompasses 11 projects with 5,500 participants, we are limited in both precision and generalizability. [2] Integrating further datasets—especially administrative data—will improve the working sample of the model significantly and is therefore our next logical step.

Future data sources: Leveraging the LMPDP

A critical way for us to improve the scope and representativeness of the model is to integrate additional administrative data from a broader range of databases held by delivery organizations and data holders at multiple levels of government—most importantly, **Employment and Social Development Canada (ESDC)** and **Statistics Canada**.

Specifically, ESDC's **Labour Market Program Data Platform (LMPDP)** contains data on the characteristics, program participation and outcomes of individuals across Canada participating in provincial workforce development interventions. While this dataset does not have the same level of detail on intervention characteristics as our FSC data, it could vastly improve the model due to its coverage of a far broader set of jobseekers with different individual and labour market contexts. Training the model on these wider datasets would help us capture a more complete spectrum of economic indicators, regional employment trends and educational outcomes. This will not only enrich our existing models but also help us identify underrepresented areas and emerging skill requirements.

We aim to integrate estimates drawn from this dataset as part of the next phase of model improvement. The BMM approach can leverage the strengths of each source of data—detailed, program-specific insights from FSC projects and broader context from national administrative records—allowing for more precise and more generalizable predictions.

2 Although there is no minimum sample size for this type of analysis, very small samples create model-fitting issues in practice.

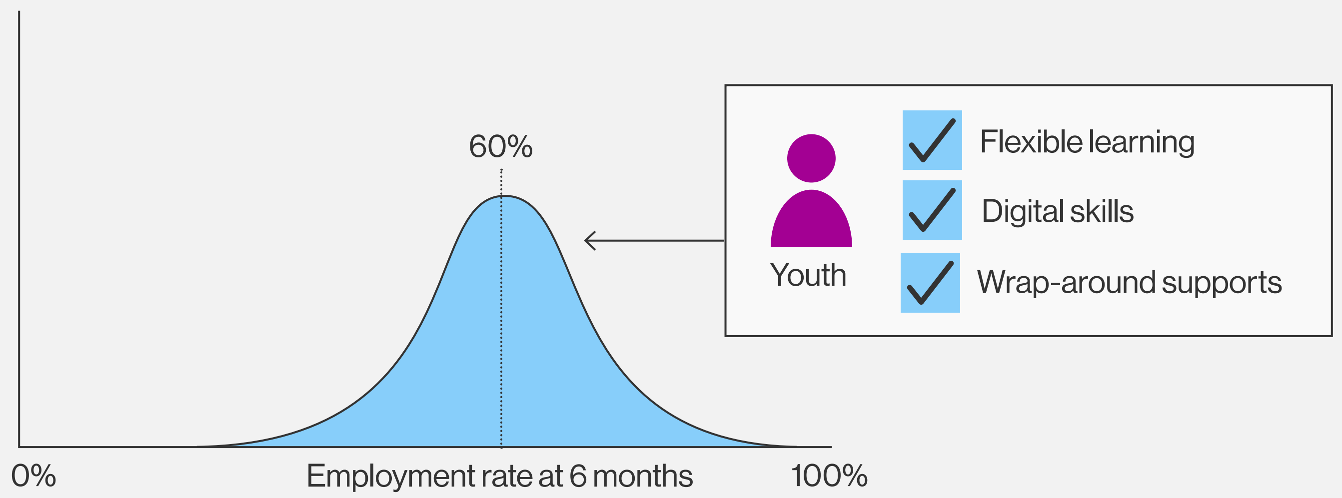
4. Example case study

The following case study illustrates how the model can be used to develop and enhance a workforce development intervention.

Phase 1: Strategic funding

A funder wishes to invest in employment programs for youth. To inform their understanding of what works, they use our model to identify program features with the greatest impact on youth employment outcomes: flexible learning schedules, integration of digital skills training and wrap-around supports. Our model allows the funder to visualize the distribution of outcomes considered plausible for youth who participate in programs with these elements. Figure 3 shows a hypothetical distribution of plausible employment rates for this program at six months, where it considers 60% to be the most plausible employment rate.

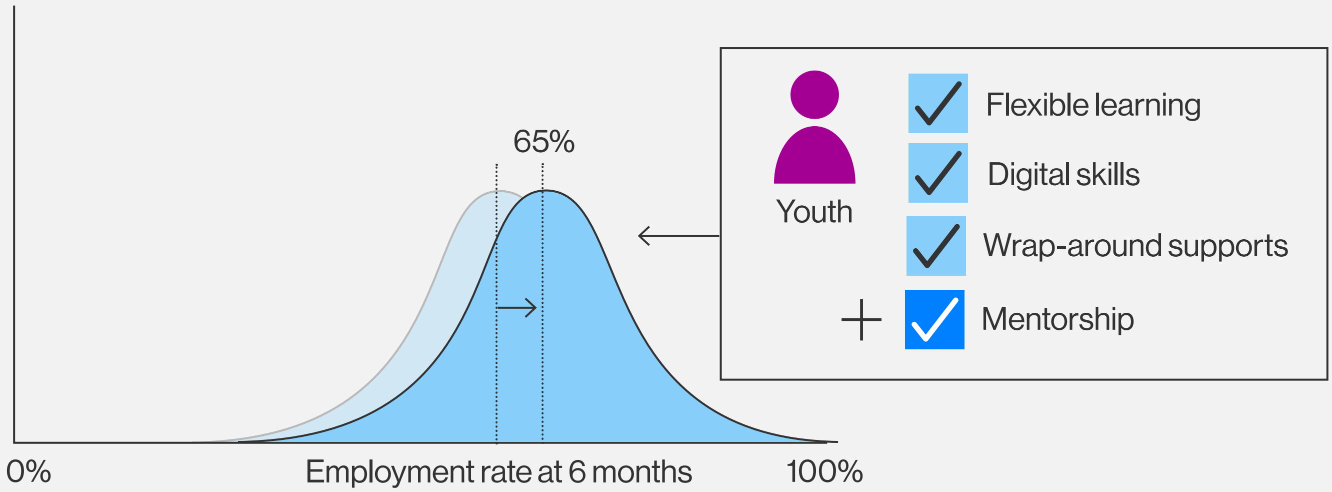
Figure 3 | Funding phase: Model's distribution of plausible outcomes given past data



Phase 2: Program implementation

One of the funded organizations, Youth Momentum, designs a program tailored to the specific needs of its youth client base. They input the characteristics of their participant demographic into the model to simulate potential outcomes. They also identify an additional feature that has historically resulted in better outcomes for youth: mentorships. Having added more specificity to the program design, our model allows the user to visualize the distribution of outcomes considered plausible for this more specific combination of program features and participant demographics. Now, as shown in **Figure 4**, the model shows the most plausible employment rate at six months is 65% rather than the 60% expected under the previous design.

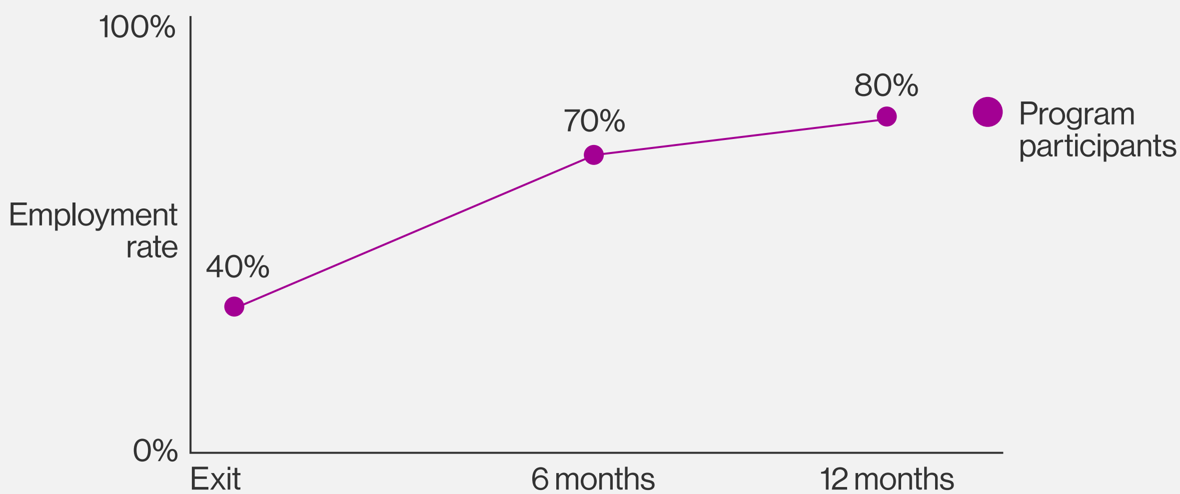
Figure 4 | Implementation phase:
Updated distribution of plausible outcomes given change in model features



Phase 3: Data collection

As the program progresses, Youth Momentum collects data on participant enrolment, engagement, satisfaction and success rates. These data include employment rates at program exit as well as follow-up employment rates at six and 12 months, illustrating the sustainability of employment gains. **Figure 5** shows these measured outcomes at three times, indicating that 40% of participants were employed at program exit, 70% were employed six months after exit and 80% were employed 12 months after exit.

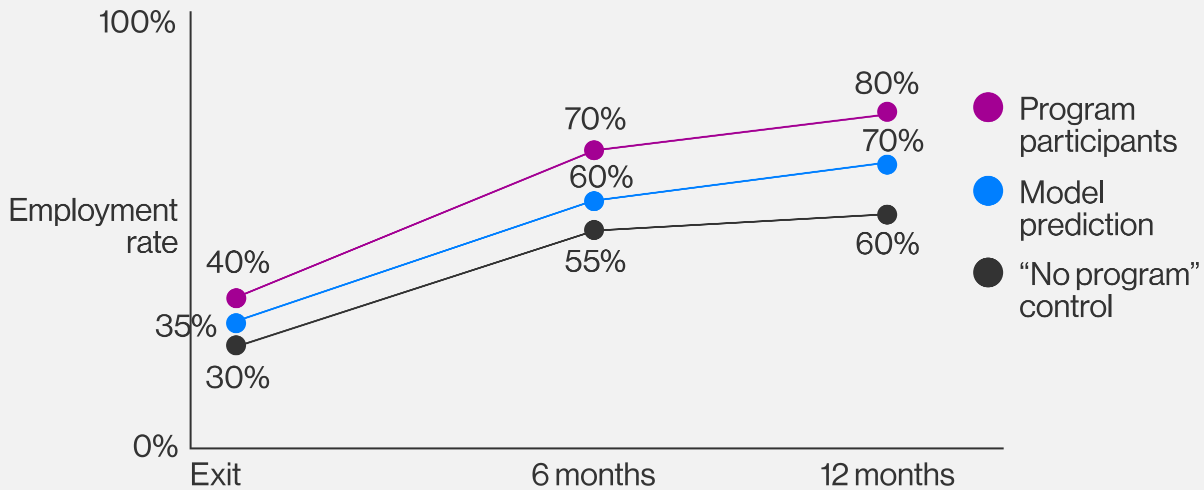
Figure 5 | Data collection phase:
Data on program completion collected during program delivery



Phase 4: Performance testing

Having collected data on participant employment rates, both the service delivery organization and the funder compare the actual outcomes with the initial model predictions and a 'no program' scenario, which serves as a control to estimate the employment outcomes of participants had they not received training through the intervention. The 'no program' scenario could be generated from historical data on similar individuals who did not engage in any training. Results show that the program not only met but exceeded the benchmark model's predictions, affirming the effectiveness of the targeted features and adjustments made during implementation. As shown in Figure 6, employment rates are noticeably higher than those of the control group, showcasing the added value of the program.

Figure 6 | Testing phase:
Data on program completion collected during program delivery



5. What's next?

Blueprint is currently refining the model, integrating data from both the **Scaling Up Skills Development** portfolio of projects and broader administrative data sets (such as the LMPDP, as they become available) to improve both the breadth and accuracy of our predictions.

Our upcoming Final Report will describe the model's initial performance and outputs across a range of contexts, supporting a stronger understanding of where its predictions can be used and at what level of confidence. The Final Report will also include a roadmap for ongoing development and dissemination of the model, focusing on use cases that sustainably support a range of stakeholders in better planning, iterating and measuring skills programming.

For a more in-depth technical discussion of the model's specification and structure, please see

Appendix A

Common Outcomes Framework

	Outcome	Indicators
Socio-demographics	Sex & Gender	Sex at birth
		Self-identified gender
	Age	Age
	Location	Province
		Region & Municipality
	Marital status	Marital status
	Children & Dependents	Children Dependents Household size
	Household Income	Household income
	Education	Highest credential obtained
		Location of highest credential attainment
	Indigenous Identity	Self-identified Indigenous identity
	Francophone status & languages spoken	First language spoken
		Official languages
		Language spoken at home
		Other languages spoken (At home)
	Citizenship Status	Place of birth
		Year of arrival
		Citizenship status
	Racial identity	Self-identification as member of racialized group
	Disability	Self-identified disability
Employment status and history	Employment	Employment status
		Nature of employment (permanent, temporary, full/part-time)
	Earnings	Hours worked / week
		Wages
		Annual earnings
	Industry and occupation of employment	NAICS code of job
		NOC code of job
	Work history	Time since last employed
		NOC code of job
		NAICS code of job
	Income source	Income sources

	Outcome	Indicators
Intermediate outcomes	Program completion	Successful completion of planned activities
	Participant satisfaction	Satisfaction with program
		Perceived Utility of Program
		Likelihood to recommend
Customized intermediate outcomes	Skills gains	Measured gains in specific skills
	Program-specific credential attainment	Attainment of program-specific credentials
Long-term outcomes	Employment and retention	Employment status
		Nature of employment (permanent, temporary, full/part-time)
		Retention
	Earnings	Hours worked / week
		Wages
		Annual earnings
	Benefits	Presence of benefits including: Paid leave, Health and dental coverage, Pension plan
	Industry and occupation of employment	NAICS code of job
		NOC code of job
	Job Satisfaction	Satisfaction with job
		Perceived opportunity for career advancement
		Perceived job security
	Enrolment in further education	Enrolment in further education
		Type of training
		Field of study
	Credential attainment	Attainment of high school or PSE credentials
		Field of study credentials

Future Skills Centre
Centre des Compétences futures

Blueprint