mdrc
BUILDING KNOWLEDGE TO IMPROVE SOCIAL POLICY

STRIVE EVALUATION ROADMAP

This brief presents a multi-year measurement, learning, and evaluation agenda to guide STRIVE's network enhancing and field building through the uptake of lessons learned and sharing of evaluation findings in a manner that can be applied by others focusing on the hardest to serve. The goal of this one-pager is to lay out a process map to get STRIVE ready for a rigorous impact evaluation. It is followed by a more fleshed out proposal.

Where does STRIVE fit in the workforce learning agenda? At its core, STRIVE can be classified as a sectoral training model. While there is a sizable body of evidence on the characteristics and effectiveness of sector programs, much still needs to be learned and STRIVE can contribute to this evidence base. One key question is: *Why do some sector programs achieve better outcomes and lead to larger economic impacts than others?* Going forward, it would be valuable to conduct research with STRIVE that focuses on how sector strategies can be more effectively and consistently implemented and what components or enhancements can strengthen the programs. This improvement strategy is a central goal of the STRIVE evaluation roadmap.

Evaluation roadmap. This learning agenda will address key questions important both to the STRIVE network and the field. It is rooted in several principles gleaned from past studies. One principle is to **not rush into an impact study**. We suggest continuing to focus on program improvement and capacity building first to develop a strong test. Several articles have documented how premature tests can do damage to the field.[1] For these reasons, we suggest starting with a formative feedback process study, which will build off work MDRC has already begun with STRIVE. A second principle is that a study is needed to help the workforce field understand **which components or enhancements can make sectoral programs like STRIVE stronger, rather than conducting yet another test of the full sectoral model.** For maximum impact, the research should focus on understanding programmatic insights rather than broad policy options. Finally, we will consider a variety of design options including quasi-experimental and experimental designs; rapid-cycle tests are preferred.

Figure 1 shows a three-step roadmap starting with a formative process study (ensuring key program improvement milestones are met), moving to a design and feasibility period for a rigorous test, followed by an impact study. Key moderators of STRIVE and MDRC's success along this journey include establishing a strong data infrastructure,[2] achieving program improvement outcomes (outlined below), obtaining stakeholder buy-in, situating the study within the workforce field, and asking the right research questions.

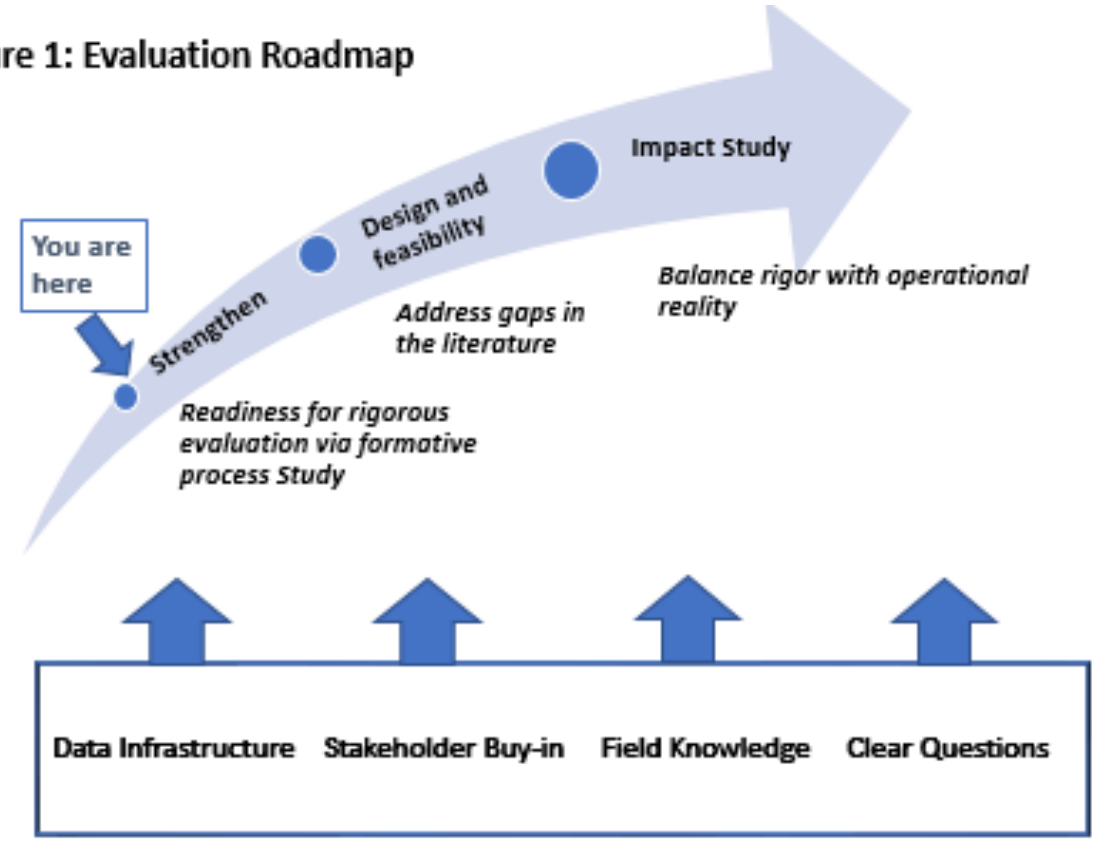
Step 1: Formative process study. A process study should come first, which would focus on strengthening mediating pathways including training completion, certifications, and job placement. **Step 2: Design and feasibility study.** After the program meets benchmarks and passes a logic model test, STRIVE can proceed to an impact design and feasibility study. Impact studies can take many forms and there will be numerous design options to consider. Given the state of evidence in sectoral training, the priority is to test one or more components of the model to make it work more consistently and to have larger effects. The goal of most rigorous impact evaluation designs is to estimate effectiveness by comparing groups of similar individuals – one group receiving the intended program model, component, or service strategy being tested (treatment group), and the other group usually reflecting the status quo (comparison group). For STRIVE, we suggest a systematic process when assessing impact evaluation designs that first considers whether a randomized controlled trial (RCT) would be appropriate, and then considers other designs. **Step 3: Impact Study.** The work of evaluation starts but does not end with design. During the

[1] Epstein & Klerman (2012), Nathan (2008).
[2] Based on our experience working with STRIVE, we believe this condition is in place.

implementation of a study, there are often numerous adaptions necessary in order to ensure that the study effectively tests the STRIVE program with as minimal effects as possible on program operations.

This evaluation roadmap is designed to contribute maximally to the workforce development field by testing STRIVE when the program is ready for rigorous evaluation, addressing the key questions with which the field is grappling, and using a design that balances programmatic needs and the knowledge-building agenda.

Figure 1: Evaluation Roadmap



Mini-Proposal Version {Funder facing}

This proposal presents a multi-year measurement, learning, and evaluation agenda to guide STRIVE's network enhancing and field building through the uptake of lessons learned and sharing of evaluation findings in a manner that can be applied by others focusing on the hardest to serve. The goal of this document is to lay out a process map for getting STRIVE ready for rigorous impact evaluation.

I. Introduction: Building Evidence of STRIVE's Impact

What is the STRIVE Program? For over 35 years, STRIVE has offered programming to foster both career development and personal development. The STRIVE network provides workforce development services in 12 cities throughout the country and has served over 75,000 participants who face significant barriers to employment, including disconnected youth, individuals living in poverty with unstable housing and substance use disorders, and individuals who were recently incarcerated. The STRIVE model focuses on placing individuals in career paths for high demand middle skilled jobs. The Career Path model has five components: 1. **STRIVE START** (Success Through Attitude & Readiness Training): This is the attitudinal training model that first made the STRIVE program nationally known. START focuses on problem solving, critical thinking, self-awareness and interpersonal skills; professional skills, including communication, decision-making, workplace ethics and professional etiquette; and work readiness, including career planning, labor market information, resume writing, and interview skills. 2. **Occupational Skills Training** (OST): This is a newer addition to the base STRIVE model. Following START, students take part in 6-8 weeks of sector-based occupational skills training where they earn industry-recognized credentials. Students choose from two industry pathways: Construction & Maintenance or Health & Office Operations. 3. **Career Coaching**: After finishing OST, STRIVE's employment specialists and career coaches match graduates with employers and develop a plan for long-term career growth. 4. **Job Placement**: STRIVE assists graduates in securing jobs that are a match with their skills and interests. 5. **Lifetime Support**: Career Coaches and the retention team actively follow up with graduates for two years to ensure continued success, and inform them about opportunities for additional training when applicable. All alumni have access to agency referrals that help them meet pressing needs that could hinder their ability to enter and thrive in the job market, including childcare, housing assistance, and access to professional clothing.

Where does STRIVE fit in the workforce learning agenda? At its heart, STRIVE can be classified as a sectoral training model. Over the past decade, the workforce development field has increasingly adopted sector strategies to meet the needs of both job seekers and employers. Sector programs like STRIVE train individuals for quality jobs in specific industries and occupational clusters where there is strong local demand and the opportunity for career advancement. The sectoral approach was initially pioneered by a few community-based organizations beginning in the late 1980s.[3] Increased interest in the approach began to take hold following the release of findings from the Sectoral Employment Impact Study (SEIS) in 2010. That study found earnings gains over a two-year follow-up period for three mature sector programs (all of which are nonprofits).[4] Following the release of those findings, sector programs have proliferated across the country and sector strategies were a key component of the Workforce Innovation and Opportunity Act passed in 2014.

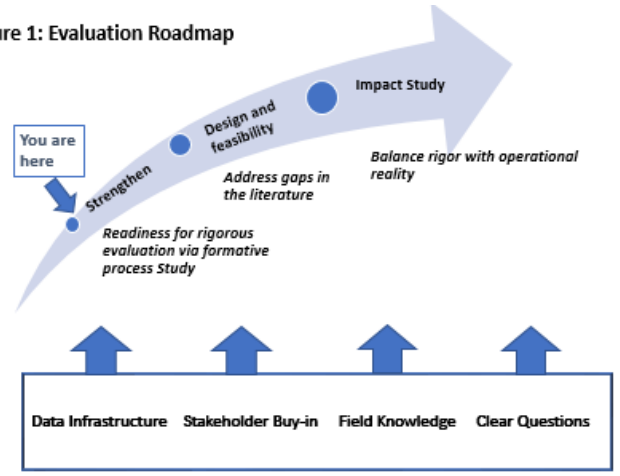
Open questions about programs like STRIVE. While there is a sizable body of evidence on the characteristics and effectiveness of sector programs, much still needs to be learned and STRIVE can contribute to this evidence base. One key question is: *Why do some sector programs achieve better outcomes and lead to larger economic impacts than others?* A recent synthesis by MDRC[5] shows that some programs produced larger earnings impacts than others. Understanding what is driving this variation – for example, certain program characteristics or the target

[3] Mangat (2007).
[4] Maguire et al. (2010).
[5] Schaberg (2020)

population – is important for both policymakers and practitioners to consider as they develop current and future sector programs. Some evidence is available on which program and organization characteristics are associated with impacts. An analysis done as part of the HPOG evaluation found that certain program characteristics were associated with the size of programs' short-term impacts. For instance, having access to employment supports and social services was associated with larger impacts on target sector employment.[6] Implementation analyses done in other evaluations, including SEIS, identified key organizational features of strong sector programs. These included having a stringent screening and intake process, developing strong employer connections, and providing individual tailored services.[7] Future research on STRIVE could further explore which participant and program characteristics are associated with better participant outcomes.

How can sector programs be strengthened to produce larger and more consistent earnings impacts? Sector strategies are widely accepted to increase participants' earnings. Yet even the programs that have the most promising evidence to date still have room for improvement. Figuring out why some participants fail to meet key program milestones will help programs better refine the services they offer. Additionally, identifying participants who are more or less likely to encounter challenges completing programs or obtaining jobs can provide insights into who is best served by sector programs and who may need additional supports to be successful. Going forward, it would be valuable to conduct research in this area that focuses on **how sector strategies can be more effectively and consistently implemented and what components or enhancements can strengthen the programs.** This improvement strategy is a central goal of the STRIVE evaluation roadmap.

Figure 1: Evaluation Roadmap



II. Evaluation Roadmap

This section highlights an **Evaluation Roadmap** that will set up STRIVE for an evaluation that will address key questions important both to the STRIVE network and the field.

Principles. The evaluation roadmap is rooted in several principles gleaned from past studies. One principle is to **not rush into an impact study**. We suggest focusing on program improvement and capacity building first to develop a strong test. Several articles have documented how premature tests can do damage to the field.[8] For these reasons, we suggest starting with a continuous improvement process study, which will build off work MDRC has already begun with STRIVE. A second principle which comes from our reading of the field is that a study is needed which helps the workforce field understand **which components or enhancements can make sectoral programs like STRIVE stronger rather than conducting yet another test of the full sectoral model.** For maximum impact, the research should focus on understanding programmatic insights rather than broad policy options. Finally, we will consider a variety of design options including quasi-experimental and experimental designs.

Figure 1 shows a three-step roadmap starting with a continuous improvement process study (ensuring key program improvement milestones are met), moving to a design and feasibility period for a rigorous test, followed by an impact study. Key moderators of STRIVE and MDRC's success along this journey include establishing a strong data

[6] Walton, Harvill, and Peck (2019).
[7] Maguire et al. (2010)
[8] Epstein & Klerman (2012), Nathan (2008)

infrastructure,[9] achieving program improvement outcomes (outlined below), obtaining stakeholder buy-in, situating the study within the workforce field and asking the right research questions.

Step 1: Continuous program improvement process and impact feasibility study

Focus on program improvement first. A formative process study should come first, which would focus on strengthening mediating pathways including training completion, certifications, and job placement. This work is already underway as described below.

Building off the current STRIVE-MDRC collaboration. The process and feasibility study can build on MDRC's ongoing STRIVE partnership, where MDRC and STRIVE have spent one year in a multifaceted partnership using behavioral insights and data analytics to improve STRIVE program processes, focusing on the flagship New York Career Path program. This collaboration identified key areas of program improvement and included several recommendations. Key focus areas identified for focus in the next stage of program improvement included: increasing rates of completion START and OST, and improving retention and reporting.

For some populations (particularly those with high rates of housing instability and substance use disorder), dropout rates are elevated. Increasing completion rates for these populations is a key goal before moving to an impact study, and would likely lead to stronger impacts as other studies have found that impacts are often largest for subgroups at higher levels of disadvantage.[10] Increasing START completion is an important priority given that an impact study (regardless of design) would likely need to include participants beginning with START and follow them from enrollment onwards. This design called "Intent to Treat" will result in a better test if more individuals are retained in START. The same logic applies to OST completion. MDRC found that while STRIVE has very high job placement numbers, OST completion numbers are somewhat lower than top tier sectoral programs. Regarding retention outcomes, MDRC included several recommendations to improve employment retention outcomes (including reporting). Finally, the MDRC study included multiple recommendations for program segmentation and tailoring. Populations of particular focus include younger participants (who are more likely to be placed, but have lower retention rates); the long term unemployed and older participants (both less likely to be placed); and, as discussed above, individuals with more housing instability and substance use challenges.

[9] Based on our experience working with STRIVE, we believe this condition is in place.

[10] https://aspe.hhs.gov/execsum/what-works-best-whom-impacts-20-welfare-work-programs-subgroup#I

Step 2: Design and feasibility study

At the conclusion of the formative feedback and program strengthening phase, we would transition to a design and feasibility study. We suggest an adaptation of a framework for conducting feasibility studies proposed by Epstein and Klerman to determine whether STRIVE meets the conditions necessary to launch a large-scale, rigorous study.[11] Epstein and Klerman warned of conducting rigorous impact evaluations on "plausible" programs too quickly, since these evaluations are costly and have high rates of negative or null results. They propose developing a "falsifiable logic model" – a logic model with specific intermediate benchmarks that can then be tested with a program pilot before a full evaluation is begun. If the program meets its benchmarks or demonstrates that the logic model has potential, then it is more likely to show positive effects in a rigorous (and more expensive) evaluation. If the pilot does not meet the benchmarks (i.e., fails its own logic model), the program would not yet warrant a rigorous evaluation. Instead, the program design may need to be refined, or implementation strengthened, and then the redesigned pilot may be retested to see if benchmarks can be met. Some common logic model failures include low program enrollment (inadequate sample sizes), low program completion rates (inadequate dosage to produce a treatment contrast), and lack of improvement over time. We will use the same principles identified in this framework to guide completion of the process study; a logic model will help clarify program benchmarks and a process study will enable formative feedback before proceeding to a summative evaluation.[12]

Research questions drive design. After the program meets benchmarks and passes the logic model test, STRIVE can proceed to an impact design stage. Impact studies can take many forms and there will be numerous design options to consider. To reduce the scope of these options, it is important to have clear research questions. A key decision is whether to test a whole model or a component. As discussed above, given the state of evidence in sectoral training, the priority is to test the components of the model to make it work more consistently and to have larger effects. The review of the literature strongly points to a need to understand "the active ingredients" of sectoral programs; that is, the components or enhancements that can result in the programs having larger and more consistent impacts. This helps to focus the impact study: because the focus is on program components which have shorter term outcomes, tests can be done more rapidly compared to tests that focus on the effects of the whole model on longer term outcomes. In recent years, there have been many developments in the use of rapid cycle tests to optimize program components.[13] Multi-arm impact designs might be appropriate, and would provide evidence on the value added of discrete service strategies (e.g., isolating the effects of STRIVE's post-placement advancement assistance to job placement program). There is very little known about effective post-placement alumni services. While most programs include these services, it is unclear what the value added overall nor how to optimize them. While the efficacy of these services seems obvious on its face, effective evidence-based models of delivering post-employment services have been elusive. This provides an opportunity gap that STRIVE can address that would be valuable for the workforce field.[14]

Choosing an impact evaluation design. The goal of most rigorous impact evaluation designs is to estimate effectiveness by comparing groups of similar individuals – one group receiving the intended program model, component, or service strategy being tested (treatment group), and the other group usually reflecting the status quo (comparison group). For STRIVE, we suggest a systematic process when assessing impact evaluation designs that first considers whether an RCT would be appropriate.[15] RCT designs, often called the "gold standard" design, assure (in a well implemented RCT) that the treatment and comparison samples are equivalent, on average, at baseline.

---

[11] Epstein and Klerman (2012).
[12] Scriven (1967).
[13] https://www.mdrc.org/sites/default/files/MDRC_BIAS_Final_Report_2017_MDRC_FR.pdf
[14] https://www.mdrc.org/publication/increasing-employment-stability-and-earnings-low-wage-workers
[15] See the Clearinghouse for Labor Evaluation and Research (CLEAR) for definitions of level of evidence.https://clear.dol.gov/reference-documents/causal-evidence-guidelines-version-21

This means the difference in outcomes between the groups represents the impact of the program intervention, and thus RCTs have high internal validity. Within RCT designs, there are numerous decisions to be made, such as the timing of random assignment (RA), the unit of RA (e.g., students versus training classes), and the statistical design (e.g., how to adjust p-values for multiple comparisons). Other considerations involve the number and composition of research groups representing different service combinations, confirmatory outcomes, subgroups, and data analysis strategies.

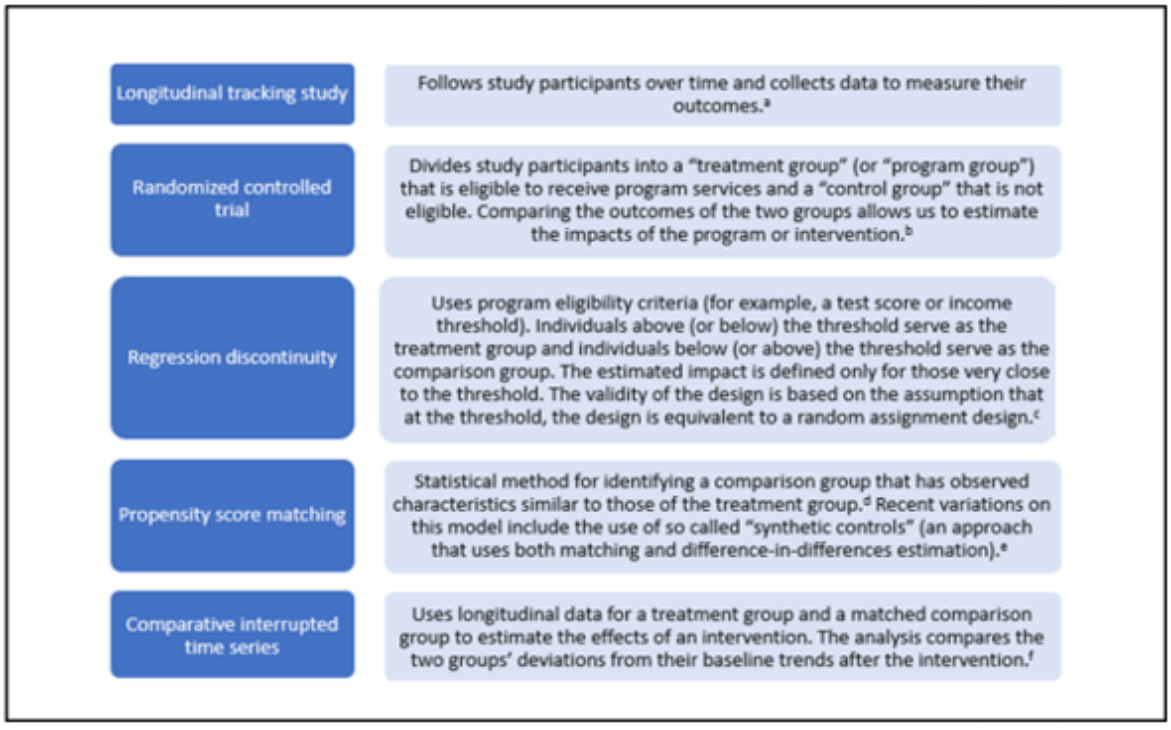
The process of choosing a research design is not always easy; each design has requirements and challenges. RCTs have practical drawbacks and are sometimes infeasible. If RCTs (including rapid cycle tests) are deemed infeasible, we would investigate the next most rigorous and appropriate design strategy. Other potential impact design options are shown in Appendix Chart 2, which shows several questions to consider when selecting a research design and which designs are affected by the answer. For example, *what is the research question of interest?* If it relates to understanding a program's effectiveness, a longitudinal tracking study would not be appropriate. Another example is: *are there clear and measurable eligibility requirements that are consistently implemented?* If so, a regression discontinuity design could be appropriate.

Step 3: Impact Study. The work of evaluation starts but does not end with design. During the implementation of a study, there are often numerous adaptions necessary to ensure that the study effectively tests the STRIVE program with as minimal effects as possible on program operations. See Schaberg (2019) for a discussion and case study of the kinds of accommodations that need to be made to successfully implement a research design.

Conclusion

This evaluation roadmap is designed to contribute maximally to the workforce development field by testing STRIVE when the program is ready for rigorous evaluation, addressing the key questions with which the field is grappling, and using a design that balances programmatic needs and the knowledge-building agenda. This plan will help establish STRIVE as a thought leader in the field by generating key lessons for policy and practice in workforce development.

Appendix Chart 1: Impact Design Options



Appendix Chart 2: Questions to Ask When Choosing a Research Design

Each design answers different research questions and use different evaluation strategies as shown in Chart 1 above. Note that there are several options within each design. Notably, rapid cycle randomized tests are recommended in this document to test components of STRIVE. This is one type of randomized controlled trial.

Question	Longitudinal tracking study	Randomized controlled trial	Regression discontinuity	Propensity score matching	Comparative interrupted time series	Adaptive design
What is the primary research question?	☑	☑	☑	☑	☑	☑
Is random assignment feasible?		☑				☑
Does the program have clear and measurable eligibility requirements that are used consistently?			☑			
Can a valid comparison group be identified?				☑	☑	
Are the key characteristics of program participants measurable?				☑	☑	
Are historical data — preferably at four or more points in time — available for both the treatment and comparison groups?					☑	