Evaluation Options for the Papers Developed Under the SSI Youth Solutions Project

November 22, 2021

Todd Honeycutt, Amal Harrati, Andrew Langan, Gina Livermore, and R. Vincent Pohl

Submitted to:

US Department of Labor
Office of Disability Employment Policy
200 Constitution Avenue, NW
Washington, DC 20210
Attn: Kirk Lew

Submitted by:

Mathematica
P.O. Box 2393
Princeton, NJ 08543-2393
Phone: (609) 799-3535
Fax: (609) 799-0005
Project Director: Todd Honeycutt
Reference Number: 50916.C1.T02.211.000

This page has been left blank for double-sided copying.

Acknowledgments

This report was prepared by Mathematica for the U.S. Department of Labor, Office of Disability Employment Policy, under contract number 1605DC-18-A-0020. The contents do not necessarily represent the policy of the U.S. Department of Labor and do not imply endorsement by the federal government, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

The authors gratefully acknowledge the input and review of staff from the Office of Disability Employment Policy and Jillian Berk, Alison Wellington, and David Wittenburg of Mathematica, along with the editing support of Susan Gonzales and production support of Sharon Clark.

This page has been left blank for double-sided copying.

Contents

This page has been left blank for double-sided copying.

Exhibits

This page has been left blank for double-sided copying.

Acronyms

ABLE	Achieving a Better Life Experience
CLEAR	Clearinghouse for Labor Evaluation and Research
CPWIC	Community Partner Work Incentives Counselor
CTE	career and technical education
CWIC	Community Work Incentive Coordinator
DDS	Disability Determination Services
DOL	U.S. Department of Labor
ED	emotional disturbance
FEAT	Family Employment Awareness Training
FEM	Family Empowerment Model
FES	family empowerment specialist
HHS	Department of Health and Human Services
IEP	individualized education program
IRT	integrated resource team
MCTI	Michigan Career & Technical Institute
MDI	minimum detectable impact
NSC	National Student Clearinghouse
ODEP	Office of Disability and Employment Policy
PERT	Postsecondary Education Rehabilitation Transition
PROMISE	Promoting Readiness of Minors in SSI
PTI	parent training and information
RCT	randomized controlled trial
ResPECT	Residential Postsecondary Education and Career Training
RSA	Rehabilitation Services Administration
SMART	specific, measurable, achievable, relevant, and timely
SNAP	Supplemental Nutrition Assistance Program
SSA	Social Security Administration
SSDI	Social Security Disability Insurance
SSI	Supplemental Security Income
TANF	Temporary Assistance for Needy Families
TESS	Transition to Economic Self-Sufficiency
TEST-CTE	Translating Evidence to Support Transitions in Career and Technical Education program
UI	unemployment insurance
VR	vocational rehabilitation
YFSN	youth and family systems navigator

This page has been left blank for double-sided copying.

I. Introduction

The Office of Disability and Employment Policy (ODEP) of the U.S. Department of Labor (DOL) initiated the SSI Youth Solutions project in 2019 to generate testable ideas for improving outcomes among youth receiving Supplemental Security Income (SSI). ODEP supported the development of 12 papers—a process that included input from independent peer reviewers and staff from ODEP and its federal partners. The ideas proposed in the papers build on the transition landscape to address some of the challenges that youth and families face in improving their employment and other outcomes. Two reports provide an overview of these papers: "Twelve Ideas to Promote Employment for Youth with Disabilities: An Introduction to the SSI Youth Solutions Project" briefly summarizes the papers developed under the project, and "Considerations for the Papers Developed for the SSI Youth Solutions Project" provides additional context and considerations for the 12 ideas described in the papers. DOL's website for the project (https://www.dol.gov/agencies/odep/program-areas/individuals/youth/ssi-youth) contains the 12 papers as well as additional resources, including recordings of webinar discussions with the papers' authors and policymakers on how to improve the employment outcomes of youth receiving SSI.

This report presents options for evaluating the interventions proposed by each paper, should they be implemented. A report from a previous ODEP project, the SSI Youth Formative Research Project (Honeycutt et al. 2018), offered a framework for policymakers to choose among potential ideas to meet specific policy objectives and fit within the existing landscape of supports, along with an appendix that proposed a broad evaluation scheme. In this report, we recommend evaluation options that policymakers and other stakeholders could consider when adapting and testing any of the ideas from the SSI Youth Solutions project. Each option describes the proposal and its associated demonstration, the evaluation design and rationale, evaluation details such as the sample and research questions, and other considerations. Throughout, we present ideal options to build evidence that, with demonstrated effectiveness of the proposal, could lead to broader implementation.

We use the following terms in this report

- **Proposal:** a policy, practice, or program change described in one of the SSI Youth Solutions papers
- **Intervention**: the implemented form of the proposal
- **Demonstration**: a test of the intervention
- **Participant**: a youth who enrolls in a demonstration
- **Lead organization**: the organization responsible for administering the demonstration

The following chapters present plans for evaluating the interventions proposed in each of the 12 papers. The report is not intended to be a primer on evaluation design or activities, but rather a guide for those who implement the interventions proposed in the papers. Because the interventions vary in their complexity and levels of evidence demonstrating their effectiveness, we recommend different evaluation approaches. Some interventions would require initial piloting to understand how they should be implemented and whether the proposed activities lead to the intended outcomes. Others have more evidence of effectiveness but would benefit from a rigorous test of their impacts. In some instances, this test could involve using existing administrative data to identify comparison groups, whereas in other instances, a randomized controlled trial (RCT) would be more appropriate.

A. Analytic and design considerations

Before presenting the framework we use to describe the evaluation options for demonstrations of the interventions proposed by the SSI Youth Solutions papers, we first discuss four key considerations for any evaluation: the analysis types, design and analytic methods, sample sizes and statistical power, and data sources. These considerations informed our thinking around the evaluation options we present in the chapters that follow.

1. Analysis types

In preparing the evaluation design options, we considered three types of analyses as being most useful for stakeholders: process, outcome or impact, and benefit-cost.

- **Process analysis.** A process analysis documents who enrolls in a demonstration, how stakeholders operate an intervention, and how well the activities match the intervention's design (Centers for Disease Control and Prevention 2011). The main focus of the process analysis is to understand whether the intended activities occur and why or why not. Aspects of an intervention's implementation include documenting organizations and partnerships involved in the intervention, funding and expenditures, staffing and staff training, program and service design, participant recruitment and enrollment, intervention dosage (such as type, amount, and duration of participant service use and participant engagement with the program), and the challenges and successes with program operations. A process analysis can also assess organization or system changes that are consistent with the intervention design, such as collaboration or policy changes. Early process analyses can document an intervention's startup activities, including initial service offerings, and can identify potential areas of improvement. It can also show how well participants' characteristics compare to both the specific population from which a demonstration draws (that is, the community) and the broader population intended for an intervention if it were scaled to a national level, along with whether a participant group reflects diverse populations and whether certain groups are overrepresented or underrepresented. Late or final process analyses show what services participants used and whether the implementation adhered to the intended model. Formative evaluation activities—describing implementation so that stakeholders can use the information to improve the intervention while they are implementing it—can be a part of the process analysis.
- **Outcome and impact analyses.** A primary aspect of an evaluation is to assess what changes as a result of an intervention and whether an intervention has its intended effects. An outcome analysis documents changes in key outcomes that are expected, as proposed by the intervention's logic model, along with the timeline of the expected changes (from short-term to long-term). An impact analysis assesses whether an intervention results in a change in outcomes by comparing the outcomes of a group that had access to the intervention with those of a comparison group that did not have access to the intervention. For the evaluation options of the SSI Youth Solutions proposals, we focus primarily on one outcome that most proposed interventions expect to affect: competitive integrated employment. Other outcomes include enrollment in postsecondary education and training and SSI benefit receipt.
- **Benefit-cost analysis.** With a benefit-cost analysis, an evaluator compares the cost of an intervention (obtained through a program's administrative records) with its benefits (calculated using findings from the impact analysis), based on various specified assumptions for the timing of the benefits. The analysis often requires a consideration of different perspectives (such as youth and family, the agency

or organization leading the intervention, and the federal government) to which different benefits and costs apply.

2. Design and analytic methods

The Clearinghouse for Labor Evaluation and Research (CLEAR) guidelines (2019) provide a useful approach for considering methods and designs tailored to each evaluation option. The guidelines emphasize the level of causal evidence in support of an intervention, or how well the results can be considered to have occurred because of an intervention. We summarize the CLEAR rating system here; readers can find more details on the CLEAR website (https://clear.dol.gov/).

CLEAR uses a three-tiered rating system to assess causal evidence for studies intended to show the effectiveness of an intervention. Studies use different methods to produce causal evidence, though most use a comparison group to document outcomes in the absence of the intervention and then compare the outcomes to those of the intervention group (even if that comparison is simply to the outcomes of a group before they use an intervention). CLEAR evaluates those studies in terms of the extent to which the relationship between an intervention and an outcome is due solely to the intervention or due to other factors in addition to (or instead of) the intervention. Also important is the extent to which the comparison (or control) group is similar to the intervention (or treatment) group. Note, too, that the CLEAR rating is independent of whether an intervention has an effect on the outcome; the rating reflects the evaluation design's potential to detect effects.

To meet the requirements for a *high rating* of causal evidence, the estimated effects must be solely attributable to the intervention. Only evaluations that use well-executed RCT or interrupted time series designs can achieve a high rating. Designs that control for some, but not all, confounding factors (those factors that might also influence the results) can receive a *moderate rating*. This rating means that although there is some confidence that the intervention caused the observed effects, there remains a question as to whether one or more factors other than the intervention led to the observed effects. Study designs that use a difference-in-differences model design, for example, might fall into this category. If the evidence does not meet the criteria for either a high or a moderate rating, then it receives a *low rating*, as the causal claims for the relationship between the intervention and the effects could be due to factors other than the intervention. For example, studies that use a pre/post design to measure changes in outcomes or that show correlations for a group of youth who do or do not use an intervention would receive a low rating.

Evaluations using RCT designs, if well designed and implemented, control for confounding factors by randomizing either individuals or groups to receive the intervention. These designs assume that those offered and not offered the intervention differ only because of chance; therefore, any differences in outcomes between the two groups can be attributed to the intervention. With clustered randomized assignment, the evaluation randomizes groups (such as school districts or vocational rehabilitation [VR] offices) to offer the intervention. This approach has the benefit of potentially being simpler to implement (because individuals are not randomized), but it increases the size of the minimum detectable impact (MDI) that the evaluation can detect because there is variation not only across individuals but also across the clusters.

Because experimental designs provide the strongest level of causal evidence of an intervention's effects, we have recommended RCTs for most of the SSI Youth Solutions evaluation options, provided the demonstration meets two conditions. First, the existing evidence in support of the intervention is low to

moderate (according to CLEAR criteria), and so the intervention would benefit from a more rigorous evaluation of its effects. Second, there is evidence that the design can be implemented well (that is, there is nothing to suggest that an experimental design is infeasible). Despite our recommendations for experimental designs, these designs can be challenging to implement, especially for those conducting the study recruitment and enrollment. Staff who recruit youth and families into an intervention might find it difficult to explain to families that they have been randomized into the control group and therefore are only eligible for usual services.

For certain evaluation options, we recommend designs other than RCT, such as those described below.

- **Pilot tests** without a valid comparison group are warranted with novel interventions that have not been implemented or tested and when stakeholders need information about whether the intervention can work as proposed. Although these designs might compare the outcomes of an intervention group with those of another group (for example, to previous or average outcomes), these comparisons are descriptive and not intended to imply a causal relationship between the intervention and outcomes. The information collected for a pilot test can be used to adapt the design of the intervention and develop more rigorous testing.
- **Difference-in-differences models** compare the change in intervention group outcomes over time with the change in those of a comparison group drawn from administrative data. This quasi-experimental design uses regression modeling to control for observed differences (that is, confounding variables) between participants and non-participants. The primary advantages of these designs are that they can be less costly and easier to implement than RCT designs. However, identifying a good comparison group can be difficult, particularly if youth and families must sign up for and enroll in an intervention (thereby indicating a level of motivation to participate or interest in obtaining the intervention's proposed outcomes). In addition, these designs do not control for all extraneous factors that could influence outcomes, and thus they can be open to criticism that the intervention might not be responsible for the observed effects.

Evaluators must weigh the level of evidence that they will generate with any one design alongside other considerations to determine the best design for an intervention. RCTs can generate the most rigorous evidence, though that design is not always practical. Identifying a comparison group that is well-matched to the intervention group by means other than random assignment might provide enough evidence for stakeholders to make decisions about an intervention. Depending on the intervention, context, and resources, evaluators might consider experimental designs using interrupted time series or other quasi-experimental designs, such as matched comparison groups, fixed-effects models, or instrumental models. If so, they might consult the guidelines presented in CLEAR (2019) to ensure that their evaluation produces results that meet CLEAR standards.

3. Sample size and statistical power

Stakeholders and evaluators must determine the right study sample sizes given the implementation options, expected impacts, and resource availability. An evaluation requires sample sizes of groups or individuals that are of sufficient size to detect impacts large enough to be meaningful to stakeholders. Evaluations should use the appropriate-sized sample for the intervention's implementation, given expected impacts. With small samples, the intervention's impact must be relatively large to be detectable, and large impacts might be implausible, thus potentially squandering the resources invested in the demonstration. With larger samples, smaller impacts can be detected, but having a sample that is larger than necessary can be costly. We do not recommend specific sample sizes for the evaluation options in

the report. Stakeholders and evaluators will need to conduct informed power calculations to make decisions about the sample size needed for an intervention that reflects appropriate assumptions and contexts.

To inform the power calculation, it is important to establish a meaningful effect size for the outcomes associated with the intervention. For example, what percentage-point change in youth employment rates generated by the intervention would stakeholders consider a success? Once that is determined, then the sample size required to detect that level of change can be calculated using assumptions and information from relevant existing studies.

To illustrate the tradeoff between sample sizes and detectable impacts, we present estimates of the MDI across three outcomes for individual- and group-level randomization scenarios where the sample sizes differ (Exhibit I.1).[1] Stakeholders can use this information as a starting point for designing a demonstration, and they would need to calculate MDI specific to their assumptions and expectations. For example, a demonstration that includes 200 youth—half in the treatment and half in the control group—can detect an impact of 17.5 percentage points on competitive integrated employment in the past 12 months from a survey, assuming that the control group has a mean outcome of 29.3 percent. Such a large impact is likely unrealistic given that few demonstrations have generated effects of this size. If the demonstration includes 1,200 youth, the MDI for the same measure is reduced considerably—to 7.1 percentage points. Similarly, with group-level random assignment (such as offices or schools), having more groups results in smaller MDIs. With eight groups (and 50 people in each group), an evaluation could detect an impact on competitive integrated employment from a survey of 19.3 percentage points; increasing the number of groups to 24 reduces the MDI to 10.1 percentage points. As documented in the exhibit, using administrative data—which will include all participants, rather than a subset who respond to a survey—allows smaller MDIs.

The calculations in Exhibit I.1 rest on four assumptions that might be inappropriate in certain situations. First, we assume a two-tailed test of size 10 percent, meaning that the difference in outcomes could be in either direction (the treatment group mean is greater or less than the control group mean) and a *p*-value of less than 10 percent is sufficient evidence of an impact. Stakeholders might be more interested in a one-tailed test (the treatment group mean is greater than the control group mean) or impacts that have *p*-values of less than size 5 percent. Second, we assume that the evaluation will focus on all participants (that is, an intent-to-treat model). If take-up of the intervention is low, then the MDIs will be greater than what is shown in Exhibit I.1 and so may be unrealistic to expect from an intervention. Third, the response rate for the use of survey data is 80 percent, so for each scenario the MDI is based on a partial sample that is 80 percent that of the full sample. Fourth, the use of administrative data assumes that an evaluator can obtain information from all participants. If consent is needed from participants to obtain administrative data, then the 100 percent response rate assumption is unrealistic; the assumption used for survey data (80 percent response) might be more appropriate.

[1] The MDI is the smallest true impact of an intervention that researchers can detect using specific sample sizes and test parameters. For example, if the true impact is a 5 percentage-point increase in the employment rate, but the MDI is 10 percentage points, researchers would not be able to distinguish the estimated impact from zero.

Exhibit I.1. Minimum detectable impacts for three youth outcomes across individual- and group-level randomization scenarios

	Percentage in competitive integrated employment in the past year	Percentage enrolled in postsecondary education or training at interview	Total SSA payments during past 18 months
Individual-level random assignment			
Assumed control group mean	29.3	45.1	$10,768
Assumed control group standard deviation	45.5	49.8	$3,708
Minimum detectable impacts using survey data			
N = 200	17.5	19.1	$1,426
N = 400	12.4	13.5	$1,007
N = 800	8.7	9.5	$711
N = 1,200	7.1	7.8	$580
N = 2,000	5.5	6.0	$449
N = 3,000	4.5	4.9	$367
Minimum detectable impacts using administrative data			
N = 200	15.7	17.1	$1,276
N = 400	11.1	12.1	$900
N = 800	7.8	8.5	$636
N = 1,200	6.4	7.0	$519
N = 2,000	4.9	5.4	$402
N = 3,000	4.0	4.4	$328
Group-level random assignment			
Assumed control group mean	29.3	15	$10,768
Assumed control group standard deviation	45.5	35.7	$3,708
Minimum detectable impacts using survey data			
Groups = 8, N = 400	19.3	21.1	$1,569
Groups = 16, N = 800	12.6	13.7	$1,024
Groups = 24, N = 1,200	10.1	11.0	$819
Groups = 40, N = 2,000	7.7	8.4	$625
Groups = 60, N = 3,000	6.2	6.8	$507
Minimum detectable impacts using administrative data			
Groups = 8, N = 400	18.2	19.9	$1,484
Groups = 16, N = 800	11.9	13.0	$969
Groups = 24, N = 1,200	9.5	10.4	$775
Groups = 40, N = 2,000	7.3	7.9	$591
Groups = 60, N = 3,000	5.9	6.4	$479

Notes: *Individual random assignment:* Assumed control group means are from the PROMISE five-year evaluation for competitive integrated employment and postsecondary education and training (Mathematica, forthcoming), and from the PROMISE 18-month evaluation for total SSA payments. The estimates assume a two-tailed test of size 10 percent and power 80 percent, response rate of 80 percent, equal probability of assignment to the treatment and control groups, and an *R*-squared of 0.05. The scenario involving administrative data makes the same assumptions, except that the response rate is assumed to be 100 percent. *N* denotes total sample size (both treatment and control groups).

Group-level random assignment: Follows the same assumptions as individual random assignment along with assuming a group-level *R*-squared of 0.05.

PROMISE = Promoting Readiness of Minors in SSI; SSA = Social Security Administration.

If a demonstration requires youth to sign up to enroll, then it will require outreach to a recruitment pool that is larger than the needed sample size. The Promoting Readiness of Minors in SSI (PROMISE) demonstration, which recruited youth receiving SSI ages 14 to 16, had enrollment rates ranging from 16 to 43 percent across the six PROMISE programs. Assuming the enrollment rate is the midpoint of that range (30 percent), a demonstration would have to conduct outreach to 667 youth to obtain a sample of 200 enrollees and 4,000 youth to obtain a sample of 1,200 enrollees.

There may be instances where the above sample sizes cannot be met. One such instance is when a demonstration involves a targeted population to determine specific issues related to recruitment, implementation, and outcomes. Because the population is limited, the demonstration cannot feasibly obtain the recommended sample size. Examples include rural youth, youth with specific health conditions, or youth who reside on tribal lands. A second instance might occur with analyses by subgroup characteristics, such as comparing differences in outcomes by race and ethnicity or by household poverty status. Dividing the participant group into subgroups increases the size of the impact that an evaluation can detect. For example, if a demonstration has 800 participants and 25 percent of them (N = 200) are youth from a minority group, the MDI for employment using administrative data is 7.8 percentage points for the full sample and 15.7 percentage points for the subgroup of youth from a minority group. A demonstration might oversample youth from specific subgroups to improve the ability of the evaluation to detect impacts or differences.

4. Data sources

The choice of data to use for an evaluation comes with certain advantages and disadvantages. The potential benefits of administrative data, for example, include having complete information for all participants and accessing them over a long time period. A disadvantage of such data is that the measures they contain may be limited, given that they are used to administer a program rather than for research purposes. Survey data can contain rich information on a range of topics but can be expensive to field and include information only on those who respond. Qualitative data offer the most detailed information about youth and staff experiences with an intervention, but they can be burdensome to collect and analyze and may not represent all staff or participants. In the evaluation options, we typically propose a combination of data sources that represent an ideal approach to obtaining information about an intervention. We describe the types and features of each source below.

a. Administrative data

Administrative data include records to track program involvement and outcomes. These records may already exist (if related to an existing program) or a lead organization may need to develop them (if related to a new intervention). Evaluators can use periodic extracts to examine service implementation and fidelity to the intervention model. In some cases, a comparison group of people not using or offered the intervention can be found in these data. In the evaluation options proposed in this report, we mention several administrative data sources relevant to evaluating the SSI Youth Solutions interventions.

Intervention program data. Any demonstration requires documentation of enrollment, services, and outcomes. A lead organization can either draw on its existing data collection tools (such as a management information system) or build a new tool to collect this information. Regular extracts of these data could

provide information on how well the agency is implementing the intervention. Program data of this sort could be relatively easy to access in close to real time and at regular intervals to inform implementation.

- *Existing tools.* Local or state agencies or programs that already collect data on services, staff, and individuals could use their existing data to identify people to recruit, track enrollment and services, understand cost, and document outcomes. For example, if a state VR agency leads an intervention, then an evaluation could rely on the agency's existing records for assessing services and outcomes. These records could also be used to identify a comparison group if random assignment is not used. The implementing agencies or programs might be able to expand their data systems to include additional measures that are relevant for the intervention evaluation.
- *New tools.* If an existing system is unavailable, a lead organization must design and implement a new tool to collect data. The tool should include measures that are relevant for staff and participants in relation to the intervention and intended services and outcomes.

Social Security Administration administrative data. The Social Security Administration (SSA) maintains comprehensive data to administer its programs, and these data include linkages to annual earnings, Medicare and Medicaid data, and VR services through the Rehabilitation Services Administration (RSA)-911 case service report data. SSA can use these data for evaluation purposes with the appropriate agreements. A lead organization might need these data to identify appropriate youth for recruitment and enrollment into the demonstration. These data can also be important for tracking SSA-specific outcomes such as SSI application, receipt, payment amounts, and use of work incentives. Access to these data requires working with SSA staff. Though SSA benefit information can be obtained in close to real time, the linkages to other data sources might have time lags. Annual earnings data is typically not complete until the early part of the second calendar year after the annual earnings period (Czajka et al. 2018).

Medicaid. Medicaid records contain information on public health coverage and enrollment, service utilization, expenditures, and provider use. With permission, state Medicaid data could be used for recruitment and enrollment into a demonstration, as youth receiving SSI in most states are eligible for and receive Medicaid. These data might be easier to access than SSA administrative data for this purpose. State records might also show outcomes related to health insurance coverage if those outcomes are relevant to a demonstration. With consent from research participants, an evaluator might also be able to access the Centers for Medicare & Medicaid Services' research identifiable data, which is available from the Research Data Assistance Center (www.resdac.org).

Public use files of the RSA-911 case service report. State VR agencies send a rich and detailed set of quarterly information about the individuals receiving their services, along with outcomes such as quarterly earnings for those who completed services. RSA (https://rsa.ed.gov/performance/rsa-911-policy-directive) creates annual public use files that can be shared for evaluation purposes. An evaluation could draw a comparison group from these data or compare measures related to a VR agency's intervention (if statewide) with those for a group of similar agencies. A limitation of these data is that they do not contain individual identifiers, so services and outcomes cannot be tracked for individuals from year to year. Because these data are linked to SSA administrative records at the individual level, an evaluation using SSA administrative data could have that level of detail.

State unemployment insurance records. States collect quarterly earnings data from employers for the purposes of administering their unemployment insurance (UI) programs (Czajka et al. 2018). These data might not cover all forms of employment (such as federal or self-employment) and have limited

information on employment and employer characteristics. If the necessary agreements can be established, evaluators can use these data relatively quickly (within six months of a quarter) to monitor quarterly earnings.

National Student Clearinghouse. The National Student Clearinghouse (NSC; https://www.studentclearinghouse.org/) collects information on education enrollment and attainment from most postsecondary education institutions. With permission from participants, evaluators can obtain individual-level extracts from the NSC that document enrollment start and end dates, enrollment status for each semester, program of study, institution characteristics, and graduation and degree information. A key advantage of this data source is the ability to observe enrollment in private or out-of-state institutions, to which a state department of education might not have access.

b. Survey data

Asking people directly about their experiences and outcomes can provide detailed information that cannot be collected through administrative data. Surveys are often the only method to obtain information about expectations, health, employment characteristics, and services that people use outside those offered through the intervention. They can be administered via paper forms (mailed to participants or handed to them by staff), telephone interviews, or web-based forms. The timing can vary based on when the type of information is needed: at enrollment, after completing services or work experiences (such as by satisfaction surveys), or at various points or milestones after enrollment or program completion. Surveys, however, can be expensive to implement, particularly if conducted by telephone (because it requires interviewers to call) or by paper (because the data have to be entered into a database by hand). Online options can decrease both of these costs, but might not be appropriate for all populations (that is, those with limited Internet access). Of notable concern is the need to ensure that response rates are high enough to be representative of the population. When not all individuals eligible to complete a survey do so, weights may be necessary to ensure that the survey responses represent the demonstration sample.

c. Qualitative data

Qualitative data can provide rich information on complex issues that cannot be obtained quantitatively through administrative or survey data. Asking open-ended questions through focus groups or structured interviews can enable staff or participants to provide their perspectives on successes, limitations, challenges, and directions for intervention adaptations. Qualitative data might also be part of administrative data, such as treatment or employment goals and case notes from staff describing their interactions with clients, and survey data, such as questions about goals or employment characteristics. One of the key disadvantages of qualitative data is that they can be difficult and time consuming to analyze, though the richness of the data and the insights they offer can offset those disadvantages.

B. Framework for the evaluation options

In the chapters that follow, we use a common framework to describe the evaluation option we recommend for each proposal (Exhibit I.2). The framework is intended to describe key elements of a demonstration and evaluation, including the features of the intervention, a rationale for evaluating it, an appropriate evaluation design, and possible refinements or other considerations in implementing and evaluating the proposed interventions.

Exhibit I.2. Evaluation framework for the SSI Youth Solutions proposals

Heading	Description
A. Proposal and demonstration description	
Proposal description	Provides a high-level summary of the proposal, including the intended outcomes for the intervention and the activities needed to obtain those outcomes.
Demonstration description	Contains details about how to implement the proposal for evaluation purposes (that is, it describes the intervention to be tested). These details may differ from or expand on the information contained in the proposal.
Lead and partner organizations	Identifies the likely lead organization (the organization that would be responsible for sample recruitment and implementation of the demonstration) along with partners needed for a successful demonstration.
B. Evaluation rationale and overview	
Evaluation rationale	Justifies why the evaluation is needed. The rationale documents to funders and stakeholders the reasons for conducting the evaluation, such as showing that the proposal has a positive impact on services and youth outcomes compared to usual practices or indicating the best way to set up the intervention. In part, the rationale should justify the evaluation by describing the knowledge gap filled by the evaluation—that is, how the results will add to the evidence base about what is known about the intervention and its effects.
Evaluation design overview	Presents a summary of the evaluation design, sample, goals, and duration.
C. Evaluation design	
Sample and recruitment	Documents the sample and how it would be recruited and enrolled.
Design	Identifies the recommended evaluation design, including timelines needed for the evaluation activities.
Research questions	Specifies the core research questions that the evaluation would answer, along with the data and analytic methods needed to answer those questions.
Data collection	Provides an overview of the specific data sources required to obtain the measures needed for the evaluation, along with a consideration of special issues associated with each (such as consent and timing).
Analyses and reporting	Details the key analytical approaches necessary for different evaluation products, along with the timing of those products.
D. Proposal refinements and other considerations	Provides final considerations implementing or evaluating the proposal.

C. Evaluation options for the SSI Youth Solutions proposals

1. Considerations for evaluating the proposals

The primary purposes in evaluating the SSI Youth Solutions proposals are to document the feasibility of implementation and provide the most rigorous test possible of the interventions' effectiveness. Many of the proposals lack rigorous evidence that connects the intervention to the expected outcomes for youth receiving SSI. Strong evidence would provide stakeholders with justification for incorporating one or more of these interventions into existing programs or developing programs centered around a proposal.

In the chapters that follow, we suggest an approach to evaluate each proposal that represents an ideal or best-case scenario. To the extent possible, we typically offer an evaluation design that involves an RCT. While we considered less rigorous approaches, such as using a matched comparison group, we found that such groups were often less than ideal (for example, the matched comparison group could not account for bias in motivation to sign up for an intervention).

Many of the evaluation options face common issues and challenges related to the interventions:

- **Detecting policy meaningful effects.** Demonstration evaluations must account for the potential for the intervention to yield effects sizes that are meaningful to policymakers relative to an intervention's potential costs. An intervention that generates a large positive impact might not offset its costs if the costs are substantial. At the same time, an intervention could generate a small impact but could be widely implemented if the associated cost in money and time are extremely low. In deciding whether to pursue a demonstration, policymakers must carefully weigh whether the potential benefits are likely to offset the potential costs as well as other factors, including the length of time that benefits are likely to accrue and the potential value of nonmonetary factors (such as equity and inclusion or quality of life).
- **Funding full-scale tests of the proposals.** The evaluation options in this report represent ideal scenarios of implementation and evaluation that are typically resource intensive. Almost all of the proposals describe long periods of implementation and therefore require even longer evaluation periods. To capture the measures described in their logic models, we propose using multiple data sources (administrative data, surveys, and staff and participant interviews). RCTs at the individual level double the cost of recruitment. Stakeholders might decide, given resource constraints, which aspects of a proposal are most important for testing, which components of an evaluation could best indicate a proposal's success, and what timelines are long enough to identify a proposal's results.
- **Testing partial proposals or combining multiple proposals.** Stakeholders might consider testing aspects of a proposal or combining parts of different proposals into a novel intervention. Rationales for such pursuits include an extended length of time for a proposal or a desire to combine parts that fit a stakeholder's environment or service perspective. In addition, policy needs and environments constantly shift, and the SSI Youth Solutions proposals—developed in response to current policies—might require adaptions to fit in new situations. Any such adaptations would begin with similar foundational work as did the authors of the proposals: specifying clear services, intended outcomes, and the target population, and developing a logic model that links services to outcomes.

To implement each of the evaluation options, stakeholders will need to make choices reflecting the following design issues:

- **Incorporating the needs of people of color and underserved communities.** Demonstrations must consider how to include youth and families of color and from other populations that might traditionally be underserved (more broadly) or who might be less likely to enroll in a research project (more specifically). Practices might include partnering with community organizations on designing the intervention and its evaluation to ensure that activities are culturally sensitive and appropriate, conducting recruitment to encourage responses from specific groups, and overseeing and delivering services. It may be important to oversample members of these populations to allow for a large enough sample to conduct subgroup analyses, which would enable an assessment of intervention effectiveness across different youth populations. In addition, the demonstration might hire staff who are demographically representative of the populations they intend to serve.

- **Accessing data.** Accessing state and federal databases can provide crucial information on outcomes, but such access requires obtaining proper agreements. Evaluators should allow for time to develop those agreements and be prepared to pursue alternatives if they cannot obtain access. This issue might be further complicated for demonstrations that run across multiple states, as each will require separate agreements.
- **Assessing intermediate outcomes for long demonstration and evaluation timelines.** Because the interventions target youth with the goal of affecting their outcomes as adults, the timelines proposed are sometimes quite long—up to 14 years. These timelines arise directly from the proposals, some of which specify long-term supports throughout the transition period into young adulthood. An accurate evaluation of these proposals would need to assess services and outcomes in line with the proposal's logic and service models. Nonetheless, assessing intermediate outcomes during the early part of some demonstrations—such as in the first year or two—could be beneficial in tracking whether the intervention is achieving its short- and medium-term outputs and outcomes. Though observing these early measures would be no guarantee of an intervention's later success, failure to do so would indicate a low probability of the intervention achieving its ultimate outcomes for employment and SSI.

Finally, all of the evaluation options present two common issues related to the demonstration sample:

- **Implementing RCT designs.** Randomizing at the individual level typically incurs challenges with enrollment, as youth and families might be hesitant to sign up for a new program when they know that there is a chance that they might not get in, even when assignment to the control group represents services they already access. Staff, too, can struggle with the enrollment process because they might believe that all interested youth should be a part of the intervention.
- **Achieving adequate sample sizes.** The sample sizes needed for a demonstration to detect meaningful impacts might be large relative to an organization's capacity to recruit participants and administer the intervention. Large sample sizes, though, as discussed in Section I.A.3, are necessary to detect impacts that are reasonable to expect. While smaller samples might be easier to manage and cheaper to implement, they also require larger impacts to detect an effect, which might be unrealistic for some interventions and outcomes.

2. High-level summary of the evaluation options for the proposals

Exhibit I.3 summarizes, at a high level, key evaluation features for each of the proposals, including the design, the organization that would lead the demonstration, the sample to be included and how the lead organization would identify the sample, the duration of the demonstration (including start-up and evaluation activities), and the primary goal for the evaluation.

Exhibit I.3. Evaluation features of the SSI Youth Solutions proposals

Proposal (Short and long title)	Evaluation design	Lead organization	Sample and recruitment identification	Demonstration duration	Primary evaluation goals
Apprenticeship Infrastructure Building an Apprenticeship Infrastructure for Youth Receiving SSI	RCT	An established apprenticeship intermediary organization in a state with an existing youth apprenticeship infrastructure	Youth with disabilities up to age 26 in a geographic area; outreach to youth participating in special education and those registered with their postsecondary education institution as having disabilities, along with those connected to partner agencies	6 years (1 year for start-up, 2 years for enrollment, an additional 2 years for services, and 1 year to conclude evaluation activities)	Gauge the appropriate scope (in size, funding, and duration) of the apprenticeship grants and measure the impact of apprenticeship on youth participants' employment and earnings outcomes
TEST-CTE Career and Technical Education for Students with Emotional Disturbance	Clustered random assignment	TransitionsARC	Youth using special education services in high schools	6 years (1 year for start-up, 4 years to stagger training across schools, and 1 year to conclude evaluation activities)	Document the effectiveness of TEST-CTE for increasing CTE participation and education, training, and employment outcomes
Delaying Substantial Gainful Activity Delaying Application of SSI's Substantial Gainful Activity Criterion from Age 18 to 22	RCT with two treatment arms (one with modified SSI criteria only and one that also includes information on supports)	Social Security Administration	Youth receiving SSI who are age 17 and approaching the age-18 redetermination in a defined geographic area and time period; recruitment is unnecessary	8 years (1 year for start-up, 6 years to follow up until participants are age 23, and 1 year to conclude evaluation activities)	Identify implementation challenges, assess whether outcomes for youth with disabilities are consistent with the proposal's logic model, and examine the effects of the policy on SSI program administration and costs
ResPECT Demonstrating the Effectiveness of Short-Term Career and Technical Training in a Residential Setting for Transition-Age Youth with Disabilities	RCT	Either a VR agency or a postsecondary education institution; whichever organization was not the lead organization would need to be a partner	Youth receiving SSI who are in or out of secondary school and young adults receiving or applying for SSI; identified through outreach to current and former clients of the state VR agency	6 years (1 year for start-up, 2 years for enrollment, an additional 2 years for services and follow-up, and 1 year to conclude evaluation activities)	Assess the benefits of combining ResPECT's two main components into a single intervention and the impacts of the intervention on outcomes

Proposal (Short and long title)	Evaluation design	Lead organization	Sample and recruitment identification	Demonstration duration	Primary evaluation goals
Employment Empowerment Employment Empowerment: A Foundational Intervention for Youth with Disabilities to Build Competitive Employment Skills	RCT, plus a pilot test to adapt the curriculum to non-college populations	One or more 2- or 4-year colleges or universities or a state department of education	Youth or young adults with disabilities attending postsecondary education institutions in their final year of studies	4 years (1 year for start-up, 1 year to offer the course, 1 year for follow-up, and 1 year to conclude evaluation activities)	Establish whether the employment empowerment curriculum improves employment attitudes and outcomes; adapt the curriculum for use with youth in non-college settings
FEAT Family Employment Awareness Training (FEAT): A Research-Based Program for Promoting High Expectations for Employment and Knowledge of Resources	RCT	A parent advocacy organization, such as the state parent training and information center, or a state developmental disability council	Youth ages 14 to 22 who use VR services; identified through outreach to current clients of the state VR agency	5 years (1 year for start-up, 2 years for enrollment and training, an additional 1 year for follow-up, and 1 year to conclude evaluation activities)	Assess the impacts of FEAT on youth's employment and reliance on SSI and other benefits
Family Empowerment Model The Family Empowerment Model: Improving Employment for Youth Receiving Supplemental Security Income	RCT	An organization that serves a broad disability community, already operates as a transition service partner, and can hire staff for FES positions, such as centers for independent living	Youth receiving SSI and their families beginning when youth are ages 14 to 16 in a single geographic location; targeted outreach to youth receiving SSI ages 14 to 16	8 years (1 year for start-up, 2 years for enrollment, an additional 4 years for services and follow-up, and 1 year to conclude evaluation activities)	Test the model's feasibility (including an assessment of challenges in recruitment, implementation, and service delivery) and effectiveness on employment and other outcomes
Integrated Treatment Team Improving Youth SSI Recipients' Employment Outcomes through an Integrated Treatment Team Intervention in a Health Care Setting	Pilot test	A multidisciplinary health clinic with a sizeable pediatric clinic	Youth receiving SSI ages 14 to 18 who use the multidisciplinary health clinic with integrated treatment teams; all eligible youth would be invited to participate in the pilot	3 years (1 year to develop the intervention, assemble partners, and enroll youth; and 2 years to deliver services and conduct evaluation activities)	Understand how an employment-focused case manager model could be implemented and adapted within a health care system that relies on an integrated resource team to offer services to youth and their families

Proposal (Short and long title)	Evaluation design	Lead organization	Sample and recruitment identification	Demonstration duration	Primary evaluation goals
YFSN Policy Considerations for Implementing Youth and Family Case Management Strategies Across Systems	RCT	Nonprofit organization with capacity to support YFSN staff	Youth receiving SSI age 14; identified through state Medicaid data	14 years (1 year for start-up, 2 years for enrollment, up to 10 additional years for services and follow-up, and 1 year to conclude evaluation activities)	Conduct a test of youth accessing YFSN services until age 24 to assess implementation challenges in connecting with youth and families over such a long period and whether the intended outcomes materialize
Progressive Education Progressive Education: Early Intervention Strategy to Improve Postsecondary Outcomes for Youth with Disabilities	Clustered random assignment	State VR agency	Youth ages 14 to 25 within state VR agency offices; half of the offices randomly assigned to offer Progressive Education, with the remaining offices assigned to offer services as usual	5 years (2 years for treatment group offices to offer services, 2 additional years for service provision and follow-up, and 1 year to conclude evaluation activities)	Evaluate whether the progressive education model can lead to improved postsecondary education and other outcomes
Transition Tracker Transition Linkage Tool: A System Approach to Enhance Post-School Employment Outcomes	Quasi-experimental difference-in-differences evaluation	State agency to coordinate efforts of multiple state agencies and school districts	All high school students with individualized education programs in school districts	6 years (1 year for start-up, 4 years for implementation and follow-up, and 1 year to conclude evaluation activities)	Quantify the benefits of an integrated data tool to transition service coordination and provision
TESS Transition to Economic Self-Sufficiency (TESS) Scholarships for Youth and Young Adults with Significant Disabilities	Pilot test using an RCT	State VR agency	Youth ages 18 to 24 who have exited high school; either receive SSI or are likely to meet the disability, income, and asset requirements for SSI; and have a desire to pursue a career	12 years (1 year for start-up, 1 year for recruitment, 9 years for services, and 1 year to conclude evaluation activities)	Follow a small group of scholars throughout their use of their scholarships (that is, through age 30) to track its employment and program participation outcomes

This page has been left blank for double-sided copying.

II. Building an Apprenticeship Infrastructure for Youth Receiving SSI

A. Proposal and demonstration description

Proposal description. The proposal on youth apprenticeship infrastructure calls for DOL to establish a grant program for apprenticeship intermediary organizations, which would then be responsible for creating apprenticeship programs for youth with disabilities (Kuehn 2021). Because youth apprenticeship programs are relatively rare in the United States and those that exist do not currently focus on serving youth with disabilities, one goal of these grants will be to explore the best models and approaches when recruiting and supporting participants and finding employers or other training providers who will make reliable, sustainable partners. The ultimate goal of apprenticeship for youth with disabilities is to provide these youth with better transitions from school to work through work-based learning opportunities, strong links to employers, and a range of coordinated supportive services.

Demonstration description. Apprenticeship programs rely on a network of partners acting in concert, with the intermediary organization overseeing and coordinating the activities of apprentices and other partner organizations. The ideal grantee would be part of an existing youth apprenticeship or other work-based learning program to avoid the need to build relationships with schools, employers, and training providers from scratch. A grantee could innovate services for the pilot by creating apprenticeship programs explicitly designed for youth with disabilities or by adding new programming or supports focused on accommodating youth with disabilities alongside current participants in an existing apprenticeship infrastructure. The new components would build on existing arrangements and relationships with employers, schools, and training providers to shorten the lead time before youth with disabilities begin training.

After a start-up period to allow the grantee to develop new services and arrangements with apprenticeship partners and make other preparations, the intermediary would invite youth with disabilities to apply for their apprenticeship programs. Researchers would randomly assign some applicants to enter the program and other applicants to a control group that was ineligible for the apprenticeship but could access existing services, including apprenticeships through other avenues. Because access to state administrative data would significantly simplify and improve the accuracy of impact evaluation, the grantee's recruitment and job placement activities should be concentrated in a single state.

Lead and partner organizations. The ideal lead organization for this demonstration would be an established apprenticeship intermediary organization in a state with an existing youth apprenticeship infrastructure, such as Colorado, Georgia, Iowa, South Carolina, or Wisconsin. Alternatively, state and local education authorities, community colleges, or career and technical education providers with established work-based learning programs and strong links to the public education system might also be well-positioned to create youth apprenticeship programs.

For youth apprenticeship programs, a critical set of partners includes secondary and postsecondary education institutions. In addition to being a source of referrals, these institutions can facilitate the transition from school to the workforce and serve as a center of support for youth with disabilities. Other key partners would include employers, labor organizations, third-party training providers, and state government agencies such as the VR agency.

B. Evaluation rationale and overview

Evaluation rationale. Apprenticeships for youth with disabilities have a strong correlational evidence base, and although some similar programs have causal evidence in support of their effectiveness, no direct causal evidence exists for the effectiveness of apprenticeships in raising employment or earnings of youth with disabilities in the United States. If a demonstration of the youth apprenticeship infrastructure could employ experimental variation in apprenticeship participation, it could build a causal evidence base for apprenticeship programs.

Apprenticeship programs of any kind are relatively rare in the United States, and more so for programs focused on youth, with or without disabilities. This study would provide foundational information about the potential scope (in size, funding, and duration) and impacts of an apprenticeship program for youth with disabilities. Accordingly, the results of this evaluation would be of interest to federal, state, and local policymakers—including workforce investment boards tasked with allocating money to training programs, work-based learning coordinators and transition professionals in high schools and VR agencies, and training providers such as labor groups and employers.

Evaluation design overview. The evaluation of the youth apprenticeship infrastructure relies on an RCT that invites youth with disabilities up to age 26 in the pilot catchment area to apply for apprenticeship. (The minimum age for apprentices would be determined by local laws covering employment for minors —likely around age 16.) A subset of applicants would be selected randomly for the treatment group and permitted to enroll as apprentices. Control group members would access usual services but have no access to the pilot program. Some control group members might obtain apprenticeship slots or other training from alternative sources, such as the VR agency or workforce center. Youth who apply would be followed for up to two years to track employment and earnings outcomes through the state's UI database, as well as through participant surveys.

An evaluation based on an RCT design would generate credible, causal estimates of the program's impacts on outcomes. The evaluation would also yield valuable information about the processes, participation patterns, and costs of implementing apprenticeship programs for youth with disabilities. Although the final impacts will require some time to manifest and estimate, early findings based on the initial stages of the pilot could suggest whether support for additional apprenticeship programs for youth with disabilities is warranted and inform the creation of new programs.

C. Evaluation design

Sample and recruitment. The pilot demonstration would enroll youth with disabilities up to age 26. To obtain a large enough sample size to detect meaningful impacts, the pilot would likely need to recruit apprentices and employers from a wide geographic area, highlighting the benefit of a state-wide program or one focused on a large city or metropolitan area. The demonstration would recruit youth participating in special education or those registered with their postsecondary education institution as having disabilities. About 14 percent of public school students and 19 percent of college students identify as having a disability (U.S. Department of Education 2021a, 2021b). To reach additional youth with disabilities not enrolled in education programs, the demonstration could recruit from lists of youth with disabilities derived from partner agencies, such as the state Medicaid agency (to identify youth receiving SSI), the VR agency (to identify current or former youth clients), or agencies such as intellectual and developmental disabilities agencies or centers for independent living (to identify other youth with disabilities).

Design. The proposed evaluation design would involve a six-year RCT that tracks apprenticeship experiences and outcomes for up to two years. The evaluation would randomize youth into the apprenticeship pilot after they express an interest in the demonstration. After an initial year of demonstration development, the pilot would enroll youth over a two-year period and provide training and support for two years after enrollment. (This assumes an apprenticeship program length of up to two years; depending on the apprenticeship programs offered by the intermediary, a shorter or longer evaluation period might be appropriate.) The process analysis would document facilitators and challenges to developing the model, along with details of the apprenticeships that youth pursue and the supports youth need to be successful. The impact analysis would document impacts on services, program participation, training, and employment in the two years following a youth's enrollment into the demonstration. In Year 6, evaluation activities would conclude with the release of final reports.

Research questions. The evaluation would address the questions in Exhibit II.1 using the data sources and analytic methods listed for each analysis type.

Data collection. The evaluation would use data from the following sources**.**

- **Baseline application data** collected through apprenticeship applications before random assignment would provide information on youth's characteristics, such as demographics, disability information, SSI recipiency status, geography, desired training program, and attitudes and expectations about employment.
- **Pilot program data** would provide information on each partner and their roles in carrying out the apprenticeship program. They would also include youth-level process data on programs and outcomes, as well as apprenticeship costs. These data would also be used to administer the intervention, assigning youth to appropriate programs for their geographic area, expressed interests, and needs. Information about the mix of available apprenticeship programs—that is, the specific industries, occupations, credentials, and career paths for which youth would be able to enroll—would provide important context for evaluating youths' employment outcomes.
- **State administrative data** on youth with disabilities gathered from secondary and postsecondary education institutions and other partner agencies would be used for recruitment and subsequently to measure participation in the apprenticeship program. The evaluation would benefit from outcome data provided by secondary and postsecondary education systems (for educational outcomes) and the unemployment insurance agency (to measure employment and earnings).
- **Staff interviews** with staff from apprenticeship intermediaries and their partners would supplement program data for the implementation and benefit-cost analyses and add qualitative perspectives in interpreting those findings.
- **Participant focus groups and structured interviews** would provide insight into participants' experiences with the program and inform the process analysis.
- **Participant survey data**, collected 24 months after enrollment, would measure outcomes unavailable through the sources above, such as enrollment in alternative training programs, experiences with apprenticeship and other programs, sources of income not covered by the unemployment insurance system, and employment characteristics.

Exhibit II.1. Evaluation questions, data sources, and analytic methods for an evaluation of the Apprenticeship Infrastructure proposal

Questions	Data sources	Analytic methods
Process analysis		
• Who participated in the program, and how did they compare with the recruitment pool? • How did participants compare to youth receiving SSI? • Which partners participated in the apprenticeship infrastructure pilot? • How did the intermediary recruit youth? How were youth and families of color or from underserved communities recruited? • What were the processes used for apprenticeship arrangements? What types of apprenticeships did youth use? Did they complete them? What supports did youth need? • What roles and responsibilities did each partner have on the apprenticeship arrangements? • What prompted employers or training providers to participate, and is there potential for a sustainable set of apprenticeship opportunities? • Was the pilot implemented as intended?	• Baseline application data • Participant focus groups or structured interviews • Pilot program data • Staff interviews • State administrative data	• Descriptive statistics • Qualitative analyses
Impact analysis		
• What was the impact of the pilot on intermediate outcomes, such as secondary and postsecondary education enrollment and attainment, self-sufficiency, and public benefit receipt? • What was the impact of the pilot on the ultimate outcomes of employment and SSI benefit receipt? • Was the program more effective with some youth and families than others?	• Participant survey data • Pilot program data • State administrative data	• Regression analyses
Benefit-cost analysis		
• What were the pilot costs? How would the costs change with the scale of the operation? • Are the benefits of the apprenticeship pilot large enough to justify its cost? • How did benefits and cost differ by stakeholder perspective?	• Participant survey data • Pilot program data • State administrative data	• Descriptive statistics • Regression analyses

SSI = Supplemental Security Income.

Analysis and reporting. The evaluation of the apprenticeship infrastructure pilot would produce early process evaluation reports to document youth participation in training, process outcomes, and program operations, and an impact evaluation report that documents 24-month impacts at the conclusion of the demonstration. Process analysis and reporting can begin as soon as youth enroll and begin their apprenticeship programs, with interim reports on subsequent participation, progress for existing cohorts, implementation successes and challenges, and cost presented periodically throughout the demonstration. One goal of these interim reports would be to identify the needs of future youth apprenticeship intermediaries, such as how many youth they can expect to recruit, realistic budgets and timelines, and key challenges in the start-up process. Another valuable aspect of these reports might be to compare

youth interests (as stated in the baseline application) with subsequent apprenticeship program enrollment, which could reflect appropriate matches. The 24-month impact analysis would be conducted in Year 6, after most youth complete their training programs and have time to secure post-training employment. This period might be too short for youth who enroll in later years of the demonstration to complete their apprenticeships, but it should provide enough time to note differences in early outcomes between treatment and control group members. This follow-up period might also be too short if the apprenticeship programs created under the grant are substantially longer than two years or if the intervention requires a longer period for employment impacts. In particular, employment outcomes at a 24-month period might reflect services due to the apprenticeship model rather than impacts that result from the model. In these cases, the evaluation duration may need to be lengthened and could rely on state administrative data.

D. Proposal refinements and other considerations

Because there is no established model for apprenticeship programs for youth with disabilities, this proposal calls for a series of multiple intermediary grants to explore many potentially promising ideas. This evaluation, as described, focuses on a single grantee for simplicity, but the recommendations around evaluation design apply equally well to a program with multiple grantee intermediaries—with each grantee creating its own apprenticeship program or programs and forming its own treatment and comparison groups.

Enrolling 400 youth into the demonstration over a two-year period, half of whom would be in the treatment group, might prove challenging for two reasons. First, creating 200 apprenticeships might create capacity issues in the overall apprenticeship pipeline. Second, a geographic area might not contain enough youth who are interested in the program. To address these issues, the demonstration could expand to other states, extend the enrollment period (thereby lengthening the demonstration period), or reduce the number of youth for the demonstration. Those overseeing the demonstration would need to weigh the benefits and limitations that come with each option.

This page has been left blank for double-sided copying.

III. Career and Technical Education for Students with Emotional Disturbance

A. Proposal and demonstration description

Proposal description. The Translating Evidence to Support Transitions in Career and Technical Education program (TEST-CTE) is a piloted intervention developed for career-technical education instructors, special educators, and transition personnel (McKay and Ellison 2021). Its focus is for those who facilitate the transition planning components of individualized education programs (IEPs) for students with emotional disturbance (ED). TEST-CTE provides staff with a half-day in-person training before the school year begins, monthly coaching sessions across one semester, and a guidebook on recruiting students with ED into career and technical education (CTE) coursework, making career plans (such as by administering assessments and setting IEP goals), and accommodating students in CTE courses. Although developed for use with students with ED, initial testing has yielded evidence that TEST-CTE can be used with students who have autism, neurological or sensory disabilities, or other health impairments.

The ultimate goals of TEST-CTE are for students to complete accredited vocational training or obtain industry-recognized credentials (including but not limited to associate's or bachelor's degrees) and become self-sufficient in their career choice. To achieve these goals, the intervention promotes students' involvement in work-based learning experiences and their completion of four CTE credits on a given career path during high school.[2] These activities, in turn, result in an increased number of students who apply to or enroll in postsecondary education or training or obtain a job within six months of high school graduation. Intermediate goals to ensure these outcomes include students' having articulated career interests and CTE participation plans in their IEP, improved social skills, self-efficacy, career expectations, and reduced stigma around vocational training.

TEST-CTE was developed at the Transitions to Adulthood Research Center (TransitionsARC) at the University of Massachusetts, which provides technical assistance to schools and districts implementing the intervention.

Demonstration description. The demonstration would assess TEST-CTE's effectiveness by recruiting a group of secondary schools to implement the training and conduct related activities. The six-year demonstration would involve a year-long set-up period to recruit schools, a four-year implementation period, and a year of follow-up data collection, analysis, and reporting.

Lead and partner organizations. The key organizations for this demonstration would include the TransitionsARC, which would coordinate provision of training and ongoing technical assistance, and the schools that sign up to implement TEST-CTE. Other partners might include state departments of education and labor, which could grant researchers access to administrative data for recruiting schools and tracking student outcomes.

[2] According to the U.S. Department of Education's Career and Technical Education Statistics (2021), one credit is awarded for passing a course that meets for one period a day for an entire academic year, or the equivalent instructional time.

B. Evaluation rationale and overview

Evaluation rationale. CTE has shown promising correlational results for special education students with ED, but there is no causal evidence of its impacts. No experimental evidence demonstrates employment and earnings impacts of CTE courses taken as part of a standard high school curriculum; however, conceptually-related programs, such as Job Corps, and more intensive vocational secondary programs, such as career academies, have experimental or other causal evidence in their support. If TEST-CTE induced experimental variation for students' enrollment in CTE or increased the number of credits they completed, it could enable stakeholders to observe the effectiveness of CTE for increasing education, training, and employment outcomes—and thereby fill this gap in the evidence base. The federal government invests significant resources in supporting CTE, and federal requirements call for state and local education agencies to improve access for students with disabilities. A demonstration such as this would also be desirable to support these policies.

Evaluation design overview. Because TEST-CTE focuses on staff training, it would require a clustered randomized design, with treatment/comparison group formation at the school or school district level.[3] Randomizing access to or support for CTE at the student or classroom level within schools could violate expectations of fairness across students and could present administrative and practical challenges, such as if students sought to switch between treated and untreated class periods, or instructors were asked to take different approaches with students in the same class. The specificity of the disabilities for which TEST-CTE was designed and the narrowness of some key outcomes of interest (such as completing four courses in CTE) could increase the sample size of schools or districts needed to obtain a large enough number of schools and students to detect meaningful impacts. Participating schools or districts would ideally be paired, based on similarity in important baseline characteristics such as school size, enrollment of students with disabilities, availability of CTE options, and local economic conditions, and randomization would occur within the pairs.

One possible way of facilitating this design would be to enlist a state department of education to identify and recruit schools throughout a single state to participate in an evaluation. State agencies must report aggregate data on special education enrollment in their schools, so they will have data to facilitate recruitment and matching of schools for randomization purposes. State agencies also control other data needed to evaluate outcomes, such as employment and earnings records and information about the use of public benefits. Conducting the research with schools recruited in a single state would therefore likely reduce the cost and complexity of the recruitment, outcome follow-up, and data analysis, but it could also raise concerns about generalizability to settings outside that state.

C. Evaluation design

Sample and recruitment. The developers designed TEST-CTE for students in special education with ED. However, they also identify TEST-CTE as appropriate for students with other disabilities, such as autism, learning disabilities, neurological and sensory disabilities, other health impairments, or the broader category of serious mental health conditions. Because of the difficulty in clearly identifying students with ED within the general student population, an intervention would likely include students

[3] The appropriate level of randomization would depend on whether CTE was administered separately in each school, or for many schools jointly in area CTE schools, with locations that provided CTE from the same source randomized as a single unit. While it would be valuable to examine the impact of TEST-CTE on regional centers such as this, as well as other specialized settings such as dedicated CTE schools (or career academies), this proposal focuses on standard high schools for simplicity.

with ED in special education. This approach could limit the number of students per school in the population of interest. One way to overcome this problem would be to expand the population of interest to include students with a broader range of related disabilities, whose participation is envisioned by the developers, and conduct subgroup analyses for students with ED or other conditions (sample sizes permitting). Although SSI participation would be a characteristic of interest in sample recruitment and as a long-term outcome, requiring SSI participation for entry into the study might be challenging because schools lack information on students' SSI status.

Schools implementing TEST-CTE would identify students by their participation in special education transition services, an IEP, or individualized learning plan. This method of sample identification could also serve as the delivery method for planning-related aspects of TEST-CTE. In the 2019–2020 academic year, more than 1.2 million students ages 12 to 21 had ED, autism, or another disability with which TEST-CTE can be applied as their primary disability diagnosis (U.S. Department of Education 2019). Also, the broad disability diagnostic group of mental disorders represents roughly four in five of all youth ages 13 to 17 receiving SSI (Social Security Administration 2020).

Design. An intervention implemented in multiple schools across a state with the assistance of state-level agencies would lend itself to implementation of the program across schools over a period of four years, with some schools being assigned at random to undergo TEST-CTE training. In some locations, multiple schools may send students to a single CTE training center, in which case all of the sending schools would be assigned in a block as a single unit.

The design would create a contrast between staff and students in adopting schools (the treatment group) versus those in schools that did not adopt it (the control group). It would provide an opportunity to conduct process and benefit-cost evaluations in adopting schools.

Because student outcomes would develop over several years, the timing of the impacts would affect the way the treatment group could be specified and change the relevant comparisons. For instance, the most straightforward comparison might be to examine students who were in grade 9 in the first year of implementation. However, their outcomes would only be observed up through and directly after high school graduation. Students in grade 12 in the first year of implementation would have a shorter exposure to the intervention (one year), but a longer observation period for impacts after high school. The timing of course and IEP planning is important to consider when developing the analyses, and implementation and evaluation of the intervention would benefit from understanding the level of exposure needed to observe measurable impacts.

Research questions. The evaluation would address the questions in Exhibit III.1 about the TEST-CTE proposal using the data sources and analytic methods listed for each analysis type.

Exhibit III.1. Evaluation questions, data sources, and analytic methods for an evaluation of the Translating Evidence to Support Transitions in Career and Technical Education proposal

Questions	Data sources	Analytic methods
Process analysis		
• How many students with disabilities participated in and completed CTE, and what were their characteristics? How were youth of color and from underserved communities included? • How did participants compare to the broader special education population in the school districts, in the state, and nationally? • How was the TEST-CTE program designed, implemented, and operated, and what factors contributed to the implementation experience? • Was the TEST-CTE program implemented as intended? How did implementation vary across schools? • What CTE courses and programs did students use? • What were student, educator, administrator, and parent attitudes toward CTE? • Did IEPs developed with TEST-CTE reflect the TEST-CTE model, such as having SMART (specific, measurable, achievable, relevant, timely) career goals? • Did CTE instructors with TEST-CTE training increase their accommodations for students with disabilities?	• Document review • School administrative records • Staff, student, and family structured interviews and focus groups • Student surveys • TEST-CTE program records	• Descriptive statistics • Qualitative analyses
Impact analysis		
• What was the impact of TEST-CTE on intermediate outcomes, such as career interests, CTE participation, social skills, self-efficacy, career expectations, postsecondary education and training, and credential attainment? • What was the impact of the policy change on the ultimate outcomes of employment and SSI benefit receipt? • Was TEST-CTE more effective with some youth than others?	• School administrative records • State unemployment insurance data • Student surveys	• Regression analyses
Benefit-cost analysis		
• Were TEST-CTE's benefits large enough to justify its cost? • How did benefits and cost differ by stakeholder perspectives (such as students, schools, and local organizations)?	• Document review • School administrative records • State administrative records • Student surveys • TEST-CTE program records	• Descriptive statistics • Regression analyses

CTE = Career and Technical Education; IEP = individualized education program; SSI = Supplemental Security Income; TEST-CTE = Translating Evidence to Support Transitions in Career and Technical Education.

Data collection. The evaluation would use data from the following sources.

- **TEST-CTE program records** (such as training logs, classroom observations and evaluations, and notes from coaching sessions) would reveal staff's receptiveness to the training, the challenges they encountered, the fidelity with which they applied the intervention, CTE course enrollment, and

enrollment in postsecondary education and training programs. These records would also inform the benefit-cost analysis by pinpointing drivers of program cost.

- **School administrative records** would be used to determine what kind of students with disabilities enrolled in CTE courses and how they compared to the student body at large. These records could also track progress outcomes, such as enrollment in and progress through CTE and other courses and ultimately high school completion.
- **Staff, student, and family structured interviews and focus groups** would measure how TEST-CTE affected attitudes towards CTE and its suitability for youth with disabilities, as well as concepts such as self-efficacy and career expectations. They could also supplement training records in showing how staff implemented the intervention.
- **Document review** covering student IEPs would measure whether they include plans to enroll in CTE and reflect TEST-CTE training, such as having specific, measurable, achievable, relevant, and timely goals (known as SMART goals). This task could be automated, with text-analysis algorithms programmed to recognize phrases related to career plans and CTE enrollment, with a subset of IEPs checked by human reviewers to confirm the accuracy of the algorithmic approach.
- **State administrative data**, such as wage records from the state UI database, would enable measurement of employment and earnings outcomes. Economic self-sufficiency can be measured using administrative data on public benefit receipt from federal or state sources (such as SSA, the Supplemental Nutrition Assistance Program [SNAP], or Temporary Assistance for Needy Families [TANF]), though these data might be more informative after youth leave high school. Data from state postsecondary education institutions would be used to measure whether students enrolled in local public colleges or universities. An advantage of these data sources is that a demonstration could use them both to observe short-term outcomes as part of the primary evaluation activities (during or right after high school graduation), and a long-term follow-up analysis could track these same outcomes into young adulthood.
- **Student surveys** at baseline and follow-up could, evaluation resources permitting, supplement student-level data to identify outcomes and measures not available in administrative data. Surveys could capture economic characteristics and outcomes such as SSI receipt, participation in training programs that do not report student enrollment to NSC, and employment activity not captured in state UI records. Surveys could also collect information on subjective intermediate outcomes, such as self-efficacy, CTE stigma, and career expectations, along with any demographic or contextual data not gathered by administrative records (such as parents' education or disability status).

Analysis and reporting. The process and cost analyses can begin as early as two years into the study, after the end of the first year of implementation. These analyses can be updated with annual reports to identify changes over time in outcomes of interest. Likewise, impact analyses for short- or intermediate-term outcomes—such as CTE inclusion in IEPs, reduction in CTE stigma, and enrollment in and progress through CTE coursework—can begin after the first year and examine impacts on all students with disabilities enrolled in the study schools. However, because of the important role that IEP preparation, course sequencing, and CTE credit accumulation play in the theory of change for this proposal, impacts on employment and other postsecondary outcomes may be muted for older cohorts in their second, third, or fourth years of high school, even when looking at short-term outcomes.

Final impacts on enrollment in postsecondary education and training programs, employment, and SSI will very likely be most evident when comparing students who spend all four years in a high school

implementing TEST-CTE to a similar group of students who spent no time in a TEST-CTE high school. This approach requires at least five years to elapse between initial implementation and the production of a final impact report examining postsecondary enrollment, employment, and earnings outcomes—four years for the youngest group of students in study schools to complete high school, plus one year in which to follow up on their outcomes, analyze the data, and prepare the report.

D. Proposal refinements and other considerations

One of the strengths of TEST-CTE is that it makes use of the widely-available CTE infrastructure already integrated within school systems. However, implementers could consider the variety of available CTE pathways (that is, different industries, credentials, and career paths, and delivery of CTE separate from academic coursework through an area CTE school versus at the same high school) as a characteristic of the intervention schools and of the students who opt to pursue them. To the extent possible, both the CTE pathways and the prevailing local economic conditions during and immediately after the intervention should be considered when matching schools or districts for comparison.

An evaluation such as the one described above could be fielded in a large urban school district or with a group of individually recruited schools, although sample size could quickly become an issue and the schools' total enrollment of eligible students would be a consideration. It could also be fielded by recruiting schools from multiple states, which could add geographical balance and improve the generalizability of the evaluation results, at the cost of a few complications. A multi-state recruitment strategy could draw on national databases for measuring postsecondary outcomes such as the National Directory of New Hires or National Student Clearinghouse. However, using these databases could complicate the data collection process, and researchers would need to ensure sufficient overlap or dispersion in the states and settings of schools selected for the treatment and control groups.

If only a small number of schools could be recruited or the random assignment of schools proved to be infeasible, a smaller, non-experimental version of the evaluation could be conducted, so long as it included an appropriate comparison group. For instance, if TEST-CTE were instituted in one school, administrative data on student progression and IEPs, teacher and student surveys, and post-graduation enrollment, employment, earnings, and benefit receipt data could be obtained from other schools with similar observable characteristics for comparison. Regardless of evaluation design, keeping the treatment and comparison schools within the same state would be important to simplify administrative data collection and reduce variation in state-level policies or characteristics that could influence outcomes.

IV. Delaying Application of SSI's Substantial Gainful Activity Eligibility Criterion from Age 18 to 22

A. Proposal and demonstration description

Proposal description. The Delaying Substantial Gainful Activity proposal by Larson and Geyer (2021) describes an approach to delay application of the adult disability criteria for new SSI and Social Security Disability Insurance (SSDI) applicants and continuing SSI recipients from age 18 until age 22. The change would extend the use of a modified child disability criteria for transition-age youth ages 18 to 21 and add the domains of self-direction, capacity for independent living, and economic self-sufficiency to the child disability criteria. The proposed change is intended to make the eligibility criteria for the disability program more consistent with evidence from the medical literature supporting continued maturation past age 18, as well as with federal legislation that established age 22 as the end of the developmental period for youth. The goal of the proposed change is to improve the transition and employment outcomes for low-income youth with disabilities by extending eligibility for cash assistance. This extension would allow youth to receive supports such as Medicaid and SNAP benefits and facilitate connections with essential support services such as VR during this critical developmental period. As a result, youth would better meet their basic needs, such as food security, housing, transportation, health care, long-term services and supports, education, and vocational and skills training. Those supports could encourage SSI recipients to develop work skills and experience, gain educational and vocational credentials, improve self-direction skills, and reach socioemotional maturity.

Demonstration description. Two-thirds of youth receiving SSI in a defined geographic area who undergo an age-18 redetermination would be subject to a redetermination that uses modified disability criteria for transition-age youth. Disability Determination Services (DDS) staff would assess all youth undergoing a redetermination with the usual criteria; for those in the treatment group, DDS staff would assess them based on both the usual and modified disability criteria (to identify which youth would and would not be allowed under both sets of criteria). Treatment group youth would receive SSI benefits based on the modified disability criteria. Upon turning age 22, if treatment group youth wish to remain on SSI, they would undergo a continuing disability review under SSA's adult disability criteria. This change will require revisions in the eligibility determination protocols and training for the state DDS staff. The youth in the treatment group would be randomly assigned into two treatment arms. In the first arm ("modified criteria plus"), sample members would receive information annually (through a mailing) about available supports and services, such as benefits counseling and contact information for the workforce center and the VR agency. The information would include resources that youth could potentially use to promote their transition. In the second arm ("modified criteria only"), sample members would not receive that information; the intervention would only be the modified criteria for the age-18 redetermination. Though the intervention proposal does not mention any activities to educate youth receiving SSI on additional supports, the demonstration could document whether a low-cost activity such as periodic mailings could improve the use of those supports and thus lead to better employment and other outcomes. To assess the impacts of the policy change, the evaluator would compare the outcomes of the treatment group members to those of an experimental control group (members of the target cohort within the geographic area randomly assigned to control status). The demonstration would occur in a geographic area large enough to provide a sample size sufficient to detect meaningful impacts on key outcomes. The site could be a single state or a large city to minimize administrative and implementation issues (such as training staff).

Lead and partner organizations. The lead organization responsible for the demonstration would be SSA, with support from the DDS and SSA field offices in the state selected for the demonstration.

B. Evaluation rationale and overview

Evaluation rationale. Because the proposal would require a change to federal law, SSA could use its SSI demonstration authority under Section 1110 of the Social Security Act to test the proposal. A program waiver is needed because the demonstration would change program rules for the age-18 redetermination and ongoing eligibility of individuals ages 18 to 21. A demonstration would be appropriate to identify implementation challenges, assess whether outcomes for youth with disabilities are consistent with the proposal's logic model, and examine the effects of the new policy on the administration of and costs to the SSI program.

Evaluation design overview. The proposal could be tested as a small-scale, RCT demonstration aimed at a cohort of youth approaching the age-18 redetermination and using a randomly selected group to assess outcomes, estimate costs, and identify potential challenges. The demonstration would assess critical factors and potential challenges to implementation related to the effect on SSI program administration, the administrative and SSI payment costs, and the outcomes of transition-age youth who experience a delay in the application of the adult criteria. To keep administration of the demonstration simple and low cost, the demonstration would not include new SSI or SSDI applicants between ages 18 and 21, though the proposal includes this group. The demonstration and its evaluation would operate for eight years to allow one year to develop procedures to operationalize administration of the new criteria and train staff, one year to identify the study cohort and implement the new criteria, five years to observe participants' outcomes one year after the treatment group becomes subject to the adult criteria, and a final year for analysis activities.

C. Evaluation design

Sample and recruitment. The demonstration would include all youth currently receiving SSI who are age 17 and approaching the age-18 redetermination in a defined geographic area and time period. Selection of the area and period would be based on having a control group and two treatment arms that are large enough to detect meaningful impacts on key outcomes. SSA would randomly assign youth into the three experimental groups; individual consent would not be necessary.

Design. Because no evidence exists about the implications and effectiveness of the proposal, a demonstration could provide evidence of its feasibility and potential. We propose an eight-year demonstration to allow one year for operational development, one year to apply the criteria, a five-year follow up through age 23, and one year to conduct final evaluation activities. Evaluators could follow this cohort for additional time to observe longer-term outcomes.

The evaluation would use an RCT to assess the impact of the intervention. SSA would randomly assign youth to a control group from among the same pool of youth from which the treatment group youth are selected.

Research questions. The evaluation would address the questions in Exhibit IV.1 using the data sources and analytic methods noted for each.

Exhibit IV.1. Evaluation questions, data sources, and analytic methods for an evaluation of the Delaying Substantial Gainful Activity proposal

Questions	Data sources	Analytic methods
Process analysis		
• How did the characteristics of participants in each treatment arm compare with the characteristics of the control group? • How did participants compare to all youth age 17 receiving SSI on their demographic characteristics? • How were DDS office staff trained on the transition-age youth eligibility criteria? Did DDS staff believe this training was enough for them to assess whether participants met the new criteria? • How were the new elements of the modified criteria implemented? Did DDS staff understand the role of these criteria in sufficient detail? • What were the facilitators and barriers to changing the eligibility criteria? • For youth who received an age-18 redetermination, what proportion met the modified criteria? What proportion would have met the adult criteria?	• SSA administrative data • Staff interviews	• Descriptive statistics • Qualitative analyses
Impact analysis		
• What was the impact of the policy change on intermediate outcomes, such as service access, basic needs, and work readiness? • What was the impact of the policy change on the ultimate outcomes of employment and SSI benefit receipt? • How did intermediate and ultimate impacts differ by treatment arm? • How did the policy change affect the proportion of disability redeterminations that youth appealed at age 18? • How did the policy change affect the proportion of continuing disability reviews that youth appealed at age 22? • How did the policy change affect the number of Section 301 determinations for those ages 18 to 22? • Was the policy change more effective with some youth than others?	• Other federal administrative data • Participant survey data • Program data • SSA administrative data	• Regression analyses
Benefit-cost analysis		
• What was the cost to change SSA's policies, and what benefits would offset those costs? How would the costs associated with a nationwide implementation of the policy change compare to the benefits?	• Other federal administrative data • Participant survey data • SSA administrative data	• Descriptive statistics • Regression analyses

DDS = Disability Determination Services; SSA = Social Security Administration; SSI = Supplemental Security Income.

Data collection. The evaluation would collect data from the following sources.

- **Staff interviews** would involve periodic discussions (in person or by phone) with SSA and DDS staff. These interviews would gather information about the training of DDS staff and experiences and challenges in the administration or implementation of the new criteria for determinations.

- **Program data** could enable DDS staff to document features and measures of their assessments using both the revised and adult criteria.
- **SSA administrative data** are necessary for all phases of the evaluation. In addition to identifying the cohort of youth who would experience the policy change and the control group, the data would enable a comparison of the characteristics (such as demographic, disability, and SSI benefit information) of youth assigned to each study arm. The data would track their experiences with SSA benefit receipt during the demonstration, including the outcomes of the age-18 redeterminations using the existing and revised criteria, appeals and new applications, and benefits awarded; work incentive use; and outcomes at age 23.
- **Participant surveys** would collect information on intermediate outcomes suggested by the proposal's theory of change: service access (such as through VR and Medicaid) and addressing basic needs (such as food security, housing, and transportation). While some measures of service access could be identified through federal administrative databases, other measures—particularly those involving basic needs and work readiness—can only be obtained by asking participants directly through a survey. The data collected through this activity could also document differences between the two treatment arms in these intermediate measures. The survey could be offered in the year before the age-22 redetermination, which would allow an assessment of participants' experiences throughout most of the demonstration period.
- **Other federal administrative data**. SSA links its data with other administrative data that could be used to track outcomes hypothesized to be affected by the policy change. These data include the Centers for Medicare & Medicaid Services' Medicaid records, RSA's VR records, and the Internal Revenue Service's annual earnings data.

Analyses and reporting. The demonstration evaluation would benefit from three sets of analyses and reports.

- A process evaluation one year after the final members of the cohort complete their age-18 redeterminations would show the equivalence of characteristics among the two treatment groups and the control group and among all participants and the broader SSI population and document the redetermination outcomes, early SSI benefit information, and qualitative findings from staff on their experiences with the new criteria and the associated training.
- An impact analysis up to one year after the age-22 redetermination would document the cohort's redetermination outcomes. Using SSA and other federal administrative data, along with participant survey data, it would also compare the intermediate and ultimate outcomes between the two treatment groups and the control group.
- A benefit-cost analysis would assess the cost to SSA of delaying the application of adult criteria and weigh that cost with the benefits to youth and SSA, based on the estimated impacts of the policy change.

D. Proposal refinements and other considerations

The evaluation design would enable SSA or other evaluators to track the long-term benefits and costs of the policy change. The effects of delaying the application of adult criteria might not occur until well beyond age 22. Thus, continued observations of earnings, SSA receipt, SSA work incentives, and Medicaid use 5 to 10 years after the end of the demonstration could provide insights into the long-term effects of the policy change if early impacts appear promising.

If a randomly selected control group is challenging or infeasible depending on how SSA implements the new policy, the evaluation could use a matched comparison group to assess the policy change's impact. The comparison group could be comprised of youth in the same state who undergo the age-18 redetermination at the same time as the treatment groups, but who reside outside of the demonstration's geographic area or who reside in the same area but underwent the redetermination in the year before the demonstration. These youth would be matched to be comparable to the treatment group across important demographic, geographic, and disability-related dimensions (including sex, race, language of preference, age at earliest SSI eligibility, impairment, living arrangements, other child and parent incomes, pre-age 18 employment, and SSI monthly payment amount). The evaluator would need to assess the options for identifying a quasi-experimental comparison group and decide which method presents the least risk of introducing confounding factors in the impact estimation.

This page has been left blank for double-sided copying.

V. Demonstrating the Effectiveness of Short-Term Career and Technical Training in a Residential Setting for Transition-Age Youth with Disabilities

A. Proposal and demonstration description

Proposal description. The ultimate goal of Residential Postsecondary Education and Career Training (ResPECT) is to launch youth receiving SSI into sustainable careers and reduce their reliance on SSI. It proposes to do so by providing youth with CTE at the postsecondary education level and self-efficacy gained through independent living and peer socialization (Hollenbeck 2021). ResPECT unifies elements from two programs into a single intervention. From the Michigan Career & Technical Institute (MCTI), ResPECT incorporates residential postsecondary CTE. From the Postsecondary Education Rehabilitation Transition (PERT) program pioneered by the Wilson Workforce and Rehabilitation Center, it draws on career interest and aptitude assessments. The major elements of ResPECT include (1) case management, (2) thorough assessment of skills and interests, (3) rigorous employer-driven career and technical curricula, (4) instructors who come from industry, (5) developmental academic coursework (as needed), (6) technological or in-person supports (as needed), and (7) independent living in a postsecondary residential setting during training. The PERT-like component lasts about two weeks, and participants would receive CTE through the MCTI-like component for at least one year. By combining the PERT-like and MCTI-like components into a unified intervention, ResPECT can offer participants a streamlined experience that is supplemented by case management. Youth who participate in the PERT-like component may enter the residential CTE program directly if indicated by their skills and interest assessment. The proposed intervention expands access to residential programs to increase the sustainable and satisfying lifetime careers and earnings of youth receiving SSI.

Demonstration description. The evaluation design presented in this chapter assumes a multi-site ResPECT demonstration as described in the proposal. The goal of the demonstration would be to assess the efficacy of the intervention by establishing the program in a select number of sites. In each site, a lead organization would establish a housing component and employ staff to monitor and supervise the student housing, conduct skills assessment, and provide CTE. The demonstration would be administered over several years, with the timeline depending on the capabilities and qualifications of the implementing institutions. If the participating organizations can build on an established initiative that includes counseling or residential components, the time needed to implement the intervention may be shorter. Each site would require outreach and recruitment mechanisms, along with training components for conducting assessments, presenting career and educational options, and providing CTE. Students who participate in the MCTI-like component would enroll in the training programs for at least one year, and then it may take several months or longer for youth to achieve labor market and other outcomes.

Lead and partner organizations. The lead organization at each demonstration site would ideally be a VR agency or a postsecondary education institution; whichever organization was not the lead organization would need to be a partner. The lead organization would administer the demonstration and monitor its implementation, but both organizations would develop the facilities and staffing necessary to provide the key elements of the intervention. If the postsecondary education institution lacks residential facilities, it could partner with another two- or four-year postsecondary education institution with such facilities. Postsecondary education institutions would host the demonstration, offer residential housing, and employ staff to administer it. VR agencies would refer customers and fund training, as appropriate.

B. Evaluation rationale and overview

Evaluation rationale. The PERT and MCTI components on which ResPECT is based have been established independently and evaluated previously (Dean et al. 2019; MCTI 2019), although the existing evidence for MCTI does not meet CLEAR's strong rating for causal evidence and has not been independently evaluated. These evaluations found that PERT increased long-term earnings and MCTI resulted in high placement rates and initial wages for youth with disabilities who completed high school. Combining the two interventions with the addition of case management constitutes a novel intervention for which existing evidence is limited. Thus, it warrants a rigorous evaluation. Housing participants in residential facilities while they attend the training, which is drawn from the MCTI program, is a key aspect of the intervention because it contributes to the goal of increasing participants' self-efficacy through independent living and peer socialization. The proposed evaluation would focus on the potential benefit of combining these programs into a single intervention and conducting a rigorous assessment of impacts on the outcomes of youth receiving SSI.

Evaluation design overview. The evaluation would use an RCT design at multiple sites to estimate impacts over a six-year demonstration period. The lead organization would randomize youth who enroll in the demonstration into a treatment group, which would use ResPECT services, starting first with the PERT-like component then moving to the MCTI-like component, or a control group, which would access usual services. The evaluation would follow participants for up to three years after they enroll to observe their short- and medium-term employment and SSI receipt outcomes. The state VR agency participating in the intervention could identify their current and former clients who meet the enrollment criteria for outreach and study recruitment.

C. Evaluation design

Sample and recruitment. The larger population intended by the proposal includes youth and young adults receiving SSI who are either in or out of secondary school. To identify the sample, the evaluation would conduct outreach to youth and young adults who are current or former clients of the state VR agency, based on youth characteristics. The size of the intervention group in each site would depend on (1) the number of suitable SSI recipients who are VR customers, (2) the capacity of the training programs, and (3) the number of sites participating in the demonstration.

Design. The evaluation would rely on an RCT design and follow participants for up to three years after enrollment. Depending on the participating agencies' experience with CTE programs and the availability of residential facilities, it could take one year or longer from the start of implementation until enrolling the first cohort of participants. Participants would enroll in the proposed intervention during a two-year period, and participants would start the PERT-like component at any point on a rolling basis and the MCTI-like component at the beginning of a semester or academic year. With a follow-up period of three years after youth enroll and one year to complete evaluation activities, the demonstration's evaluation would take at least six years, though the evaluation could produce interim findings as early as two years after the first youth enroll.

Research questions. The evaluation would address the questions in Exhibit V.1 using the data sources and analytic methods listed for each analysis type.

Exhibit V.1. Evaluation questions, data sources, and analytic methods for an evaluation of the Residential Postsecondary Education and Career Training proposal

Questions	Data sources	Analytic methods
Process analysis		
• What were the characteristics of youth who enrolled in ResPECT? How many were youth of color or from underserved communities? • How did participants compare to all youth receiving VR agency services or receiving SSI? • How many referred youth declined to participate in the program and for what reasons? • What funding did the participating agencies use for the program? • How was ResPECT designed, implemented, and operated and what factors contributed to the implementation experience? • Which specific CTE programs did participants attend? Which youth completed the entire program? Which completed only one of the components? Why did participants leave the program before completing it? • Which services beyond the skills assessment and training did participants use, such as case management, developmental academic coursework, technological supports, and job placement? • Did participants improve their skills through CTE? • To what extent did participants obtain employment in the field in which they were trained?	• Baseline application or survey data • Participant focus groups or structured interviews • Participant survey data • Program administrative data • Staff interviews • State or SSA administrative data	• Descriptive statistics • Qualitative analyses
Impact analysis		
• What was the impact of ResPECT on intermediate outcomes, such as education, living arrangement, and self-efficacy? • What was the impact of ResPECT on the ultimate outcomes of employment and SSI benefit receipt? • Was the program more effective with some participants than others?	• Participant survey data • SSA administrative data • State administrative data	• Regression analyses
Benefit-cost analysis		
• Do the benefits of ResPECT have the potential to offset its cost in the long term?	• Participant survey data • Program administrative data • SSA administrative data • State administrative data	• Descriptive statistics • Regression analyses

CTE = career and technical education; ResPECT = Residential Postsecondary Education and Career Training; SSA = Social Security Administration; SSI = Supplemental Security Income; VR = vocational rehabilitation.

Data collection. The evaluation would use data from the following sources.

- **Baseline data** collected through an application form would obtain information—at the time of enrollment—from youth interested in ResPECT.

- **Program administrative data** from the VR agency and postsecondary education institutions, such as management information system extracts, would enable the evaluation to track program implementation and CTE services. This information would inform the implementation, impact, and benefit-cost analyses.
- **Staff interviews** would include periodic interviews with ResPECT staff about the program and its implementation.
- **Participant focus groups or structured interviews** would provide an opportunity to solicit their opinions on the quality of the CTE and the residential component.
- **A participant survey** would collect information on key outcomes, such as employment, earnings, attainment of credentials, and benefit receipt, from ResPECT participants and their control group counterparts three years after enrollment. This information would enable the evaluation to gauge impacts and inform the benefit-cost analysis.
- **State administrative data** are necessary to track outcomes for ResPECT treatment and control group members. Data sources could include VR agency records (to track employment services and outcomes), unemployment insurance records (to observe quarterly earnings), and state education data (to track postsecondary education outcomes).
- **SSA administrative data** would provide accurate data on SSI-related characteristics and outcomes and annual earnings for all enrollees.

Analyses and reporting. The ResPECT evaluation would produce reports about findings that inform stakeholders interested in pursuing residential CTE at other sites. The evaluation could include the following analyses and reports.

- An early process assessment would document the fidelity of the intervention implementation in each site, after the lead organization establishes the model and while the first cohort of youth enroll in ResPECT. These descriptive analyses, which rely on program administrative data, project meetings, and staff interviews, would be conducted once during the demonstration.
- A final process evaluation would assess program outcomes, including CTE completion and skills attainment, for youth two years after enrolling in ResPECT. It would document who participated in the demonstration and include descriptive analyses of enrollment data and qualitative analyses of interview and focus group data.
- A final impact analysis would assess treatment and control group outcomes three years after enrollment using SSA administrative, state administrative, and participant survey data.
- A benefit-cost analysis would compare the impacts documented in the final impact analysis with cost data (from the program administrative data and other sources). This comparison would show whether the benefits of ResPECT outweigh its costs, considering different perspectives (such as the youth and the program) and what timeline might be needed for any benefits to outweigh the cost.

D. Proposal refinements and other considerations

A potential implementation issue for ResPECT is establishing new sites that do not already offer PERT or residential postsecondary CTE. The residential component, in particular, may be difficult to accomplish because of the many regulatory and procedural hurdles related to the housing needed for one- or two-week campus visits for youth under age 18 (for the PERT-like component) and multi-semester length residential housing for adults (for the MCTI-like component).

An expensive part of the intervention is the residential component, and so stakeholders may be interested in the efficacy of a nonresidential CTE program that is otherwise identical to ResPECT. The differential impact of the residential component could be estimated by randomly assigning some participants within a site to a nonresidential training program and comparing impacts with participants at the same site who were assigned to the residential component. In addition, the process analyses would consider differences in challenges implementing the intervention with and without the residential component. This change in the evaluation design would necessitate expanding the number of sites and participants included in the demonstration.

The evaluation design for ResPECT assumes that the demonstration includes multiple sites in a single state. The advantages are that implementation and data collection are simplified; a disadvantage is that a single state might not have enough sites or large enough sample to detect impacts. A demonstration could span multiple states; if so, then it would require a lead organization in each state, as well as multiple partnership agreements, memoranda of understanding, and data use agreements.

This page has been left blank for double-sided copying.

VI. Employment Empowerment: A Foundational Intervention for Youth with Disabilities to Build Competitive Employment Skills

A. Proposal and demonstration description

Proposal description. Employment empowerment is an orientation in services and job-skills education for youth with disabilities (Hippolitus 2021). It emphasizes a positive outlook on individuals' employment potential and training that focuses on (1) building self-confidence, aspirations, and self-efficacy; (2) imparting knowledge of spoken and unspoken job interview and workplace rules; and (3) building career-sustaining skills. This proposal consists of three parts—a call for broad cultural change to promote an employment empowerment mindset among disability policymakers and service providers, a federal cross-agency initiative to encourage that change, and an existing curriculum available for use by organizations interested in providing employment empowerment instruction.

The curriculum originated at the University of California-Berkeley, where it was designed as a semester-long course by the director of the Disabled Students Program. It has subsequently been used at other higher education institutions in California and Florida, with the help of the Interwork Institute at San Diego State University. The curriculum addresses the goals of employment empowerment by focusing on the needs of students with disabilities, such as strategies for how to disclose disabilities that make them attractive employees. Materials for instructors and implementation support for the curriculum are publicly available from the Interwork Institute.

Demonstration description. The demonstration proposed here addresses only the curriculum component of the proposal because it is the most straightforward to evaluate. Compelling findings from a rigorous evaluation of the intervention and demonstrated capacity to adapt it for wider audiences might prompt or contribute to the systems change envisioned by the proposal author.

This demonstration consists of two separate parts. First, to establish the effectiveness of the curriculum with its initial focal population, an impact evaluation could be carried out with students at two- or four-year colleges. The evaluation would compare employment empowerment training to the standard set of career services offered to students with disabilities by randomly assigning some students to be eligible for (or specially invited to) the training. Second, separate from the postsecondary education impact evaluation, this curriculum could be adapted for youth in a broader range of settings outside of higher education. This adaption, which is a key need, requires a pilot study in a setting such as high schools or VR agencies to document the time and resources needed. Though this pilot would not use an experimental design, it would provide correlational evidence on student outcomes in new settings.

Lead and partner organizations. Evaluating the curriculum's effectiveness would require the cooperation of one or more two- or four-year colleges or universities, with a state department of higher education or researchers from the participating schools leading the study. Drawing these schools from a single state could simplify aspects of the data collection or institutional recruitment process, as described below. Likewise, the pilot study would require the lead organization to recruit high schools, a VR agency, or other organizations interested in adapting the curriculum for their service population. The participation of a state or local education authority could bolster this process.

B. Evaluation rationale and overview

Evaluation rationale. Previous implementations of the curriculum have provided proof of concept by demonstrating improvements in attitudes among class participants. However, a rigorous evaluation is needed to provide evidence that the curriculum improves subsequent employment outcomes relative to existing services. The evaluation results would be of interest to secondary and postsecondary education staff, transition professionals, career counselors, and others responsible for job skills training for students with disabilities.

Evaluation design overview. A rigorous evaluation of the employment empowerment curriculum's employment impacts requires an RCT with at least a four-year timeline. The evaluation could be conducted with students in coordination with the career services offices at one or more postsecondary education institutions. Multiple sites might be needed to obtain a large enough sample, perhaps drawing on public institutions within a single state or campuses within a university system.

The nonexperimental pilots of the curriculum outside of higher education, such as in high schools or a VR agency, would involve a set of locations that span the range of institutions offering the curriculum, with students or clients comparable to the population for whom the curriculum would be adapted. Because some state agencies (such as Medicaid agencies) have access to SSI program participation data, depending on state data-sharing regulations, it might be possible for an agency to pilot the curriculum exclusively with youth receiving SSI. The piloting agency would record the process and measure the costs of adapting the curriculum, training instructors, recruiting participants, and delivering the curriculum.

C. Evaluation design

Sample and recruitment. The employment empowerment curriculum can be offered to youth or young adults with a disability attending postsecondary education institutions and adapted for other populations. In 2018, according to the American Community Survey, there were about 2.5 million people with disabilities ages 16 to 24 in the United States. Youth would be recruited by colleges or universities (for the impact evaluation) or other institutions and agencies (for the pilots) implementing the curriculum, drawing either from their student bodies or their existing service population. Because these institutions already identify youth with disabilities and have established means of communicating with them, they have a straightforward route to recruiting participants.

In fall 2019, US public two- and four-year postsecondary education institutions enrolled around half a million undergraduate students with disabilities. The average state has about 30 such public institutions, each enrolling around 300 students with disabilities on average.[4]

The pilot studies would not require a large sample of organizations or youth because they would primarily focus on implementation and cost outcomes. If possible, comparing participants' outcomes to those of similar youth who did not receive the curriculum would give researchers a rough counterfactual for pilot participants' outcomes and add context to the pilot's findings.

Design. Under an RCT approach to the postsecondary education impact evaluation, the evaluator would randomly assign a subset of the students who had registered with their respective institutions as having a disability to receive an offer for employment empowerment training. This offer would be in addition to

[4] These estimates are based on the authors' calculations using data reported to the Integrated Postsecondary Education Data System, based on mid-range assumptions regarding enrollment at institutions where precise figures were too low to publicly report.

other career services typically offered by the university. The control group would also be eligible for those usual services, but not employment empowerment. By randomizing on strata defined using available characteristics where possible, such as disability or academic major, evaluators could ensure a balance of the treatment and control groups on these characteristics. The demonstration would offer the training to students in or entering their final year of school to reach students at a time when they were likely to be actively searching for jobs and shorten the follow-up period.

This evaluation design would estimate the impact of being *offered* employment empowerment, rather than actually receiving it (that is, the "intent to treat" measure). This impact is likely the most policy-relevant measure, but it would be relatively straightforward to estimate the impact on those who actually participated based on program information. (The approach highlights another benefit of recruiting multiple institutions to participate in the study, because obtaining a sufficient number of youth participants in the treatment group is contingent on the rate of response to the enrollment invitation.)

The purpose of the nonexperimental adaptation pilots is to assess implementation feasibility and costs rather than impacts, reducing the importance of an experimental design. However, a sufficiently large state or local agency might be able to assemble a comparison group of youth for whom the curriculum was not offered; the employment, earnings, and other outcomes for these youth could serve as a benchmark or counterfactual for pilot participants. One benefit of this set-up would be that the implementing agency would have access to administrative data and contact information for intervention participants and comparison group members, and thus a student survey could be used for gathering outcome data. This benchmarking approach could build a correlational evidence base around the curriculum to inform future studies.

Research questions. The evaluation would address the questions about employment empowerment posed in Exhibit VI.1 using the data sources and analytic methods listed for each analysis type.

Data collection. The evaluation would use data from the following sources**.**

- **Baseline data** collected through an application form would provide information on a student's characteristics before signing up for the class, including attitudes and expectations about employment.
- **Program administrative data** from educational institutions participating in the impact evaluation would provide participant demographic characteristics, disability information, and other background data. They would also include academic measures, such as major and grade point average; academic outcomes, such as on-time graduation and credential attainment; and program data, such as participation in career readiness activities. For the pilot, participating agencies would provide similar information on participant characteristics, along with relevant education and service measures (depending on the type of agency).
- **Document reviews** of lesson plans, classroom presentations, and class materials would enable a comparison of the curriculum in practice with the intended design. It could also provide information about the counterfactual or existing services and trainings to which both treatment and control group members have access.
- **Staff interviews** would provide information on experiences with learning and applying the curriculum along with impressions of student receptivity with the content.
- **Classroom observations** of a subset of employment empowerment course sessions would supplement the measures of fidelity implementation and student engagement.

- **State administrative data**, particularly from the UI system, would measure employment and earnings outcomes for participants.
- **Participant survey data** would document outcomes not available through program or state administrative records. Surveys are the only possible source for measuring attitude and knowledge outcomes for the impact evaluations, as well as aspects of the participant's job application process (such as disclosure of a disability to a prospective employer and the student's approach to framing it). Questions measuring attitudes and knowledge would ideally be administered both before the intervention (at baseline) and after the intervention to measure changes in attitudes rather than just post-intervention levels.

Exhibit VI.1. Evaluation questions, data sources, and analytic methods for an evaluation of the Employment Empowerment proposal

Questions	Data sources	Analytic methods
Process analysis		
• Which students participated in employment empowerment courses? • How were participants different from the general population of students with disabilities? How were students recruited, particularly those of color or underserved communities? • How did postsecondary education institutions select and train instructors and develop the classes? • How was the curriculum implemented? Was it implemented with fidelity? • What types of organizations were interested in adopting the curriculum for the pilot, and for what types of students? • How was the curriculum adapted for the pilot?	• Baseline data • Classroom observations • Document reviews • Participant survey data • Program administrative data • Staff interviews	• Descriptive statistics • Qualitative analyses • Regression analyses
Impact/outcome analysis		
• What was the impact of the curriculum on intermediate outcomes, such as attitudes about employment (self-efficacy, employment prospects, disability disclosure change), education, and employment? • What was the impact of the policy change on the ultimate outcomes of employment and SSI benefit receipt? • Was the curriculum more effective with some students than others? • What were the outcomes of pilot participants, such as attitudes about employment, education, and employment?	• Participant survey data • Program administrative data • State administrative data	• Regression analyses
Benefit-cost analysis		
• Are the benefits of the curriculum large enough to justify its cost? • How do benefits and cost differ by stakeholder perspective (such as students and postsecondary education institutions)? • What are the costs to organizations working with youth outside of higher education to adopt the curriculum for their population?	• Participant survey data • Program administrative data • State administrative data	• Descriptive statistics • Regression analyses

SSI = Supplemental Security Income.

Analysis and reporting. The evaluation of the employment empowerment curriculum would include the following analysis and reports.

- For the impact evaluation, the timeline would require about four years: one year to identify a sample and prepare to deliver the curriculum, one academic year to offer the course to students in their final year of school, a one-year follow-up period after which surveys were fielded, and a subsequent period for analysis and write-up of the results. A single report would include findings from the process, impact, and benefit-cost analyses. Education, employment, earnings, and attitude/knowledge outcomes would be measured using a combination of administrative data and surveys conducted 18 months after random assignment; this period allows time for students to graduate and find a job.
- The pilot evaluation would unfold on a similar timetable: one year to adapt the curriculum, one year to offer the curriculum, a follow-up window in which to observe youths' outcomes, and a year for analysis and reporting. The report could assess the implementation of the curriculum adaptations, youth participation patterns, and agency costs, and it could include a comparison of participants' outcomes against those of similar youth who did not use the curriculum.

D. Proposal refinements and other considerations

In the future development of this intervention, a logic model would be useful to understand the relationship between intervention elements and outcomes and to benchmark the curriculum's performance with respect to those outcomes. It would also support adaptation to other populations.

Evaluators might consult policymakers and practitioners to prioritize other settings and populations of youth with disabilities for which to adapt the curriculum in the pilot studies. While we propose high schools and VR agencies, other organizations could include parent training and information centers, workforce centers, and developmental disability agencies.

This evaluation design focuses on one aspect of employment empowerment—testing the effectiveness of the curriculum on youth outcomes. The other two aspects of the proposal—a call for broad cultural change among disability policymakers and service providers and a federal cross-agency initiative to promote that change—could be assessed through a developmental evaluation that tracks the training on and use of the curriculum by the intended stakeholders. These systems-change promotional efforts could increase the impact of employment empowerment instruction if they encourage service providers and youth to be more open to adapting, offering, and enrolling in the curriculum.

This page has been left blank for double-sided copying.

VII. Family Employment Awareness Training (FEAT): A Research-Based Program for Promoting High Expectations for Employment and Knowledge of Resources

A. Proposal and demonstration description

Proposal description. Family Employment Awareness Training (FEAT) is a two-part training for youth with disabilities transitioning from school to adulthood, their families, and the professionals who support them (Gross et al. 2021). FEAT's purpose is to raise attendee expectations in obtaining competitive employment and increase their knowledge of how local, state, and federal resources can make employment a reality. FEAT targets transition-age youth ages 14 to 22 with significant support needs. The training bridges the gap between secondary school and adulthood by organizing and simplifying issues related to finding jobs, accessing services and supports, and offering youth and those who support them guidance to facilitate successful school-to-employment transitions. The training addresses transition issues such as customized employment, the family's role in transition, and available resources, and it is tailored to the state where it is offered. Attendees develop plans for employment along with next steps to use what they learned about available resources. To overcome employment barriers they might encounter, FEAT offers attendees opportunities to participate in individual follow-up technical assistance activities within one to two months after the training (via phone or email) and group technical assistance activities at a local workforce center three and nine months after the training. FEAT has been implemented in five states to date. States wanting to offer the training, which is typically coordinated by a parent training and information center, need to license the FEAT curriculum from the University of Kansas, which developed the program.

Demonstration description. An evaluation of FEAT could either draw from newly participating youth in states with an existing program (Kansas, Indiana, Nebraska, Oklahoma, and Rhode Island) or from one or several states that would newly implement the intervention. Conducting the demonstration in a new state would be informative for a process analysis of facilitators and challenges surrounding a new implementation and a fidelity analysis of how well states can adhere to the model when designing their training. If the demonstration is conducted in a state with an existing FEAT program, participants could enroll immediately. For newly implementing states, lead and partner organizations would need to collaborate with the FEAT consulting team to develop the curriculum and identify speakers for the training sessions. The curriculum must be licensed from the University of Kansas, whose staff developed FEAT; licenses cost $3,000 for the first year and $1,000 for each subsequent year. Starting up the program in a new state would take about half a year, after which participants could start enrolling in the training. The demonstration would enroll participants for about one or two years, depending on the number of training sessions, to ensure a large enough sample for an evaluation.

Lead and partner organizations. In states with existing FEAT programs, a parent advocacy organization, such as the state parent training and information (PTI) center or a state developmental disability council, is the lead organization. These types of organizations would likely take the lead in a newly implementing state. Partners that could also be involved in the intervention include workforce development centers, state VR agencies, state disability services or Medicaid agencies, developmental disability councils, community rehabilitation providers, and state departments of education. These partners would advertise FEAT to their clients, potentially present about their services at the training, and provide financial support to the lead agency adapting and implementing FEAT. State VR agencies would

play an important role because they would identify transition-age clients and refer them to FEAT training; the evaluation would also rely in part on VR data around employment and other outcomes. The intervention is delivered through a train-the-trainer model, so the lead organization would work to adapt the curriculum to a new state or locality together with the FEAT consulting team, which operates out of Indiana University and is led by one of the proposal's authors.

B. Evaluation rationale and overview

Evaluation rationale. FEAT has been evaluated in two studies: (1) a mixed-methods pilot study that assessed the intermediate-term (one-year) outcomes of FEAT attendees in 2012 (Francis et al. 2013) and (2) a quasi-experimental mixed-methods study comparing short- and intermediate-term outcomes of attendees and non-attendees from 2013 to 2016 (Gross and Francis 2016; Francis and Gross 2017). These studies considered the effect of FEAT on employment expectations; they did not evaluate its impact on employment outcomes or benefit use. Although the existing evidence shows that attendees reported higher expectations for employment, the studies did not assess whether these increased expectations led to significantly higher employment levels among attendees compared with similar youth who did not attend the training. Understanding whether FEAT can be an effective intervention to improve employment outcomes of transition-age youth and reduce their reliance on SSI and other benefits requires a rigorous evaluation of long-term outcomes.

Evaluation overview. Because FEAT has been implemented in five states to date and the authors of the proposal previously conducted a pilot study, no additional feasibility or pilot study is necessary. Instead, a rigorous impact evaluation relying on random assignment would provide evidence on the employment and program participation impacts of FEAT for participating youth in states with established FEAT programs or in one or several states where the program will be newly implemented. An RCT would compare the employment and SSI outcomes of transition-age VR clients who enroll into the demonstration; these clients would be randomly assigned to either a treatment group (and thus participate in a FEAT program) or a control group (and thus use usual services). To assess the intervention's impacts, outcomes of FEAT attendees and the control group would be measured for two years to estimate how the intervention affects employment and SSI outcomes. The evaluation sample would be drawn from youth referred by local VR agencies in states with an existing training program and one or several states that will newly implement the intervention in the future. Using data from the five states with existing FEAT programs and possibly additional states where FEAT may be implemented in the future over a two-year period should yield a large enough sample size to detect meaningful differences between the treatment and control groups.

C. Evaluation design

Sample and recruitment. The population for this demonstration consists of youth ages 14 to 22 who use VR services. Although potential FEAT participants might include non-VR clients, randomly assigning youth using VR services into treatment and control groups has the advantage of facilitating access to rich administrative data from the state VR agency. It would also enable comparisons with a group of youth similarly motivated to seek employment-promoting services as those who attend FEAT. The VR agency in the state where FEAT is implemented would identify clients who might benefit from attending FEAT, conduct outreach to youth and families, and randomly assign youth who enroll to a treatment or control group. The VR agency would then refer clients assigned to the treatment group to the training. Clients who were randomly assigned to the control group would not receive a referral.

Design. A five-year evaluation would rely on an RCT design to assess the two-year impacts of FEAT. The evaluation would include an analysis in Year 2 to understand the demand for the intervention and the characteristics of the youth who sign up for it. It would also explore the willingness of partnering providers to attend the training. A process analysis in Year 4 would document the experience of FEAT teams with established programs and the facilitators and challenges with a newly implemented training (if the demonstration involves new states that adopt the program). An impact analysis—conducted in Year 5 to allow a two-year observation period after youth enroll—would compare the service, employment, expectation, and SSI outcomes of FEAT attendees with those of the control group. Based on the employment and SSI impacts and considering the cost of the training, a benefit-cost analysis would show whether increased employment can offset the cost.

Research questions. The evaluation would answer the questions in Exhibit VII.1 using the data sources and analytic methods listed.

Data collection. The evaluation would require different types of data to address the research questions.

- **Baseline data** from a survey or application would collect information on the characteristics of youth who apply for FEAT, including an assessment of employment expectations and resource knowledge.
- **Program administrative data** collected by the VR agency—either parallel to or as a part of their existing management information system—would track referrals, enrollment, and use of technical assistance services after the training. The process and benefit-cost analyses would rely on this information. Service data that the VR agency typically collects would provide information on VR service use and outcomes (such as employment) for both treatment and control group members.
- **Staff interviews**, conducted throughout implementation, would contribute to understanding FEAT implementation, changes over time, and partner relationships.
- **Focus groups and structured interviews** with participating youth and their families would provide an opportunity to solicit their opinions on the program and about service use.
- **Participant surveys** would inform the ultimate outcomes relatively early, collecting information from youth about employment expectations, employment, and SSI outcomes. The surveys could be fielded immediately after the training (for treatment group members only) and two years after training (for both treatment and control group members).
- **State administrative data** would provide information on key outcomes for treatment and control group youth. Data would include earnings from state unemployment insurance and education attainment from state education databases.
- **SSA administrative data**, provided through a partnership with SSA, could provide accurate information on SSA program enrollment and benefit data for all members of the treatment and control groups.

Exhibit VII.1. Evaluation questions, data sources, and analytic methods for an evaluation of the Family Employment Awareness Training proposal

Questions	Data sources	Analytic methods
Process analysis		
• How many and what were the characteristics of youth who expressed interest in attending FEAT? • How many participants and families attended FEAT trainings? How many used the technical assistance offered after the training? • Which organizations were involved in the implementation of FEAT in the state(s) on which the evaluation is based, and what were their roles? What training was needed? • How was FEAT implemented? How was it adapted to its state or locality? • How were FEAT attendees recruited? How were youth and families of color or from underserved communities recruited?	• Baseline application data • Participant focus groups and structured interviews • Program administrative data • Staff interviews	• Descriptive statistics • Qualitative analyses
Outcome/impact analysis		
• Did FEAT affect intermediate outcomes, such as employment expectations, service use, resource knowledge, and postsecondary education? • What was the impact of the training on the ultimate outcomes of employment and SSI benefit receipt? • How did employment expectations and resource knowledge resources change among FEAT attendees from before to directly after the training? • Was the program more effective for some participants than others?	• Participant survey data • SSA administrative data • State administrative data	• Regression analyses
Benefit-cost analysis		
• Do the benefits of FEAT have the potential to offset its cost? • Did the benefits and cost differ by stakeholder perspective (such as participants, the lead organization, and state and federal agencies)?	• Participant survey data • Program administrative data • SSA administrative data • State administrative data	• Descriptive statistics • Regression analyses

FEAT = Family Employment Awareness Training; SSA = Social Security Administration; SSI = Supplemental Security Income.

Analyses and reporting. The FEAT evaluation would observe both early and long-term employment and SSI outcomes. The evaluation could include the following analyses and reports.

- Formative assessments, using program administrative data, would document intervention implementation and changes to the FEAT model.
- An early process analysis would document participant enrollment and service use, enrollment of service providers in the training sessions, and partner engagement. It would include descriptive analyses of training attendance and qualitative analyses of interview and focus group data.

- Early outcome analyses conducted immediately after each training session for treatment group members would contrast the immediate outcomes of participants in response to the training. This information could also be combined with the two-year survey data to show temporal patterns in changes for these outcomes.
- A final impact analysis, conducted five years after demonstration start, would assess treatment and control group outcomes from administrative and survey data.
- A benefit-cost analysis would assess whether differences in earnings and benefit levels between treatment and control groups, extrapolated into various points in the future, could justify the cost of FEAT, and if not, what impacts would be needed to offset program costs.

D. Proposal refinements and other considerations

The intervention consists of a two-part training and follow-up technical assistance, so there is relatively limited scope to change any intervention components for the purposes of an evaluation. FEAT has been implemented in five states to date, so there are no concerns about the feasibility of its implementation. Given the horizon required to evaluate the employment and education outcomes of youth who attend FEAT at a young age, it might make sense to either restrict the training to youth ages 18 to 22 or to evaluate its impact among attendees in this age range in a subgroup analysis.

Because the demonstration would be conducted in multiple states, the evaluation could consider whether the program is implemented differently across states. The process analysis could assess fidelity of the model across states through interviews with lead agency and VR staff. Depending on potential differences in implementation, evaluators could consider conducting separate impact analyses for each state, provided sample sizes are large enough.

This page has been left blank for double-sided copying.

VIII. The Family Empowerment Model: Improving Employment for Youth Receiving Supplemental Security Income

A. Proposal and demonstration description

Proposal description. The Family Empowerment Model (FEM) proposes a combination of evidence-based and novel solutions to improve the outcomes of youth receiving SSI beginning as early as age 14 and continuing through age 24. The FEM incorporates three main components (Anderson et al. 2021). The first component is targeted direct and indirect outreach to inform youth and their families about FEM and encourage their take up of services. The second component relies on a family empowerment specialist (FES), a person with lived experience who connects with youth and families, informs them about available school-to-work transition services, and raises their expectations for youth outcomes. Dependent on youth needs, but typically within two years, the FES connects youth to the third component, integrated resource teams (IRTs) comprised of transition, employment, and education service providers. With meetings led by the youth, IRTs can connect youth with relevant employment and other supports and track progress toward program commitments and youth goals.

FEM intends to bolster the expectations of youth and families regarding what youth can achieve and to increase service engagement and connections throughout a youth's transition process. Youth would then have greater access to postsecondary education or training, long-term employment, greater financial stability, and reduced poverty —thus improving outcomes related to educational achievement, employment, and financial self-sufficiency.

Demonstration description. For evaluation purposes, an eight-year demonstration of FEM would develop a program within a single state. The eight-year period allows for program development, a two-year enrollment period, and sufficient time to support cohorts of youth and their families through FESs and IRTs for four years plus conduct evaluation activities. Youth and families would begin working with an FES when youth are ages 14 to 16 for a two-year period. During that time, FESs would assemble an IRT specific to each youth; the youth would begin working with the IRT approximately two years after they begin working with the FES (or when youth are ages 16 to 18). Youth would then work with IRTs for an additional two-year period or until they no longer need support. FESs will need training and technical assistance in various areas, such as person-centered planning; motivational interviewing; trauma-informed care; self-determination; advocacy; service navigation; and diversity, equity, and inclusion. Members of the IRT would need training to support their roles for the FEM.

Lead and partner organizations. The lead organization should be one that serves a broad disability community, is already present as a transition service partner, and can flexibly hire staff for FES positions. Centers for independent living may be uniquely positioned to take the lead organization role, particularly given their experiences with youth with disabilities and their families, their role in the community as a resource hub, and their focus on self-advocacy and independent living skills. Partner organizations will include the entire range of those already involved with transition, as their staff will need to participate in IRTs. These organizations include secondary and postsecondary schools, the VR agency, workforce center, Medicaid, SNAP, TANF, and disability-specific organizations such as developmental disability or mental health agencies. To obtain lists of youth receiving SSI, a demonstration would employ SSA or the state Medicaid agency as a partner.

B. Evaluation rationale and overview

Evaluation rationale. Comprehensive case management as proposed through FEM, employing persons with lived experience, has not been tested explicitly. An RCT could provide a test of the model's feasibility and effectiveness. A demonstration could also assess and identify any challenges in recruitment, implementation, and service delivery. Though PROMISE tested a similar model, FEM differs from PROMISE in that it (1) relies on a case manager with lived experience, (2) emphasizes culturally responsive training, and (3) offers connections through a two-stage process, with the FEM handing the youth to an IRT, which the youth leads.

Given the upcoming release of the PROMISE five-year impact findings, the results could affect the need to test FEM using an RCT. As discussed in Chapter X for the Youth and Family Systems Navigator proposal, if PROMISE has positive impacts, a rigorous evaluation of FEM might not be necessary because PROMISE would have provided evidence of the benefits of a case management model connecting youth and families to services. A feasibility study for FEM could then document the mechanisms, facilitators, and challenges with the model's implementation.

Evaluation design overview. The intervention would be tested through an eight-year evaluation using an RCT. Enrollees would include youth between the ages of 14 to 16 currently receiving SSI and their families. The demonstration would occur in a single state or geographic area for ease of administration, given the relationships needed by the lead organization and its partners. The lead organization would recruit youth and families based on lists obtained through SSA or the state Medicaid agency; youth assigned to the treatment group would have access to FEM supports for up to four years, while youth assigned to the control group would have access to usual transition services. The evaluation would follow youth over this period using administrative and survey data.

C. Evaluation design

Sample and recruitment. The FEM would offer services to youth receiving SSI and their families beginning when youth are ages 14 to 16 and continuing through a four-year period. Targeted outreach, as noted, is a key component of the model. A demonstration would involve outreach to youth receiving SSI ages 14 to 16 using a two-pronged outreach approach. For the first approach, the lead organization would obtain contact information for youth receiving SSI and their families through agreements with either SSA or the state Medicaid agency. FESs would then mail information packets and postcards to youth and conduct other individualized outreach to encourage their enrollment into the demonstration. For the second approach, FESs would partner with local community organizations to educate families and providers about the demonstration and its services, both to encourage youth and families to sign up for the demonstration and to build needed relationships with these organizations. Half of those who sign up to enroll would be assigned to a treatment group that would receive the offer of FEM services; the remainder would be assigned to a control group that would receive usual services.

Design. We propose an eight-year RCT to take place in a single geographic location. In Year 1, the lead organization would operationalize the intervention, hire and train FESs, and develop partnerships specifically for the FEM with other organizations. In the second and third years, the lead organization would recruit youth and families to enroll in the demonstration; upon enrollment, the FESs would initiate their work with youth and families. In the fourth and fifth years, youth and families would transition from the FES to the IRT. By Year 7, the last enrolled youth will have finished working with their IRTs. In addition to early or interim reports, the evaluation would produce a final impact report in Year 8. The

evaluation would assess the impacts of the FEM on service use, educational attainment, employment outcomes, and financial self-sufficiency. It would also measure aspects related to implementing the three FEM components (outreach, FESs, and IRTs).

Research questions. The evaluation would address the questions in Exhibit VIII.1 using the data sources and analytic methods listed for each analysis type.

Data collection. The FEM evaluation would use data from the following sources.

- **Baseline data** collected through a survey or application form would obtain information on the characteristics of youth and families who enroll in the demonstration at the time of enrollment.
- **Program administrative data** from the lead organization, such as case notes, documents, and management information system extracts, would enable the evaluation to track program implementation, enrollment, staffing, and service delivery. Depending on the FEM phase for the youth, either the FES or the lead contact for the IRT would input required information into the lead organization's system database. This information would inform the process analyses and provide the detailed cost data necessary for the benefit-cost analysis.
- **Staff interviews** would include periodic interviews (in person or by phone) with FESs, IRT members, and administrators of the lead organization about their involvement with the FEM and its implementation.
- **Focus groups or structured interviews** with participating youth and their families would provide their insights about the program and usual services.
- **Participant surveys** would collect information from treatment and control group members on key outcomes, such as employment, earnings, education, service use, financial self-sufficiency, community integration, empowerment, and expectations. This information would enable the evaluation to gauge FEM's impacts and inform the benefit-cost analysis. The surveys could be offered at two and four years after enrollment, which would track the changes from baseline through participants' involvement with the FES and then again through their involvement with the IRT. A lower cost approach could omit the two-year survey.
- **State administrative data** could supplement other data sources to inform earnings outcomes (through unemployment insurance records), public health coverage and expenditures (through Medicaid data), secondary and postsecondary education outcomes (through education databases), and employment services (through vocational rehabilitation agency data).
- **SSA administrative data** would provide accurate information on SSI-related outcomes, such as benefit amounts, age-18 redetermination outcomes, work incentive use, and annual earnings.

Exhibit VIII.1. Evaluation questions, data sources, and analytic methods for an evaluation of the Family Empowerment Model proposal

Questions	Data sources	Analytic methods
Process analysis		
• Who participated in the FEM and how were participants different from other youth receiving SSI? • How were youth recruited? Which outreach methods were most effective? How were youth and families of color or from underserved communities recruited? • How was the FES component implemented and what factors contributed to the implementation experience? • How was the IRT component implemented and what factors contributed to the implementation experience? • What services did participants use and how often did they engage with the FES and IRT? How did those patterns change as participants aged? How did services vary for youth and families of color and from underserved communities? How did FESs engage with participating youth and families? • Was the intervention implemented as intended? • What program or policy changes did the lead organization and its partner organizations make in response to the intervention?	• Baseline application or survey data • Participant focus groups or structured interviews • Participant survey data • Program administrative data • Staff interviews • State or SSA administrative data	• Descriptive statistics • Qualitative analyses
Impact analysis		
• What was the impact of the FEM program on intermediate outcomes, such as service use, expectations, resource navigation, work experiences, measurable skill gain, community integration, and educational attainment? • What was the impact of the FEM program on the ultimate outcomes of employment and SSI benefit receipt? • Was the program more effective with some youth and families than others?	• Participant survey data • State administrative data • SSA administrative data	• Regression analyses
Benefit-cost analysis		
• Are the benefits from the FEM large enough to justify its cost? • How did benefits and cost differ by stakeholder perspectives (such as participants, the lead organization, and local, state, and federal governments)?	• Participant survey data • Program administrative data • SSA administrative data • State administrative data	• Descriptive statistics • Regression analyses

FEM = Family Empowerment Model; FES = family empowerment specialist; IRT = integrated resource team; SSA = Social Security Administration; SSI = Supplemental Security Income.

Analyses and reporting. The evaluation of the FEM would produce reports of findings throughout the demonstration period and could include the following analyses and reports.

- A formative assessment using program administrative data would continually document how FESs and IRTs implement the intervention. It would also track changes or improvements to the FEM model.
- An early process evaluation conducted two years into the demonstration (after the first year of enrollment) would document initial enrollment and recruitment activities, the initial activities of the FESs, and the partnerships that the lead organization develops in support of the model. The lead organization could use this information to improve or adjust its approach to outreach and FES services.
- A final process evaluation conducted in Year 8 would document the participation of youth, youth and family engagement and service use, FES and IRT practices, training, and organization partnerships. It would include descriptive analyses of program data and qualitative analyses of interview and focus group data. An important aspect of the process evaluation is an assessment of how FESs incorporated their training (such as the use of culturally-appropriate or trauma-informed approaches) into their service models.
- An early impact analysis two years after enrollment would compare the treatment and control groups' outcomes as they complete FES services by assessing intermediate outcomes, such as expectations, service use, and postsecondary education and training, through participant survey and state administrative data.
- A final impact analysis, conducted four years after enrollment, would assess treatment and control group outcomes as they complete IRT services. The analysis would draw on participant survey data, state administrative data, and SSA administrative data to assess intermediate and ultimate outcomes.
- A benefit-cost analysis would use the final impact analysis findings and cost data (from the program administrative data and other sources) to calculate whether and when the benefits that can be attributed to the FEM outweigh the program costs.

D. Proposal refinements and other consideration

The proposal by Anderson et al. (2021) suggests two treatment arms for a test of the FEM. For the first arm, participants would have access to the FEM and not the IRT. For the second arm, participants would have access to both the FEM and the IRT. We did not include this approach for the proposed evaluation design because results from PROMISE will show results for a model that is similar to the FEM alone. However, an evaluation with these two arms could show the added value of the IRT.

If the demonstration uses an organization that has limited experience with partnerships and transition services as the lead organization, the involved staff might benefit from technical assistance to build the service model and connect to other organizations. The proposed demonstration and evaluation design assumes that a CIL would be the appropriate organization to lead the effort. As the proposal documents, CILs might encounter challenges with leading an employment-oriented model, but all organizations that could lead the FEM face potential (and different) challenges.

This page has been left blank for double-sided copying.

IX. Improving Youth SSI Recipients' Employment Outcomes Through an Integrated Treatment Team Intervention in a Health Care Setting

A. Proposal and demonstration description

Proposal description. The Integrated Treatment Team intervention proposed in Taylor et al. (2021) aims to improve employment and other outcomes for youth receiving SSI between the ages of 14 and 18 who use multidisciplinary health clinics with integrated treatment teams. Pediatric multidisciplinary clinics offer a setting where patients with disabilities, including youth receiving SSI, work with integrated treatment teams. These teams often comprise doctors, nurses, nutritionists, social workers, and a variety of therapists. Hallmarks of the teams are a high degree of collaboration and communication among team members and a focus on improving coordination among the different systems of care often faced by patients with complex medical needs. These settings provide an opportunity to offer coordinated services designed to promote the employment of youth with disabilities.

The proposed intervention adds an employment-focused case manager to established integrated treatment teams that care for youth receiving SSI in multidisciplinary health clinics. The goal of the case manager is to (1) coordinate health care between medical providers and public benefits programs and (2) increase provider and family awareness of employment-related issues and opportunities as a part of youth's health care treatment through provider training and youth referrals to local resources. The intervention intends to improve the long-term employment outcomes of youth with disabilities by providing pathways for youth and families to utilize resources that promote employment (specifically VR services and benefits counseling) and using a health care setting to emphasize employment as a long-term, attainable goal.

Demonstration description. For the purposes of an evaluation, a three-year pilot demonstration of the Integrated Treatment Team intervention would occur in a multidisciplinary health clinic. The timeline would include one year to develop the intervention, assemble partners, and enroll youth and two years to observe system implementation and youth outcomes. During that period, employment-focused case managers would be added to already-existing treatment teams and patients who sign up for the program would be assigned to case managers. Through monthly engagement meetings, case managers would provide ongoing support, referrals, and assistance to the youth receiving SSI. The case manager would provide referrals, coordinate care, and maintain communication between state agencies and programs, including VR, Medicaid, and Community Work Incentive Coordinators (CWICs) or Community Partner Work Incentives Counselors (CPWICs). To maintain engagement across the integrated treatment team, questions about the patient's use of employment and transition services would be added to standard appointment forms. Case managers would receive 40 hours of initial training on the intervention processes, benefits counseling, public supports and referral pathways, and other employment-related information.

Lead and partner organizations. The lead organization would be a multidisciplinary health clinic that has a sizeable pediatric clinic. Multidisciplinary clinics already have an infrastructure of treatment teams, who work together to manage patients' complex medical conditions. Community partnerships will be essential to the successful implementation of the intervention. Partnerships would include the VR agency, the organization that offers benefits counseling, secondary and postsecondary education institutions,

Medicaid agencies, and workforce development centers. The lead organization must collaborate with these partner organizations to connect participants to their services.

B. Evaluation rationale and overview

Evaluation rationale. Given the limited evidence about the effectiveness of the intervention and potential challenges in implementing it in a health care setting, a pilot test could provide early evidence of an integrated treatment team's feasibility and potential. Although similar models exist to coordinate health and employment services for injured adult workers, the intervention is relatively untested for youth and therefore would benefit from an initial test of its feasibility in this setting. The PROMISE demonstration tested the effectiveness of case management and other services for youth, but this intervention differs in that the setting is specific to multidisciplinary health clinics with an existing treatment team in place and a more targeted set of case management goals. A pilot could identify benchmarks to measure success, gauge the youth and system outcomes that the intervention generates, and identify opportunities to improve the service model. A novel component of the intervention is integrating an employment focus into a health care setting. Although the intervention may not be generalizable to all health care settings, evaluation findings could provide initial evidence about whether the general approach holds promise. It would also provide lessons on training and orienting clinic staff toward an employment focus and establishing data systems to track services and outcomes. The primary audiences for the evaluation include health care providers in multidisciplinary clinics and hospitals, state VR agencies, entities providing benefits counseling services, and youth receiving SSI and their families.

Evaluation design overview. We propose a three-year pilot test that includes one year to develop the intervention, assemble partners, and enroll youth and two years to observe system implementation and youth outcomes. One of the primary goals of the pilot is to understand how an employment-focused case manager model could be implemented and adapted within a health care system that relies on an integrated resource team to offer services to youth and their families. The results from the pilot could therefore inform the development of a systematic method for establishing the case manager role, training clinical staff, and engaging youth and families. A large hospital with multidisciplinary clinics, similar to that described in the proposal, with a sufficient number of youth receiving SSI, could implement the pilot. The pilot would cover a 36-month period, which would provide sufficient time for youth receiving SSI to engage in case management services, act on referrals to VR and benefits counseling, and potentially engage in work-based learning experiences. The pilot would assess critical factors and potential challenges to implementation related to (1) patient enrollment, (2) engaging patients with case management, (3) successful pathways for referrals to benefits counseling and employment services, and (4) partnerships with organizations such as local VR offices and benefits counseling providers.

C. Evaluation design

Sample and recruitment. The proposal aims to promote employment and other outcomes for youth receiving SSI between the ages of 14 and 18 who use multidisciplinary health clinics with integrated treatment teams. While not all youth receiving SSI use such clinics, a sizeable portion likely represents the proposal's intended population. For example, 7 percent of youth using care through children's hospitals receive SSI. A single hospital whose medical care includes integrated treatment teams and with a large youth population would implement the pilot. Participant recruitment would come from the universe of eligible patients at that hospital. To identify youth receiving SSI for the pilot, the hospital would add a question to its standard health screening forms that asks about SSI receipt. Patients and families identified through the health screening question would be offered access to a pilot program that

features employment-focused case management. All interested youth would enroll in the pilot, with a goal of at least 50 youth participating in the pilot.

Design. The evaluation of the integrated team model would involve a pilot test of the model's feasibility. Because identifying an appropriate comparison group will likely be infeasible, we do not propose that the pilot include one. Youth receiving services through a multidisciplinary clinic likely have more significant or specialized health conditions and needs, which makes constructing a comparison group with existing administrative data (such as youth receiving SSI or using VR services) difficult. Similarly, existing benchmarks, such as employment outcomes of special education students or VR clients, might not provide useful comparisons for the population targeted by this proposal. The pilot could position itself to collect information about the youth who enroll, which would enable future evaluators to consider whether a valid comparison group could be created using existing data sources, especially if a state has a database that combines Medicaid, VR, and UI records. The pilot test would explore the demand for the intervention and the characteristics of the youth and families who enroll; document the intervention rollout, focusing on staff training and engagement, partner relationships, the case management role, and youth and family service use; and measure participant outcomes. It could also explore system outcomes, such as staff attitudes toward employment, staff discussions of employment and other changes to service delivery in their interactions with youth and families, and the integration of the case manager into the treatment team.

Research questions. The evaluation would address the questions in Exhibit IX.1 using the data sources and analytic methods listed for each analysis type.

Data collection. The pilot test of the Integrated Treatment Team proposal would require the following types of data to address each analysis component.

- **Baseline data** collected through a survey or application form at enrollment would obtain information on the characteristics of youth and families who enroll in the demonstration.
- **Hospital administrative data** would include case notes, program documents, and management information system extracts and enable the evaluation to track program implementation, enrollment, staffing, and service delivery. This information would inform the process analyses and provide the detailed cost data necessary for the benefit-cost analysis. These data could include information on health care visits and case management services.
- **Staff interviews** would involve periodic interviews with case managers (both existing and new employment-focused case managers), hospital administrators and other clinical staff, and staff from partner organizations about their involvement in and perceptions of the program. These interviews could occur periodically (every 6 to 12 months) and inform intervention delivery as a formative exercise.
- **Focus groups or structured interviews** with participating youth and their families could obtain insights and impressions about the program and services to improve the service model during the pilot and provide qualitative information to supplement the evaluation's quantitative findings.
- **Participant surveys** would collect information from participants on employment and education outcomes and knowledge of SSI work incentives. This information would enable the evaluation to gauge the Integrated Treatment Team program's outcomes. The evaluation could offer surveys in Years 2 and 3 of the demonstration to track changes that occur from the baseline through participants' involvement with the case managers and a post-intervention follow-up period.

- **State administrative data** could supplement other data sources to inform employment services (through VR agency data), public health coverage and expenditures (through Medicaid data), and benefits counseling (if available from the agency providing benefits counseling services).
- **SSA administrative data** would provide accurate information on SSI-related outcomes such as benefit amounts, work incentive use, and annual earnings.

Exhibit IX.1. Evaluation questions, data sources, and analytic methods for an evaluation of the Integrated Treatment Team proposal

Questions	Data sources	Analytic methods
Process analysis		
• How did the pilot identify youth receiving SSI and conduct recruitment? Was the health screener a useful recruitment tool? • How many patients enrolled in the pilot and how did enrollee characteristics compare with members of the clinical population receiving health care for their disability? With youth receiving SSI in general? How were youth and families of color or from underserved communities recruited? • Who were the primary partners in the intervention? Who initiated these partnerships and what were each partner's roles? How well did the partners communicate, collaborate, and engage with program goals? • How were providers trained on the intervention, how did their attitudes and behaviors change, and what factors contributed to the implementation experience? • How was the Integrated Treatment Team intervention designed, implemented, and operated, and what factors contributed to the implementation experience? • What were the patterns of case manager engagement with youth and their families? • What percentage of youth received referrals to VR services and benefits counseling? What percentage of youth used those services? • What program or policy changes did the hospital administration and its partners make in response to the intervention? • Was the intervention implemented as intended?	• Baseline data • Hospital administrative data • Participant and family focus groups and structured interviews • Participant survey data • Program administrative data • Staff interviews	• Descriptive statistics • Qualitative analyses
Outcome analysis		
• What were the intermediate outcomes of pilot participants, such as knowledge of work incentives and health insurance options, health coverage, employment service use, and benefits counseling? • What were the ultimate outcomes of pilot participants regarding employment and SSI benefit receipt?	• Participant survey data • SSA administrative data • State administrative data	• Descriptive analyses

Questions	Data sources	Analytic methods
Benefit-cost analysis		
• What were the program costs, and what benefits would be needed to offset those costs?	• Participant survey data • Program administrative data • SSA administrative data • State administrative data	• Descriptive statistics • Regression analyses

SSA = Social Security Administration; SSI = Supplemental Security Income; VR = vocational rehabilitation.

Analyses and reporting. The evaluation of the Integrated Treatment Team intervention would produce the following reports of findings.

- A process evaluation during Year 2 of the pilot would document the characteristics of the youth and families that participate in the demonstration, the participation rate of patients, experiences in recruiting patients and training case managers, partnership building, and qualitative findings from treatment team members, hospital staff, and participants and families on their experiences with the program.
- An outcome analysis at the end of the intervention would document participants' outcomes in employment and transition service use, knowledge of SSI work incentives, knowledge of their medical conditions and how it relates to potential employment, and other outcomes.
- A cost analysis would use the data from the program administrative data and other sources to document the cost of the Integrated Treatment Team program and assess what benefits would be needed to outweigh those costs.

D. Proposal refinements and other considerations

Without a comparison group, the evaluation cannot assess whether the outcomes of the pilot group represent any potential benefit of the program. Use of benchmarks—perhaps from the PROMISE evaluation, the VR agency statistics, evaluations for similar programs and populations, or youth using services in the multidisciplinary health clinic in prior years—could offer useful context for interpreting the outcomes of pilot participants.

An important aspect of the evaluation will be to consider that the organizations that bear the cost of implementation may not necessarily be those that would accrue the eventual benefits, if any, of this intervention. In other words, costs would be incurred primarily by the health clinic, and benefits would go to the youth and to the federal government (through reduced SSI and Medicaid). Therefore, an evaluation might consider how clinic reimbursement through Medicaid or other sources along with other incentives can offset the cost of case management services for health care clinics.

This page has been left blank for double-sided copying.

X. Policy Considerations for Implementing Youth and Family Case Management Strategies Across Systems: Youth and Family Systems Navigator

A. Proposal and demonstration description

Proposal description. The overall goal of the youth and family systems navigator (YFSN) is to encourage greater financial independence and self-sufficiency among youth receiving SSI through securing and maintaining employment (Karhan and Golden 2021). The YFSN could achieve this by connecting youth and families to the existing scaffolding of transition-related supports and services through comprehensive case management. The proposal targets youth ages 14 to 24 who receive SSI, along with their families. Its goal is to engage youth at age 14, two years before the Individuals with Disabilities Education Act's requirement to begin transition planning at age 16, and offer supports and services until youth turn age 24. The core functions of the case manager—to educate youth and families about transition and connect them to services—take place in three phases.

1. The YFSN completes a family systems assessment to analyze factors for a successful transition, including basic needs, cultural norms, past trauma and adversity, and familial crisis.
2. The YFSN then develops a multi-pronged intervention plan to address the needs and points of adversity as they emerge.
3. The YFSN, in collaboration with youth and their families, sets and implements short-, medium-, and long-term goals toward employment and independent living. To do so, the YFSN creates linkages to local and state resources to address needs and goals.

Demonstration description. For purposes of an evaluation design for a demonstration of the YFSN model, we assume that a state would implement a demonstration under a Department of Health and Human Services (HHS) 1115 waiver, as described in the proposal. This approach would enable an entity funded through Medicaid to provide the YFSN services to Medicaid-eligible youth receiving SSI and their families. A single agency, such as a nonprofit organization with the capacity to bill for Medicaid services, would deliver the YFSN services. Implementation of the model would occur in a single or small number of geographic locations to simplify implementation and limit the resources and number of connections required (for example, to high schools, postsecondary education institutions, and workforce and VR agencies). After an initial start-up year to hire and train staff and build relationships within the community for the program, the organization would conduct outreach to youth receiving SSI who are age 14 over a two-year span with the offer of using YFSN services through age 24. The demonstration period would cover 14 years as the YFSN works with the youth and family to offer case management services, including (1) advocacy and self-sufficiency, (2) assessments, (3) resource brokering, (4) case planning, (5) crisis management, (6) education engagement and preparation, (7) employment preparation, and (8) meeting basic needs. The geographic location(s) would require a large enough number of youth receiving SSI to meet the evaluation sample size requirements. The intervention might also be easier to implement

in states where SSI eligibility automatically qualifies a youth for Medicaid and Medicaid enrollment is automatic (Rupp and Riley 2016).[5]

Lead and partner organizations. The lead organization responsible for the demonstration could be a nonprofit organization, such as a center for independent living or parent training and information center. The key requirements are that it has the capacity to support the YFSN staff for the demonstration period and that it can bill Medicaid for its services. Partner organizations necessary for implementation include those typically involved with transition service coordination and delivery: secondary and postsecondary education institutions, workforce partners such as vocational rehabilitation agencies and American Job Centers, and developmental disability agencies. For the evaluation, the demonstration would need partnerships for identifying youth receiving SSI (through data provided by the state Medicaid agency) and obtaining administrative data on disability benefits, education, and earnings outcomes (with state agencies and SSA).

B. Evaluation rationale and overview

Evaluation rationale. No consistent vehicle to offer comprehensive case management services to the broader SSI population exists under current disability, education, and employment policies. However, the PROMISE demonstration tested a similar case management model for youth receiving SSI, with youth and families receiving the services only for the limited duration of the demonstration (at most, about four years). PROMISE did not provide services until the youth reached age 24, as proposed for the YFSN.

Measuring YFSN's results through an RCT would provide rigorous evidence of its effectiveness and support for broader adoption. An assessment of the full model—in which youth can access YFSN services until age 24—is necessary to test the implementation challenges in connecting with youth and families over such a long period and whether the intended outcomes materialize. The evaluation can identify facilitators, challenges, and outcomes related to both policy and practice and develop stronger approaches for future implementation, if warranted.

As with the evaluation option detailed in Chapter VIII for the Family Empowerment Model, the PROMISE five-year impact findings are forthcoming and could provide information relevant to the evaluation of YFSN. If PROMISE has positive impacts on education and employment outcomes, then an RCT evaluation of YFSN might not be necessary because PROMISE will have shown the advantages of providing holistic case management to youth receiving SSI and their families. Instead, a YFSN feasibility study could document approaches to offer and fund case management services over a longer duration than shown by PROMISE. If the PROMISE evaluation does not have consistent impacts on desired outcomes, then an RCT evaluation might still be warranted to test the efficacy of YFSN's long-term support for youth navigating the transition to young adulthood. Alternatively, an RCT could document whether the longer period of the YFSN intervention results in greater impacts than observed with PROMISE.

Evaluation design overview. The evaluation design for the YFSN involves a long-term (14-year) RCT. Participants would consist of youth receiving SSI and Medicaid who are age 14 at enrollment and their families. The lead agency would conduct outreach to these families based on lists of eligible youth

[5] SSI eligibility automatically qualifies a person for Medicaid in 40 states and the District of Columbia; in seven of these states, Medicaid requires a separate application (Rupp and Riley 2016). A separate eligibility determination for Medicaid is required in the remaining 10 states. Testing the YFSN model in a state where eligibility is automatic might facilitate identification of a broader group of youth receiving SSI through Medicaid program data and funding the intervention through a 1115 waiver.

obtained from the state Medicaid agency. A random assignment process would place those interested in services into either a treatment group with access to the YFSN or a control group with access to usual transition services. The lead organization would offer youth and their families services through the YFSN for up to 10 years (when participants turn age 24). The evaluation would follow youth during this period through surveys and administrative data to observe outcomes. Interim reports would document implementation and early impacts.

C. Evaluation design

Sample and recruitment. The YFSN intervention would offer services to youth receiving SSI and Medicaid and their families beginning when youth are age 14 and continuing through age 24. Youth's receipt of SSI and Medicaid would determine the eligibility for YFSN case management services; Medicaid eligibility is necessary because the demonstration would receive funding under an HHS 1115 waiver. The state Medicaid agency could provide lists of eligible SSI youth in the demonstration state to the lead organization, and that organization could then recruit youth and families primarily through the direct mailing of letters to eligible youth and their families, as well as other outreach methods (such as through schools, community organizations, and social media campaigns).[6] The demonstration would have no exclusion criteria; all youth receiving SSI and Medicaid could benefit from the intervention. Youth and their families who agree to participate would be randomly assigned to access comprehensive case management services from the YFSN (the treatment group) or the existing transition services (the control group).

The demonstration will need a mechanism for ensuring that youth continue to receive Medicaid in the event of SSI cessation. Youth could stop receiving SSI due to medical improvement, having sufficient earnings, or an unsuccessful age-18 redetermination, and they might lose Medicaid as a result. Youth could continue to qualify for Medicaid under other state-specific requirements, or the state Medicaid agency could offer a waiver for demonstration purposes that would allow continued Medicaid eligibility for participants through age 24.

Design. Over a 14-year period, a rigorous YFSN evaluation design would entail conducting an RCT and tracking enrollees for 12 years (2 years for enrollment and up to 10 years of observation after the final youth enrolls), plus additional time for final data collection and analysis. In the first year, the lead organization would develop the intervention, hire and train staff, and assemble partners. This process would include collectively building uniform practices, identifying appropriate data collection elements, and developing measures of success for the evaluation design. It could also involve a community of practice comprised of various stakeholders to offer input and feedback on the YFSN design and evaluation. Over the next 12 years, the evaluation would observe system implementation and youth outcomes and generate interim reports on the YFSN intervention. A final year would allow for a comprehensive examination of outcomes as youth complete the full complement of YFSN services.

Research questions. The evaluation would address the questions in Exhibit X.1 about the YFSN proposal using the data sources and analytic methods listed for each analysis type.

[6] An alternative to the state Medicaid agency providing lists of eligible youth is using lists obtained from SSA under a data use agreement with the agency.

Exhibit X.1. Evaluation questions, data sources, and analytic methods for an evaluation of the Youth and Family Systems Navigator proposal

Questions	Data sources	Analytic methods
Process analysis		
• Who participated in YFSN and how were participants different from other youth receiving SSI? For other youth receiving both SSI and Medicaid? • How were youth recruited? How were youth and families of color or from underserved communities recruited? • How was the YFSN program designed, implemented, and operated, and what factors contributed to the implementation experience? • What services did participants use and how often did they engage with the YFSN? How did those patterns change as participants aged? How did services vary for youth and families of color or from underserved communities? • How did YFSNs establish trust and rapport with participating youth and families? What were the challenges in the relationships that YFSNs had with youth and families across the duration of the intervention? • Was the intervention implemented as intended? • How did the education, workforce, and transition providers change in response to their involvement with the YFSNs? How did those changes affect the usual transition environment? • What program or policy changes did the lead organization and its partner organizations make in response to the intervention?	• Baseline application or survey data • Participant focus groups and structured interviews • Participant survey data • Program administrative data • Staff interviews • State or SSA administrative data	• Descriptive statistics • Qualitative analyses
Impact analysis		
• What was the impact of the YFSN program on intermediate outcomes, such as transition planning, service use, financial literacy, education, youth and family employment, and peer networks? • What was the impact of the policy change on the ultimate outcomes of employment and SSI benefit receipt? • Was the program more effective with some youth and families than others?	• Participant survey data • SSA administrative data • State administrative data	• Regression analyses
Benefit-cost analysis		
• Are YFSN's benefits large enough to justify its cost? • How did benefits and cost differ by stakeholder perspectives (such as participants, the lead organization, and local, state, and federal governments)?	• Participant survey data • Program administrative data • SSA administrative data • State administrative data	• Descriptive statistics • Regression analyses

SSA = Social Security Administration; SSI = Supplemental Security Income; YFSN = youth and family systems navigator.

Data collection. The evaluation would use data from the following sources.

- **Baseline data** collected through a survey or application form would obtain information—at the time of enrollment—on the characteristics of youth and families who enroll in the demonstration.
- **Program administrative data** from the lead agency (such as case notes, documents, and management information system extracts) would enable the evaluation to track program implementation, enrollment, staffing, and service delivery. This information would form the foundation for the process analyses and provide the detailed cost data necessary for the benefit-cost analysis.
- **Staff interviews** would include periodic interviews (in person or by phone) with YFSNs, administrators of the lead organization, and staff of partner organizations to obtain their perspectives on the program and its implementation.
- **Focus groups and structured interviews** with participating youth and their families would provide an opportunity to solicit their opinions on the program and the usual service environment from a select group of participants.
- **Participant surveys** would collect information on intermediate outcomes, such as employment, education, basic needs, financial literacy, and service use, for both treatment and control group members. This information would enable the evaluation to gauge short-, medium-, and long-term impacts and inform the benefit-cost analysis. The surveys could be offered at set intervals (such as at 5 and 10 years after enrollment) to enable an assessment of outcomes throughout the demonstration.
- **State administrative data** could include Medicaid data (to identify a sample for recruitment and monitor health expenditures), VR agency records (to track employment services and outcomes), unemployment insurance records (to observe quarterly earnings), and state education data (to track secondary and postsecondary education outcomes).
- **SSA administrative data** would provide accurate information on SSI-related characteristics and outcomes and annual earnings for all enrollees.

Analyses and reporting. The YFSN evaluation would benefit from periodic reporting of findings—including early assessments of implementation and impacts—to inform stakeholders interested in pursuing similar approaches to promoting the outcomes of youth with disabilities sooner than the release of the final evaluation results. The evaluation could include the following analyses and reports.

- A formative assessment would periodically document the intervention's implementation and track changes or improvements to the YFSN model. These descriptive analyses, which rely on program administrative data, could be conducted throughout the demonstration period.
- A process evaluation four years into the study (after enrollment concludes) would document who participated in the demonstration, early service use by youth and service practices by YFSNs, and training and engagement issues. It would include descriptive analyses of service data and qualitative analyses of interview and focus group data. Given the proposal's emphasis that YFSNs can be a catalyst for provider connections, the evaluation might consider using social network analysis techniques to map YFSN and participant connections to education, workforce, and transition providers and track the development of these networks over time.
- An early impact analysis five years after enrollment would contrast the treatment and control groups' outcomes (when youth are age 19) by analyzing participant survey, state administrative, and SSA

administrative data. It would also extend the process analysis findings during this period of program operations.

- A final impact analysis, conducted 10 years after enrollment, would assess treatment and control group outcomes when participants are age 24 and have finished with YFSN services.
- A benefit-cost analysis would compare the impacts documented in the final impact analysis with cost data (from the program administrative data and other sources). This comparison would document whether the YFSN benefits outweigh its costs, with consideration of different perspectives (such as the youth and family, the program, and SSA) and what timeline might be needed for any benefits to outweigh the cost.

D. Proposal refinements and other considerations

An advantage of the evaluation approach is that state and federal agencies could continue to track outcomes for many years after the demonstration end date. SSA, for example, could use its administrative data to track SSDI and SSI receipt, earnings, and VR agency service use for participants, as it has done for its Youth Transition Demonstration.

The Karhan and Golden proposal suggests that SSA could demonstrate and test a Youth Partnership Plus option under the Ticket to Work and Self-Sufficiency Program as a funding mechanism. This systems-change approach would allow incentive payments across providers in the Medicaid, workforce development, and VR system. For the evaluation proposed for YFSN, we intentionally focused the intervention and its evaluation on the case management aspect of the YSFN proposal and not the funding systems-change aspect. However, policymakers could consider and test that approach separately from—or in addition to—the YFSN model.

The Karhan and Golden proposal also offers an alternative to the HHS 1115 waiver for funding: using the demonstration authority under Workforce Innovation and Opportunity Act. YFSN implementation could occur under Title I American Job Centers or Title IV VR Agencies. This alternative offers several potential advantages, such as developing a staffing and service model that could be supported through existing systems, but requires overcoming administrative challenges with embedding the model into existing programmatic structures.

XI. Progressive Education: Early Intervention Strategy to Improve Postsecondary Outcomes for Youth with Disabilities

A. Proposal and demonstration description

Proposal description. Progressive Education is a model for public VR programs to increase participation in postsecondary education and training for students with disabilities (Howe et al. 2021). It builds on the premise that maximizing opportunities to experience postsecondary education while in high school leads students to greater participation in postsecondary education and training programs. The model offers a graduated series of activities for students to experience postsecondary education opportunities before leaving high school. These activities include career guest speakers, career fairs, career assessments, informational interviews, college campus tours, course audits, dual enrollment in college courses, career and technical training enrollment, apprenticeships, and postsecondary program enrollment. Participating youth attend these activities progressively from ages 14 to 25 (grade 9 until well after graduating from high school). The program also provides a menu of supports to promote youth's success in postsecondary education programming, such as coaching or tutoring, collaborations with school staff and agency partners, and access to assistive technology. The VR agency in Vermont is in the early stages of implementing this intervention, derived from its demonstration project, Linking Learning to Careers, which had positive impacts on education outcomes (Sevak et al. 2021).

Demonstration description. For evaluation purposes, the VR agency in Vermont would be well suited to implement a test of the intervention, given it developed the proposal and gained experience implementing a similar intervention through Linking Learning to Careers. Vermont VR staff would counsel participating youth and coordinate postsecondary education opportunities with other state institutions. To enable a rigorous evaluation, the agency could randomly select 6 of its 12 district offices to offer the program for a period of two years. In regions where the program is available, the agency would train VR counselors to expand the postsecondary education services they offer to their transition-age VR clients. Transition-age clients receiving services from staff at the offices where the intervention is not yet implemented would serve as a control group.

Lead and partner organizations. The VR agency in Vermont would implement Progressive Education in collaboration with other state agencies. The VR agency staff would establish or strengthen relationships with school staff, particularly those involved with developing IEPs and Section 504 plans — and so could provide students with information about suitable postsecondary education opportunities. The program also requires a strong working relationship with the state's two- and four-year college systems, including postsecondary education transition programs for students who qualify for developmental services.

B. Evaluation rationale and overview

Evaluation rationale. Progressive Education complements existing VR services but is a novel intervention that should be evaluated rigorously to understand whether it can lead to improved postsecondary education and other outcomes. Relevant outcomes include obtaining credentials, education attainment, exiting the VR program with employment, earnings, and SSI and SSDI receipt. Because the intervention would use resources that VR agencies could deploy elsewhere, VR stakeholders might value an assessment of the effectiveness of the intervention, including a benefit-cost analysis.

Evaluation overview. The demonstration and its evaluation build on the development of the intervention by the Vermont VR agency to assess its implementation and outcomes. To determine whether Progressive Education leads to improved postsecondary education, employment, and program outcomes, an evaluation would compare the outcomes of youth served by VR offices that offer Progressive Education activities with those of VR clients in control group offices in the same age range. To enable a rigorous evaluation, the Vermont VR agency would first introduce the intervention in six randomly selected offices for a two-year period. In addition to an impact analysis, the evaluation would include a process analysis to fully document and describe the intervention as it is deployed, including training needs and use of specific postsecondary education services. This analysis would also include the development of a fidelity tool to help VR staff understand how the approach extends usual VR services.

C. Evaluation design

Sample and recruitment. The target population for this intervention comprises students and youth with disabilities, including SSI recipients, between the ages of 14 and 25, regardless of their career goals and past education and employment experiences. The spread in ages enables the program to support both those who are beginning the process of determining postsecondary possibilities and those who may have already missed but could still benefit from such opportunities. This population includes individuals who dropped out of school and high school graduates whose postsecondary plans were never formulated or were unsuccessful. Youth participating in the intervention would consist of existing VR clients. Recruitment would not be necessary; the intervention group would consist of all youth in the six treatment group offices, and the comparison group would consist of VR clients in the six remaining offices. Although the intervention is not intended exclusively for SSI recipients, youth receiving SSI while participating in the intervention are an important part of the target group. The evaluation could conduct subgroup analyses for SSI recipients among participating youth.

Design. The evaluation is based on a clustered random assignment design in which half of the state's offices are randomly assigned to offer Progressive Education, with the remaining offices assigned to offer services as usual for a period of two years. The clustered random assignment would be conducted using strata comprised of at least two areas that are similar to one another in terms of their populations, economic environments, and relevant pre-intervention outcomes for youth receiving VR services. The evaluation would be conducted over a five-year period: two years during which Progressive Education activities are offered to VR clients in treatment group offices; two additional years to observe client service, program, education, and employment outcomes; and a fifth year to conduct evaluation activities.

Research questions. The evaluation would address the questions included in Exhibit XI.1 using the data sources and methods noted for each analysis type.

Data collection. The demonstration would require different types of data to address the evaluation questions.

- **Program administrative data** collected by the VR agency would track youth characteristics, youth use of postsecondary education services through Progressive Education, other VR agency services, and VR outcomes. The process and benefit-cost analyses would rely on this information.
- **Staff interviews**, conducted throughout implementation, would contribute to understanding Progressive Education implementation, changes over time, and partner relationships.

- **Focus groups and structured interviews** with youth and their families would provide insights into their use of VR services broadly and how they used the postsecondary education opportunities specifically.
- **State administrative data** would provide information on key outcomes for youth in the treatment and control groups. Data would include earnings from state unemployment insurance and education from state education databases (or alternatively, the National Student Clearinghouse).
- **SSA administrative data**, obtained by the VR agency, would identify youth use of SSI and SSDI throughout the demonstration.

Exhibit XI.1. Evaluation questions, data sources, and analytic methods for an evaluation of the Progressive Education proposal

Questions	Data sources	Analytic methods
Process analysis		
• What were the characteristics of youth in the offices assigned to offer Progressive Education activities? What were the characteristics of youth who used Progressive Education activities? What share received SSI? • How do youth compare to state and national youth with disabilities ages 14 to 25? • How did the intervention differ from usual VR practices? • Which of the Progressive Education activities did youth use? How did this use vary by youth characteristics, such as race and ethnicity? • What specific approaches did VR counselors use in offering services to youth? • How did implementation compare to the model? • How did VR counselors collaborate with secondary and postsecondary education staff to support clients?	• Program administrative data • Staff interviews • Youth and family focus groups and structured interviews	• Descriptive statistics • Qualitative analyses
Impact analysis		
• What was the impact of Progressive Employment on intermediate outcomes, such as VR service use and postsecondary education and training? • What was the impact of Progressive Employment on the ultimate outcomes of employment and SSI benefit receipt? • Was Progressive Employment more effective with some youth than others?	• Program administrative data • SSA administrative data • State administrative data	• Regression analyses
Benefit-cost analysis		
• Do the benefits of Progressive Education have the potential to offset its cost in the long term? • What employment impact is needed to offset program costs? How long would these employment impacts need to be sustained?	• Program administrative data • SSA administrative data • State administrative data	• Descriptive statistics • Regression analyses

SSA = Social Security Administration; SSI = Supplemental Security Income; VR = vocational rehabilitation.

Analyses and reporting. The Progressive Education evaluation would observe short- and medium-term service, education, employment, and program outcomes. The evaluation could include the following analyses and reports.

- Formative assessments, using program administrative data, would document intervention implementation and changes to the Progressive Education model.
- Early process analyses one year into the study would document the involvement of youth (including service use), VR counselor practices, and partner engagement. It would include descriptive analyses of service data and qualitative analyses of interview and focus group data.
- Early impact analyses conducted three years after implementation start would contrast the outcomes of the treatment and control groups. These analyses would draw on program and state administrative data.
- A final impact analysis, conducted five years after implementation start, would extend the early impact analyses to assess treatment and control group outcomes for up to four years using administrative data.
- A benefit-cost analysis would assess whether differences in education, earnings, and benefit levels between treatment and control groups, extrapolated into various points in the future, could justify the cost of Progressive Education, and if not, what impacts would be needed to offset program costs.

D. Proposal refinements and other considerations

Although the proposed intervention is expected to lead to improved employment outcomes in the long run, short- and medium-term employment impacts could be negative if youth in the treatment group are more likely to enroll in postsecondary education instead of taking a job, compared to the control group. Depending on the ages of participating youth, an evaluation may not detect positive employment impacts within the proposed five-year time frame. Instead, evaluators could consider combining enrollment in postsecondary education and employment into a single outcome and estimate the impact of the intervention on this combined desirable outcome.

XII. Transition Linkage Tool: A System Approach to Enhance Post-School Employment Outcomes

A. Proposal and demonstration description

Proposal description. The Transition Linkage Tool is a joint effort across government agencies within a state to link their technology systems and data recording procedures for secondary educators and transition service providers (Gingerich and Crane 2021). The goal of these linkages is a centralized, accessible way of tracking key information on high school students with disabilities as they prepare for transition out of high school. Information gathered by this tool could include (1) service referrals, application dates, and eligibility determinations; (2) services that students have been offered or used from their schools, VR agencies, intellectual and developmental disabilities agencies, and state workforce investment agencies; (3) students' high school graduation, post-school enrollment, and employment outcomes; and (4) the student's or family's consent to data sharing between agencies.

This tool would link information across schools and agencies in such a way that staff could better coordinate services for students across agencies, increase student participation in programs and services, and ultimately increase students' likelihood of finding employment after graduating high school or enrolling in postsecondary education or training.

Demonstration description. This proposal was developed and is being piloted in Maryland, where it grew out of the Maryland PROMISE initiative. There, the system is in the early stages of rollout in eight school districts as a prelude to state-wide rollout. This ongoing effort provides an opportunity for retrospective analysis of the linkage tool's implementation, student participation trends, and costs and benefits—focusing both on the state agencies that developed the tool and on districts that have begun using it. Retrospective and prospective impact evaluation is also possible based on this staggered rollout by examining either changes to date in locations that have already adopted the tool or future changes in outcomes at locations adopting the tool at a later date. Moreover, the power of the demonstration to produce useful research results could be improved by shifting Maryland's future rollout schedule to follow a random assignment approach and proceed at the school rather than district level.

Because the tool links data systems across agencies, it must be developed and deployed at the state level, or possibly within a large city with independent agencies. Because institutional structures and arrangements differ from state to state, each location would need to design its own version of the linkage tool. Any evaluation of a location's implementation would follow a similar staggered rollout model, given the linkage tool is designed for comprehensive statewide use.

Lead and partner organizations. This proposal involves two phases: development and deployment. During the development phase, state agencies collaborate on institutional arrangements and build the software and other infrastructure needed to store and share data. During the deployment phase, school districts, schools, and other agency front-line staff receive training and begin to use the tool in their interactions with students and clients. In Maryland, the state's Department of Disabilities and Department of Education spearheaded the development of the tool, with system design support from the University of Maryland. The cooperation of these agencies—and others who participate in data sharing through the linkage tool (such as state VR agencies or developmental disability agencies)—would be necessary to facilitate staff interviews and sharing of administrative data for evaluation. Other important partners would include districts and schools, whether they were early adopters of the linkage tool or were in the comparison group, to provide data on student participation and access to staff, students, and families for

interviews and surveys. In addition, adjusting the rollout schedule in Maryland (as described below) would require significant cooperation both from participating state agencies' leadership and from school and district leadership.

B. Evaluation rationale and overview

Evaluation rationale. Maryland developed the proposed linkage tool in response to two problems: (1) uncoordinated services for youth and families and (2) gaps in knowledge and use of available resources by families due to a lack of awareness about services for which they were eligible. An evaluation of the tool would quantify the benefits of a more integrated approach to transition service coordination and provision. By providing information about strong and weak spots in the transition process, the tool could also be used for program improvement and to maintain the documentation needed to meet federal requirements around equal access to services.

For these reasons, the evaluation would be of interest to state agency administrators and staff involved in youth transition to young adulthood. Understanding the potential benefits of using the linkage tool and the associated costs could prove useful for states interested in linking state data systems to track the service use and outcomes of transition-age youth with disabilities.

Evaluation design overview. A process analysis of Maryland's experience in designing the linkage tool and rolling it out in schools would be straightforward and low-cost. It would provide useful insights for other localities in Maryland rolling out the tool, as well as for other states interested in taking up the proposal. This portion of the evaluation could be accomplished through document review and retrospective interviews with state, district, and school staff.

Maryland's plan for a staggered rollout of the program across districts offers a ready-made comparison group for a quasi-experimental difference-in-differences evaluation of the program's impacts on districts, schools, and youth and for calculating program benefits. This approach would compare outcomes for districts, schools, and graduates in locations that already have begun using the tool against (1) outcomes in other locations and (2) outcomes in the same locations from before the tool's adoption. This analysis could be implemented going forward and possibly also include retrospective analysis of the rollout so far using state administrative data. Depending on data availability and the ability of researchers to link student data to other records, this evaluation design can draw on previous adoptions of the linkage tool by the eight Maryland districts that have implemented the tool and/or future adoptions by remaining districts. (Additional discussion of retrospective data availability versus prospective data collection follows in the data collection section below.)

The non-random ongoing rollout schedule across districts in Maryland reduces the confidence with which an evaluator could attribute changes in outcomes to the adoption of the linkage tool. (For instance, if districts were prioritized for rollout because they had simultaneously undertaken other program improvement efforts besides the linkage tool, the changes might actually be due to those initiatives.) However, the rollout process could be adjusted for the remaining Maryland schools (or for future implementation in other states) to enable evaluators to directly attribute impacts to the intervention by pairing districts or schools based on similar observable characteristics and randomly choosing one early- and one late-adopting school from each pair.

C. Evaluation design

Sample and recruitment. High school students with IEPs are the focus of this intervention. In the 2019–2020 school year, about 2.3 million youth ages 14 to 21 nationwide were served under the Individuals with Disabilities Education Act and were eligible for IEPs—about one in seven students in grades 9 to 12. During the same period, Maryland's public schools enrolled about 50,000 students with disabilities ages 12 to 21. Through existing mechanisms, schools already identify these students. The linkage tool requires students and families to volunteer to have their data included in the linkage tool by providing consent for the students' program participation data to be linked across state agencies. However, depending on data availability and collection methods, it would be possible to include student outcomes in an evaluation even if they opted against participating in the linkage tool or attended a school that had not yet adopted the tool.

Maryland has 24 public school districts containing 181 high schools; 8 districts have thus far adopted the linkage tool. With adoption proceeding district-by-district, even with about 50,000 affected students, a sample of this size would likely be able to detect only very large changes to key outcomes (such as transition program participation and post-graduation employment).

Under a district-level design, an employment outcome with a baseline prevalence of 33 percent or 67 percent would have an MDI of 9.2 percentage points.[7] One approach to improving the power of the evaluation to detect impacts would be to switch the level of adoption in remaining districts to the level of an individual high school, as described in the design section below. This strategy would reduce the MDI significantly, but it would require significant cooperation from state and local education officials and could have added complications stemming from the need to coordinate rollout across a large number of individual schools rather than a small number of districts.

Design. Maryland has already developed its own linkage tool and is piloting it in selected school districts in preparation for a statewide rollout. Two small school districts began using the tool in the spring semester of the 2018–2019 school year, and six others began participating in July 2021.

Data collection to address questions about agency participation, implementation, and benefit-cost analysis can begin immediately through interviews and document review among participating agencies and state staff. Likewise, evaluators can interview teachers and administrators in districts that have already begun using the tool and expand to other districts as they join the pilot. Meeting with educators and state agency staff throughout the initial implementation year can allow stakeholders to troubleshoot issues, develop best practices, and provide insights into the implementation process to contextualize other evaluation findings and simplify future rollouts.

[7] The MDI assumes a two-sided *t*-test of a non-random difference-in-differences design with a power of 80 percent and *p*-value of 0.10, a treatment rate of 33 percent, and generous assumptions about other sources of variance (an intra-class correlation of 0.02, and 25 percent of all variance—within- and between-group outcome variance, variance in treatment condition assignment, and variance in outcomes due to persistent factors—explained by covariates). The baseline outcome rate is similar to the nationwide employment rate at age 19 for individuals with disabilities based on the 2018 American Community Survey (32 percent) and the postsecondary education enrollment rate within 8 years of graduation based on the National Transition Longitudinal Survey 2 (60 percent). For an outcome with a 25 percent (or 75 percent) baseline prevalence, the MDI would be 9.6 percentage points, and for an outcome with 10 percent (or 90 percent) baseline prevalence, the MDI would be 6.7 percentage points.

After each district's initial year using the linkage tool, their students' program participation and other outcomes can be observed. Many impact outcomes can be measured either at school year's end or after a relatively brief follow-up period (such as three or six months after graduation for the first cohort of graduating seniors, in the case of employment and postsecondary enrollment outcomes).

The tool's impact on any outcomes collected over time at the school- or district-level can be assessed with a difference-in-differences methodology. The same approach would work for any student-level outcomes in which an individual's outcome data can be retroactively linked to their high school and graduation year. (In such a case, outcomes for students at high schools adopting the tool would be compared with those for students at other schools, as well as against outcomes for students from the same school prior to the tool's adoption.) If retrospective linking is not possible, difference-in-differences analysis could still be used prospectively by collecting baseline and post-implementation data for students in all schools and observing changes that occur in remaining districts as they adopt the linkage tool.

To improve the power of the evaluation to detect impacts of the linkage tool and the ability of evaluators to attribute those changes to the tool, rollout in Maryland's remaining districts could proceed at the school level using random assignment. With this approach, schools would be matched on their observable characteristics (such as local characteristics and economic conditions, prevalence of students with disabilities, and past performance on targeted outcomes) and one school chosen at random from each pair to adopt the tool early, with the second school starting one or more years later.[8] If two-thirds of Maryland's schools and students remain in districts that have not adopted the linkage tool, using this strategy would improve the detectable impact threshold by more than a factor of two, relative to a district-level rollout.[9]

This approach implies an evaluation timeline of up to six years—one year for set-up of the implementation and planning, up to four additional years for the intervention to run in early-adopting schools with interim reports each year, and a final year for follow-up and final evaluation. Researchers have a critical choice about the timing of the implementation and evaluation. If they choose to delay linkage tool rollout in the late-adopting group by four or more years, they would have at least one cohort of students for whom they could estimate the tool's cumulative impacts across a full standard high school career, comparing treated students to similar students whose schools never had access to the tool. However, in that length of time, outside trends may have made weakened the direct comparability between the paired treatment and control schools. Alternatively, researchers could delay rollout in the late-adopting group by less than four years. This shorter delay would likely maintain better comparability between treatment and control schools and speed the availability of final evaluation results, but the impact estimates would rely more heavily on comparisons between students who differed only in their degrees of exposure to the linkage tool (for example, two versus three years of exposure), which may understate the tool's impact.

[8] Other designs might also be possible, such as choosing staff at some schools to receive earlier training or heavier promotion and encouragement to use the linkage tool than staff at other schools.

[9] Under the same assumptions noted before—an evaluation with 33,333 students in 120 schools with 50 percent assigned to the early-adopting group, the MDI for an outcome with 33 percent (25 percent/10 percent) baseline prevalence would be 4.0 (3.7/ 2.6) percentage points, a sizeable change in what the evaluation can detect and a more realistic expectation of the potential impact. See footnote 7 for baseline assumptions.

Exhibit XII.1. Evaluation questions, data sources, and analytic methods for an evaluation of the Transition Linkage Tool proposal

Questions	Data sources	Analytic methods
Process analysis		
• How did students and families react to requests for cross-agency data-sharing consent? What share declined to give consent? • How did the characteristics of consenting students compare to those of the full population of students with disabilities? Were youth and families of color or from underserved communities more or less likely to provide consent (and if so, why)? • How did students in the treatment group districts compare with students in the control group districts? • Which agencies used the linkage tool? Which did not, and why? What other data would be useful to include in the linkage tool, and why did the data owners choose not to participate? • What personnel and resources were needed to design and implement the linkage tool? What unique conditions in the state facilitated or hindered linkage tool development? What key issues arose, and how were they resolved? • What was the final design of the linkage tool, and did it operate as intended? • What training did staff need to use the tool? • Did school and agency staff use the tool? For what purposes did they use it? Why did some staff not use the tool? • How did staff's interaction with students and families change as a result of using the tool?	• Linkage tool data • School administrative data • Staff surveys • Structured interviews or focus groups with students and families • Structured interviews with staff	• Qualitative analyses • Descriptive statistics
Impact analysis		
• What was the impact of the tool for students' intermediate outcomes, such as service awareness, service use, linkages to adult services, and educational attainment? • What was the impact of the tool for the system's intermediate outcomes, such as data sharing, service coordination, and compliance with federal law? • What was the impact of the tool on students' ultimate outcomes related to employment? • What was the impact of the tool on the system's ultimate outcomes related to interagency collaboration and services? • Was the tool more effective for some students and families than others? • How did agencies and staff change in response to their use of the tool in terms of services, outcomes, and collaboration?	• School and state administrative data • Survey data	• Descriptive statistics • Regression analyses

Questions	Data sources	Analytic methods
Benefit-cost analysis		
• Are the linkage tool's benefits large enough to justify its costs? • How did benefits and cost differ by stakeholder perspectives (such as students, schools, and participating organizations?	• Linkage tool data • School and state administrative data • Survey data	• Descriptive statistics • Regression analyses

Research questions. The evaluation would address the questions in Exhibit XII.1 using the data sources and analytic methods listed for each analysis type.

Data collection. The evaluation would use data from the following sources.

- **Structured interviews** with members of the state-level linkage tool working group, linkage tool designers, selected administrators and staff in school districts and state agency offices, and educators would be used to understand their perceptions and use of the tool.
- **Structured interviews and focus groups** with students and families could also be used to understand student and family choices about granting data-sharing consent and how their transition experiences have been affected by their schools' adoption of the tool.
- **School administrative data** from schools, state agencies, and central state databases would show district- or school-level outcomes that have been collected consistently over time (such as graduation rates).
- **State administrative data** could be used to measure individual-, school-, and district-level outcomes. The evaluation requires both pre- and post-intervention data for its difference-in-differences design. The evaluator's success will be contingent on its ability to link data from sources such as the unemployment insurance program, VR agency, and postsecondary education institutions to individual-level secondary school data.
- **Linkage tool data** would document the data collected and staff access of the tool and its reports, which would inform the understanding of the tool's implementation.
- **Survey data** could be an alternative source of student-level outcomes. Student outcomes would need to be collected via survey in both early- and late-adopting districts. Some analysis of retrospective survey data might shed light on changes in the early-adopting districts (such as if past graduates from the full sample of schools could be surveyed about some easily recalled outcome, such as a first job).

Analysis and reporting. Data collection and analysis covering the implementation process and the measurement of the implementation's direct costs would ideally begin quickly, as some processes to be studied have already taken place. These results would ideally be reported soon after to inform future implementation efforts in Maryland and elsewhere.

In each year, participation data could be analyzed and reported for ongoing monitoring of the students who consent—information that the state could use to better promote the tool. Analysis of high-level trends in overall program participation would also provide early indications of how state agencies' funding and personnel needs might change in response to increased demand for services.

Annual analyses of system and student-level impacts each year after random assignment of the remaining schools would enable researchers to observe implementation and impact results from use of the linkage

tool throughout students' high school experiences. These analyses would build on the prior year's analysis and show the effects of increased exposure to the tool.

A final report in the sixth year would provide four-year process, impact, and benefit-cost results. It would draw on the full set of data and analyses by showing the impact of exposure to the tool from one to four years. A critical aspect of this evaluation would be whether use of the tool and its impacts varied across student characteristics.

D. Proposal refinements and other considerations

Although Maryland's experience can offer important lessons and model guidance, other states or large cities implementing their own transition linkage tools will have different and possibly unique factors to consider. For instance, each state must develop its own interagency data sharing and systems integration approach. There is not likely to be an off-the-shelf approach to the arrangements necessary for the linkage tool. Some states may choose to involve a different set of agencies than will contribute to Maryland's tool. Because each state that implements the linkage tool will have its own idiosyncrasies, it will likely be advantageous to stagger the tool's rollout in their implementations; this process would enable them to identify state-specific implementation issues and assess their tool's effectiveness. Rollouts in other states could be staggered as proposed for Maryland, perhaps randomizing the rollout phase-in timing across schools, districts, counties, or other jurisdictions; this process also would enable them to generate causal impact estimates if further impact analyses were desired—perhaps to extend the impact evidence to settings outside of Maryland—and the necessary outcomes linkages could be achieved.

Increasing the number of state agencies that participate will increase its value to users but could also add to its development challenges. States that take up this proposal must consider which agencies are must-haves and which might be included in later implementation.

Maryland stakeholders are interested in identifying youth receiving SSI for inclusion in the group served by the transition linkage tool. However, none of the participating agencies in Maryland have access to SSI program data. Some state agencies—such as Medicaid agencies—have access to these data and might be able to share them for the purpose of administering their programs, but whether the transition linkage tool is an allowable use could vary from state to state based on legal regulations and data agreements.

This page has been left blank for double-sided copying.

XIII. Transition to Economic Self-Sufficiency (TESS) Scholarships for Youth and Young Adults with Significant Disabilities

A. Proposal and demonstration description

Proposal description. Transition to Economic Self-Sufficiency (TESS) Scholarships would provide reliable income support to low-income youth with significant disabilities for an extended period after high school (Stapleton et al. 2021). Upon enrolling after high school completion or shortly thereafter, TESS scholars would receive cash support (approximately $10,000 per year) through age 30 as long as they actively pursue a career. The scholarship would be paid into an Achieving a Better Life Experience (ABLE) account, which is not counted for purposes of determining SSI eligibility or calculating SSI benefit amounts; SSA excludes the first $100,000 from the means test. Scholars could only use the funds for career-related expenses. Additional supports include professional career coaching, benefit and financial counseling, vocational services, financial aid for postsecondary education, health insurance, and long-term home-based support services.

The intervention aims to support youth and young adults in overcoming the barriers and disincentives that impede sustained employment and successful careers by offering income support not tied to earnings along with other opportunities for human capital development. The primary goals of the interventions are for scholars to establish successful careers by age 30 and to reduce their reliance on SSI and other public programs.

Demonstration description. For evaluation purposes, TESS would be implemented in a single state using funding through a private–public partnership. It would combine existing state programs (such as a state VR agency to deliver transition services and coordinate services across agencies) with private funding. Career coaches, located within the VR agency, would coordinate the activities of others involved in executing the scholars' plans and monitor their progress toward economic independence and compliance with the scholarship terms. The private sector partner, such as a nonprofit organization, would fundraise and administer the scholarship program. After an initial start-up year to develop partnerships and identify funding, a single cohort of youth receiving SSI ages 18 to 24 would enroll into the demonstration over a one-year period and have access to TESS services and supports, including the scholarship, for a 10-year period (or until enrollees are age 30) as long as they remain successfully engaged in the program.

Lead and partner organizations. A state VR agency is well-positioned to lead a TESS demonstration. It has the ability to identify youth interested in enrolling in the program, and its employment and training services are a crucial component of the intervention. It could also employ career coaches to assist youth in achieving their career goals. Partner agencies to support TESS include a private sector partner, such as a center for independent living, to oversee the scholarship program; local high schools, the state education agency, and postsecondary education institutions; business organizations and employers; organizations that offer benefits and financial counseling; the state Medicaid agency; an institution administering ABLE accounts; and other nonprofit organizations, such as foundations, to contribute to fundraising for the scholarships.

B. Evaluation rationale and overview

Evaluation rationale. TESS being a major departure from the current policy environment and having a sizeable cost suggest that evidence meeting the highest standard of rigor might be needed to justify large-scale implementation. In addition, many of its components are untested, and it thus requires an understanding of how to implement the intervention and an assessment of whether it can produce expected outcomes. Although existing interventions offer career coaching and other services, the scholarship that scholars would receive unconditionally on earnings is a key part of the intervention, and no evidence exists on how such a scholarship would affect outcomes. Before widespread adoption of the intervention, stakeholders would need evidence of its effectiveness.

Evaluation design overview. Because TESS is untested and requires significant infrastructure development, a pilot test is warranted. The pilot test, conducted over a 12-year period, would follow a small group of scholars throughout their use of their scholarships (that is, through age 30). The pilot would use an RCT, likely with a small number of enrollees and a lottery to randomly assign youth who sign up for TESS into a treatment group that would be eligible to participate in TESS or a control group that would have access to usual services. An RCT design, even on a small scale, would provide strong evidence on the effectiveness of TESS in supporting youth in their transitions through young adulthood. The remainder of this chapter describes features of a pilot test as a first step along the proposed continuum.

C. Evaluation design

Sample and recruitment. The larger population targeted by the proposal includes youth receiving SSI or youth who are likely eligible to receive SSI based on their disability conditions and their income and assets. Applicants would need to satisfy the eligibility criteria laid out in the TESS proposal or some variant: (1) be ages 18 to 24; (2) exited high school; (3) either receive SSI or be likely to meet the disability, income, and asset requirements for SSI; and (4) have a desire to pursue a career. Study participants would typically apply for a TESS Scholarship during their last year of high school and begin the program upon leaving high school. Suitable candidates (such as students with significant disabilities) could be identified by guidance counselors, special education staff, or VR transition counselors and encouraged to apply for a TESS Scholarship. Those who enroll would be randomly assigned to have access to TESS or usual services. All enrolled youth would need to consent to have their administrative data used for evaluation purposes, regardless of being randomly assigned to the treatment group.

The number of participants and, thus, the recruitment pool would depend on available scholarship funds; full program funding would require $10,000 in scholarship funds for each scholar per year, plus additional funds for the career counseling staff and support services that are not offered through existing programs. For example, a demonstration sample size of at least 200 youth, of whom 100 are in the treatment group, would require up to $10 million in scholarship funds over a 10-year period.

Design. Given the duration of the intervention (which scholars could use as early as age 18 and through age 30), a comprehensive evaluation must span that same period and track outcomes for 5 to 10 years after youth complete the program. Waiting more than a decade for results to make decisions related to the proposal is not practical, nor is implementing a costly, large-scale effort without any evidence of effectiveness. A pilot test, with one year to develop the intervention and assemble partners, one year for scholar recruitment, and 10 years to offer the scholarships and observe scholar service use, education,

benefit receipt, and employment outcomes, could provide incremental evidence throughout the demonstration period on whether TESS works as intended and thus justify further investigation.

Research questions. The evaluation would address the questions included in Exhibit XIII.1 using the data sources and analytic methods listed.

Exhibit XIII.1. Evaluation questions, data sources, and analytic methods for an evaluation of the Transition to Economic Self-Sufficiency Scholarships proposal

Questions	Data sources	Analytic methods
Process analysis		
• How many and what were the characteristics of youth who applied for a TESS Scholarship? How many applicants were youth of color or from underserved communities? • How did participants compare to all youth receiving SSI ages 18 to 24? • Did any applicants decline the scholarship offer? If so, what were their reasons? • How was the TESS program designed, implemented, and operated, and what factors contributed to the implementation experience? • Was the intervention implemented as intended? • Which public and private partners participated in the intervention? What were their respective roles? How were partners identified? What was the nature of the relationships among the partner organizations? • How much funding for the scholarships was obtained from private versus public partners? What were the challenges in obtaining or maintaining funding? • What challenges did scholars encounter in satisfying the ongoing scholarship requirements? • How many accepted scholars did not complete the program? What were the reasons? How did they compare to those who completed the program? • For what expenses did scholars use their scholarships? • What services did scholars use and how often did they engage with the career coach? How did those patterns change as scholars aged? How did services vary for scholars of color or from underserved communities?	• Baseline application data • Participant focus groups and structured interviews • Participant survey data • Program administrative data • Staff interviews	• Descriptive statistics • Qualitative analyses
Impact analysis		
• What was the impact of TESS on intermediate outcomes, such as service use, career expectations, employment, postsecondary education, and health coverage? • What was the impact of the policy change on the ultimate outcomes of employment and SSI benefit receipt? • Was the program more effective with some scholars than others?	• Participant survey data • SSA administrative data • State administrative data	• Regression analyses

Questions	Data sources	Analytic methods
Benefit-cost analysis		
• Do the benefits of TESS have the potential to offset its cost in the long term? • What employment impact is needed by the time scholars turn age 30 to offset program costs at the end program? Within 10 years of program end? 20 years?	• Participant survey data • Program administrative data • SSA administrative data • State administrative data	• Descriptive statistics • Regression analyses

SSA = Social Security Administration; SSI = Supplemental Security Income; TESS = Transition to Economic Self-Sufficiency.

Data collection. The pilot test would require different types of data to address the evaluation questions.

- **Baseline data** collected through an application form would collect information on applicant characteristics.
- **Program administrative data** collected by the lead agency would track recruitment efforts, applications, acceptances, withdrawals, scholarship funds, and participant service use, perhaps in conjunction with VR agency administrative data systems if a VR agency is overseeing TESS implementation. The process and benefit-cost analyses would rely on this information.
- **Staff interviews**, conducted throughout implementation, would contribute to understanding TESS implementation, changes over time, and partner relationships.
- **Focus groups and structured interviews** with participating scholars and their families would provide an opportunity to solicit their opinions on the program and how they used their scholarships.
- **Participant surveys** would inform the intermediate and ultimate outcomes relatively early, collecting information from scholars and control group members about early career pathways (job level, occupation, benefits, and tenure with an employer), postsecondary education and training, and other outcomes. The surveys could be offered at set intervals (such as at 2, 5, and 10 years after enrollment) to enable an assessment of outcomes throughout the demonstration.
- **State administrative data** would provide information on key outcomes for both scholars and control group members. Data would include earnings from state unemployment insurance or the VR program, services from the VR program, and education from state education databases or the NSC.
- **SSA administrative data**, provided through a partnership with SSA, could provide accurate information on SSA program enrollment and benefit data for all scholars and control group members.

Analyses and reporting. The TESS evaluation could provide needed information on the mechanics of implementation and on the early career pathways of scholars. It would observe both early and long-term outcomes, as well as the proposal's ultimate outcome of stable employment and earnings that exceed SSI and SSDI eligibility thresholds enabling scholars to achieve self-sufficiency by age 30. The evaluation could include the following analyses and reports.

- Formative assessments, using program administrative data, would document intervention implementation and changes to the TESS model.
- Early process analyses two and five years into the study would document the enrollment and involvement of participants (including early service use), career counselor practices, scholarship management, and partner engagement. It would include descriptive analyses of service data and qualitative analyses of interview and focus group data.
- Early impact analyses conducted at two and five years after enrollment would contrast the outcomes of scholars with those of the control group. These analyses would draw on state and SSA administrative data to document service use, education, employment, and public program participation outcomes. They would also document additional intermediate outcomes for participants using survey data. Stakeholders might consider whether the evidence at these early periods warrant expanding TESS to other locations, rather than waiting for the pilot evaluation to conclude.
- A final impact analysis, conducted 10 years after enrollment, would assess scholar and control group outcomes from administrative data and additional participant outcomes from survey data.
- A benefit-cost analysis would assess whether impacts for education, earnings, and benefit levels—extrapolated into various points in the future—could justify the cost of TESS, and if not, what impacts would be needed to offset program costs.

D. Proposal refinements and other considerations

Sponsors of a TESS pilot could consider two refinements to testing the intervention. First, as proposed, TESS includes scholarship payments to scholars of about $10,000 per year through age 30. Alternative considerations for the scholarships could involve a reduction in either the annual payment amount or the duration of the scholarship. Second, policymakers might consider testing two variations of TESS: one that includes the scholarship and supporting services, and one that includes only the supporting services without the scholarship. An RCT design could randomize eligible applicants into three arms: those receiving the scholarship and additional services (representing the full intervention), additional services only (representing a partial intervention), and no services or scholarship (the control group). Such a test could provide evidence on the effectiveness of the income supports in navigating the transition to adulthood.

If random assignment is not feasible, the evaluation could use a matched set of youth with which to compare the service use, employment, and education outcomes of TESS participants. An evaluator could identify a comparison group matched to TESS scholars on age, sex, disability type, SSI receipt, and other relevant characteristics available in state VR agency administrative data. The service, education, and employment outcomes of that group could be tracked alongside those of scholars using the same administrative data sources. The differences in outcomes would reflect a counterfactual of those interested in and using VR services. If the comparison group is a good match with TESS participants (that is, the TESS and comparison group members have similar observed characteristics at baseline and the study lacks confounding factors), the evidence generated from the pilot study would meet CLEAR's moderate rating for causal evidence. This approach has the advantages of showing the value of the scholarship on top of existing VR services and relying on a rich administrative data source that can be tied to state unemployment insurance and NSC data to assess earnings and postsecondary education. The approach would be limited, however, in its ability to contrast the experiences of scholars with the comparison group on many intermediate outcomes, such as career expectations. Alternative data sources for matching, such

as through national VR records or SSI administrative records, have limitations in terms of the data available and lag times, but they could be used either as an alternative or supplement to state VR agency data.

References

Anderson, C., E. Hartman, and D.J. Ralston. "The Family Empowerment Model: Improving Employment for Youth Receiving Supplemental Security Income (SSI)." Report prepared for the U.S. Department of Labor, Office of Disability Employment Policy, April 2021. Available at https://www.dol.gov/sites/dolgov/files/ODEP/pdf/SSI_Youth_Anderson_Final_Proposal.pdf.

Centers for Disease Control and Prevention. "Program Evaluation for Public Health Programs: A Self-Study Guide." Atlanta, GA: Centers for Disease Control and Prevention, 2011. Available at https://www.cdc.gov/eval/guide/index.htm. Accessed September 17, 2021.

Clearinghouse for Labor Evaluation and Research. "CLEAR Causal Evidence Guidelines, Version 2.1." 2019. Available at https://clear.dol.gov/reference-documents/causal-evidence-guidelines-version-21. Accessed September 19, 2021.

Czajka, J.L., A. Patnaik, and M. Negoita. "Data on Earnings: A Review of Resources for Research." Final report for the U.S. Department of Labor. 2018. Available at www.dol.gov/sites/dolgov/files/OASP/legacy/files/Data-on-Earnings-Report.pdf. Accessed September 21, 2021.

Dean, David, John Pepper, Robert Schmidt, and Steven Stern. 2019. "The Effects of Youth Transition Programs on Labor Market Outcomes of Youth with Disabilities." *Economics of Education Review* 68, no. 2019: 68–88.

Francis, G.L., and J.M. Gross. "The Effectiveness of the Family Employment Awareness Training Program: Implications for Practitioners." Paper presented at the Council for Exceptional Children Convention, Boston, MA, April 19–22, 2017.

Francis, G., J.M. Gross, A.P. Turnbull, and H. Rutherford Turnbull. "The Family Employment Awareness Training (FEAT) Program in Kansas: A Mixed Method Follow-Up." *Journal of Vocational Rehabilitation*, vol. 39, no. 3, 2013, pp. 167–181. https://doi.org/10.3233/JVR-130652.

Gingerich, J.A., and K. Crane. "Transition Linkage Tool: A System Approach to Enhance Post-School Employment Outcomes." Report prepared for the U.S. Department of Labor, Office of Disability Employment Policy, April 2021. Available at https://www.dol.gov/sites/dolgov/files/ODEP/pdf/SSI_Youth_Gingerich_Final_Proposal.pdf.

Gross, J.M., and G.L. Francis. "Short-Term and One-Year Outcomes for Family Employment Awareness Training Attendees." Paper presented at the TASH Conference, St. Louis, MO, November 30–December 2, 2016.

Gross, J., G. Francis, and S. Gage. "Family Employment Awareness Training (FEAT): A Research-Based Program for Promoting High Expectations for Employment and Knowledge of Resources." Report prepared for the U.S. Department of Labor, Office of Disability Employment Policy, April 2021. Available at https://www.dol.gov/sites/dolgov/files/ODEP/pdf/SSI_Youth_Gross_Final_Proposal.pdf.

Hippolitus, P. "Employment Empowerment: A Foundational Intervention for Youth with Disabilities to Build Competitive Employment Skills." Report prepared for the U.S. Department of Labor, Office of Disability Employment Policy, April 2021. Available at https://www.dol.gov/sites/dolgov/files/ODEP/pdf/SSI_Youth_Hippolitus_Final_Proposal.pdf.

Hollenbeck, K. "Demonstrating the Effectiveness of Short-Term Career and Technical Training in a Residential Setting for Transition-Age Youth with Disabilities." Report prepared for the U.S. Department of Labor, Office of Disability Employment Policy, April 2021. Available at https://www.dol.gov/sites/dolgov/files/ODEP/pdf/SSI_Youth_Hollenbeck_Final_Proposal.pdf.

Honeycutt, T., D. Wittenburg, K. Crane, M. Levere, R. Luecking, and D. Stapleton. "SSI Youth Formative Research Project: Considerations for Identifying Promising and Testable Interventions." Washington, DC: Mathematica Policy Research, 2018.

Howe, T., C. McCarthy, J. Smith, and R. Tulikangas. "Progressive Education: Early Intervention Strategy to Improve Postsecondary Outcomes for Youth with Disabilities." Report prepared for the U.S. Department of Labor, Office of Disability Employment Policy, April 2021. Available at https://www.dol.gov/sites/dolgov/files/ODEP/pdf/SSI_Youth_Howe_Final_Proposal.pdf.

Karhan, A.J., and T.P. Golden. "Policy Considerations for Implementing Youth and Family Case Management Strategies Across Systems." Report prepared for the U.S. Department of Labor, Office of Disability Employment Policy, April 2021. Available at https://www.dol.gov/sites/dolgov/files/ODEP/pdf/SSI_Youth_Karhan_Final_Proposal.pdf.

Kuehn, D. "Building an Apprenticeship Infrastructure for Youth Receiving SSI." Report prepared for the U.S. Department of Labor, Office of Disability Employment Policy, April 2021. Available at https://www.dol.gov/sites/dolgov/files/ODEP/pdf/SSI_Youth_Kuehn_Final_Proposal.pdf.

Larson, S.A., and J. Geyer. "Delaying Application of SSI's Substantial Gainful Activity Eligibility Criterion from Age 18 to 22." Report prepared for the U.S. Department of Labor, Office of Disability Employment Policy, April 2021. Available at https://www.dol.gov/sites/dolgov/files/ODEP/pdf/SSI_Youth_Larson_Final_Proposal.pdf.

Mathematica. "Promoting Readiness of Minors in SSI: Final Impact Evaluation Report." Washington, DC: Mathematica, forthcoming.

McKay, C.E., and M.L. Ellison. "Career and Technical Education for Students with Emotional Disturbance." Report prepared for the U.S. Department of Labor, Office of Disability Employment Policy, April 2021. Available at https://www.dol.gov/sites/dolgov/files/ODEP/pdf/SSI_Youth_McKay_Final_Proposal.pdf.

Michigan Career & Technical Institute. 2019. *Annual Report: August 2018–July 2019*. Plainwell, MI: MCTI.

Rupp, K., and G.F. Riley. "State Medicaid Eligibility and Enrollment Policies and Rates of Medicaid Participation Among Disabled Supplemental Security Income Recipients." *Social Security Bulletin*, vol. 76, no. 3, 2016, pp. 17–40. https://www.ssa.gov/policy/docs/ssb/v76n3/v76n3p17.html.

Sevak, P., K. Feeney, T. Honeycutt, and E. Peterson. "Linking Learning to Careers Demonstration. Impacts 24 Months After Enrollment." Princeton, NJ: Mathematica, August 2021.

Social Security Administration. "SSI Annual Statistical Report, 2019." 2020. Available at https://www.ssa.gov/policy/docs/statcomps/ssi_asr/2019/sect04.html. Accessed October 19, 2021.

Stapleton, D., J. Smith, and T. Howe. "Transition to Economic Self-Sufficiency (TESS) Scholarships for Youth and Young Adults with Significant Disabilities." Report prepared for the U.S. Department of Labor, Office of Disability Employment Policy, April 2021. Available at https://www.dol.gov/sites/dolgov/files/ODEP/pdf/SSI_Youth_Stapleton_Final_Proposal.pdf.

Taylor, A., T. Nguyen, and M. Honsbruch. "Improving Youth SSI Recipients' Employment Outcomes through an Integrated Treatment Team Intervention in a Health Care Setting." Report prepared for the U.S. Department of Labor, Office of Disability Employment Policy, April 2021. Available at https://www.dol.gov/sites/dolgov/files/ODEP/pdf/SSI_Youth_Taylor_Final_Proposal.pdf.

U.S. Department of Education. "IDEA Section 618 Data Products: Collection Documents, Part B, 2018-2019, Child Count and Educational Environments." 2019. Available at https://www2.ed.gov/programs/osepidea/618-data/state-level-data-files/part-b-data/child-count-and-educational-environments/bchildcountandedenvironments2019-20.csv. Accessed August 10, 2021.

U.S. Department of Education, Career and Technical Education Statistics. "Secondary/High School Level Glossary." 2021. Available at https://nces.ed.gov/surveys/ctes/tables/glossary_secondary.asp. Accessed October 18, 2021.

U.S. Department of Education, National Center for Education Statistics. "Students with Disabilities: How Many Students in Postsecondary Education Have a Disability?" 2021a. Available at https://nces.ed.gov/fastfacts/display.asp?id=60.

U.S. Department of Education, National Center for Education Statistics. "Students with Disabilities: How Many Students with Disabilities Receive Services?" 2021b. Available at https://nces.ed.gov/fastfacts/display.asp?id=64.

Mathematica Inc.

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org