

**Promoting Readiness of Minors in
Supplemental Security Income
(PROMISE):**

**Recommendations of the
Technical Advisory Panel
Regarding the Use of Incentive
Payments and the Evaluation
Design**

Appendices

February 28, 2012

Thomas Fraker
Todd Honeycutt



MATHEMATICA
Policy Research

Contract Number:
SS00-06-60110

Mathematica Reference Number:
06970.300

Submitted to:
Social Security Administration
Office of Retirement and Disability Policy
Office of Program Development and
Research
496 Operations
Pole 3-C-25
Baltimore, MD 21235

Project Officer: Jeffrey Hemmeter

Submitted by:
Mathematica Policy Research
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director: Thomas Fraker

**Promoting Readiness of Minors
in Supplemental Security Income
(PROMISE):**

**Recommendations of the
Technical Advisory Panel
Regarding the Use of Incentive
Payments and the Evaluation
Design**

Appendices

February 28, 2012

Thomas Fraker
Todd Honeycutt

MATHEMATICA
Policy Research

CONTENTS

APPENDIX A: TAP MEMBER COMMENTS ON PROMISE EVALUATION CRITERIA	A.1
APPENDIX B: PROMISE TECHNICAL ADVISORY PANEL MEETING NOTES	B.1
APPENDIX C: PROMISE TAP MEMBER POSTMEETING COMMENTS	C.1

APPENDIX A

TAP MEMBER COMMENTS ON PROMISE EVALUATION CRITERIA

Criterion 1: Interaction Between SSI Incentive Payments and the Local Interventions

1.1 How should incentives be structured to achieve transformational impacts (better use of resources, increased coordination among programs, and better outcomes for SSI youth)?

- A financial framework for assessing public/private provider costs and revenues based on specific program/process activities (similar to the Ticket to Work (TTW) milestones, and outcomes) should be fully considered. The program/process activities, in general, should include (1) outreach/recruitment, (2) intake (assessment, goal-setting, planning, etc.), (3) provision/implementation (interagency coordination and/or private source of service delivery), (4) follow-up services, and (5) management and coordination functions related to payment tracking and other administrative overhead. The outcome and outcome/milestone payment system used by TTW may serve as a general model; however, lessons learned from the TTW program should be reviewed to produce an adequate level of financial incentive. Cash flow to provide services and achieve milestones must be sufficient to cover costs, which may extend for more than two years, to achieve desired results. If the four primary state-level agencies are to be involved (Education, Labor, Human Services, and SSA), is one of them, then, to be identified as the primary entity responsible for managing and tracking participant progress and tracking payments?
- Are the incentive payments in addition to direct funding of program intervention sites? What is the relative level of the two—that is, are the incentives likely to be enough to alter program behavior? Timing of incentive payments (reduction in SSI benefits) relative to intervention costs could be critical.
- I'd recommend testing two approaches. In one approach, states are given flexibility to use funds in a more coordinated way to achieve specific performance targets. In this approach, there may not need to be financial incentives. In the second approach, there would be incentive payments.
- As much as possible, tie incentives directly and simply to better outcomes for SSI youth—by far the most important result. Better use of resources and increased coordination are not ends in themselves, but potential means to better outcomes, and if they really do work to that end, then you won't need additional incentives to get programs to do them. And if, in reality, they don't enhance results for youth, then they shouldn't be incentivized. By tying incentives directly and simply to the highest-priority outcome, it seems to me there are at least several benefits:
 - the efforts of providers will be focused on the critical goal;
 - there will be less chance that significant funds will be spent encouraging activities that may or may not yield substantive outcomes for youth; and
 - the activities of (and learnings from) these pilot projects will be unadulterated by systemic forces (incentives) that artificially skew project approaches.
- I believe there is much to be learned from the Ticket to Work Program and the new regulations that improved the payments for both outcome payments and milestone

payments. The incentive amounts must be structured in a way that adequately resources the work to be done to earn the incentive. I’m not arguing for a choice between the two models (outcome and milestone)—but perhaps there is a hybrid where both types of incentive payments are utilized. Another option would be to think about a “bonus formula strategy.” In such a scenario, states or organizations (whoever is the grantee?) involved would receive a bonus payment based on a formula in which each bonus is multiplied by the percentage that the organization has managed to exceed its baseline outcome (i.e., job placement rate). The grantee must maintain this increase in each successive year of the outcome (like compound interest).

1.2 Who should be eligible to receive incentives—private providers, state agencies, and/or participants?

- To achieve success, all parties will have to work toward the program’s goals. It is particularly important to have incentives for the participants as they are likely to be nervous about losing benefits. A key incentive for participants is maintaining their ability to continue receiving health services if they leave SSI and to return to SSI with relative ease should their transition fail.
- My bias is to strongly incentivize private providers that, in the best case, may be more sensitive to (and more nimble about) fully delivering the desired results. (This view is, no doubt, significantly informed by the fact that I’m rooted in a private provider and, before that, in private industry.) I would think that the primary incentive for participants (youth) would/should be meaningful work and that the more important issue relative to them (and their families) is the removal of real or imagined factors that might discourage their pursuit of work—in particular, loss of benefits that penalize them for being employed.
- As a general rule, it would be best to position the financial incentives as close to recipient service implementation as possible. While the PROMISE program is intended to operate at the state level, state administrative overhead costs associated with direct operation of the program should be minimized. It would also be of interest to explore how financial incentives can be shared with families or primary caregivers to support outcomes. This may entail the use of waivers that extend benefits to support meaningful employment outcomes. I believe the YTD demonstration made extensive use of a wide range of waivers to support participant participation in these projects. The YTD program may also serve, in part, as model.
- Incentive payments may be most effective when provided to participants and their families. Improved coordination among vocational rehabilitation agencies may result in enhanced reporting for SSA reimbursements. Similarly, local agencies supporting youth with disabilities and their families may benefit from some incentive support though, again, incentives paid to SSI participants directly may encourage involvement in employment activities.
- I’d suggest focusing on two models within the incentive approach. In one model, the state is SSA’s main partner and agrees to undertake some activities that are costly to the state in exchange for sharing of savings with SSA. In a second model, it is the private providers (or intermediaries in a pay for success model) who share in the savings.

- Like in the Ticket to Work program, I think that it would be useful to open the competition to an array of providers. I believe we should also explore the possibilities around participants receiving part of the payment based on outcomes (like AAA Take Charge). It seems to me that the funder should consider funding a combination of entities (both private providers and state agencies) perhaps even mandating that grantees form consortiums in order to apply.

1.3 How should the counterfactual outcomes be incorporated into incentive payment calculations?

- I think we should explore using a benchmarking model where the base year is the floor from which payment calculations are based on improvement for the treatment groups.
- In theory it is desirable to link incentive payments to program impact, but I think this might not be a good idea in this project. First, obtaining impact data with a reasonable follow-up period often requires several years. For the incentives to be meaningful, they must be provided relatively quickly. Second, incentives linked to impacts cannot be specified in advance to participants, providers, or state agencies; moreover, the incentives would have to be specified relative to performance of some other group, so they would lack specificity. Third, unless an experimental design is used, there are likely to be alternative approaches to computing the impact and specifying the comparison group, leading to further complications.
- If we start with the premise that counterfactual outcomes are generally viewed in relation to an assessment of program impact, then the counterfactual outcomes would serve as a comparison to outcomes achieved through the PROMISE program. The counterfactual outcomes could be used as an estimate of “added benefit” (lifetime earnings vs. SSA benefits paid out over a lifetime) as part of a cost-benefit analysis that would provide an estimate of longer-range costs that would be incurred by taxpayers/SSA if the PROMISE program was not initiated. The counterfactual outcome costs could be examined in relation to the financial incentives or investments made in the program.
- It is essential that incentive payments be based on outcomes relative to a counterfactual. This is important both for motivating good performance and for transparency in demonstrating to the public that the initiative did in fact achieve the claimed results.

1.4 What outcomes should warrant incentive payments? That is, given that payments should reflect actual savings, should payments be based on paid competitive employment, earnings, and/or reductions in SSI payments?

- All of the monetary outcomes identified here, including paid competitive employment, earnings, and/or reductions in SSI payments, are clearly important outcomes that warrant incentive payments. One additional consideration would be to think about the types of jobs and wage structures currently achieved by individuals who either do or would qualify for SSI benefits. Most employment occurs within generally low-skilled occupations, for which it has been difficult to achieve income levels sufficient to substantially reduce or eliminate SSI payments. Options and opportunities for individuals with disabilities to access postsecondary education training and/or other public/private vocational/technical training programs have been limited. If we assume

by all other available data that individuals who have participated in postsecondary education achieve higher lifetime earnings, then potentially attaching a financial incentive to service providers that successfully enroll PROMISE program participants in postsecondary education and/or public/private vocational/technical training programs could be considered.

- I'd recommend that the primary outcome be employment/earnings. If appropriate, educational attainment could be an outcome as well. Whether reductions in SSI payments would be an appropriate outcome depends on whether this target would be met solely through achieving greater self-sufficiency or whether it might also be achieved by discouraging people from receiving the benefits they are eligible for. In the latter case, it would be problematic to base incentives payments on reductions in SSI payments.
- If all you care about is savings, you could just use SSI savings, but I think this would not send the best message. At a minimum, I suggest including paid competitive earnings as well.
- Payments should represent a mix of outcomes, but priority should be given to paid competitive employment (of which earnings and reductions in SSI payments will follow). I think the standard should be competitive integrated employment. Increases in competitive, integrated employment will result in increased earnings and reductions in SSI payments.
- There will have to be some early- and mid-term benchmark payments to assure providers have the resources to deliver services over an extended intervention with the youth, and it makes sense to tie those payments to various interim outcomes that lead, ultimately, to reductions in SSI payments. Outcomes such as preemployment interventions with the youth, placement, retention benchmarks, and vocational growth. But ultimately, incentive payments have to yield reductions in SSI payments—otherwise, especially in the current fiscal environment, this effort will be indefensible. In order to assure that such reductions occur in adequate amounts, perhaps there should be some kind of outcome benchmark(s) that providers would have to meet in order to continue participating in the program. Without this, you run the risk that some providers continue to draw down interim payments for work with participants without ever delivering the ultimate outcome (savings to SSA).
- The summary says incentive payments are based on “reduced SSI dependence and other positive outcomes for project participants.” Yet it goes on to say the payments are based on actual savings in SSI benefits. Is it possible for a pilot intervention to be focused on increasing high school graduation (or staying in school)? Increasing postsecondary schooling, reducing teen pregnancy/parenting, reductions in arrests or other negative behaviors? It isn't clear that any of these would result in decreased SSI payments. In fact, is there an outcome other than employment that would lead to decreases in SSI?
- It seems that health interventions are a possibility for helping reduce SSI payments because you might reduce the impact of disability for some young people. This includes behavioral issues/treatment of mental health. Are there lessons on this from the Montgomery County YTD? This also seems to be a possible targeted subgroup intervention.

1.5 Should incentive payments be linked to group-level reductions in SSI payments (e.g., a reduction of state-level SSI payments by a certain percentage) or to reductions in SSI payments for individuals (similar to Ticket payments)?

- I do not think this can be adequately addressed until the design is more developed. For example, an important issue is whether states are expected to serve all those eligible.
- Since the PROMISE program is implemented at the state-level, states may want to link this to group-level reductions in SSI payments. However, for research/intervention purposes, linking reductions in SSI payments to individuals seems like a better choice. As noted earlier, this would be similar to Ticket payments, with modifications.
- In order that incentives deliver as much punch as possible, I think they need to be tied to individuals. To link them to group-level reductions would significantly dilute their impact on a given provider.
- Could we test both models concurrently or is that pie in the sky with the amount of funding available? What is to be learned from bonus payments to states and outcomes achieved (Personal Responsibility and Work Opportunity Reconciliation Act; Out of Wedlock Birth Bonus; Adoption and Safe Families Act; Adoption Bonus) versus Ticket payment structure? What does the data show us on most effective employment network's to date and their payment mechanisms (isn't AAA Take Charge one of the most successful?)
- It depends on the initiative. If a state decides to reform its approach statewide, then the group-level reductions might be appropriate. But most often, I would expect there to be a subpopulation within the state who was being served with the new approach and it would be the outcomes for the individuals in that subpopulation relative to an appropriate counterfactual that should be the basis for payments.

1.6 How should the requirements for obtaining incentive payments (documentation of outcomes achieved, administrative processing of payments) be defined? How can the burden on all parties of administering the payments be kept reasonably low relative to the size of the payments?

- Interim payments (for things like preemployment interventions, placement, etc.) would obviously require documentation from the provider—perhaps in conjunction with a quarterly system of submitting requests for payment? But from my largely uninformed perspective it seems to me that the beauty of payments for hitting given retention or earnings targets, or for reductions in SSI payments, lies in SSA already having all the necessary information. Could incentive payments be triggered automatically by SSA's system for tracking employment status, earnings, and benefit payments?
- One of the four state-level agencies could be deemed the administrative entity responsible for documentation of outcomes achieved, administrative processing of payments, and other federal reporting, as required. An independent private/public entity (community-based rehabilitation program, Workforce Center, DD Planning Council, others) could similarly manage the process and impact/outcome evaluations. Each state also has at least one federally designated University Center for Excellence in

Developmental Disabilities (UCEDD) that could serve as a host administrative entity at the state-level. UCEDDs routinely work collaboratively with their state agencies in managing state evaluations and administering new initiatives.

- The outcomes need to be measurable with administrative data—state UI earnings records or SSA earnings histories, SSI claim data, an appropriate federal or state education database.
- Again, I think we can learn from the experience from Ticket to Work and the documentation requirements for payments. I believe asking for pay stubs is arduous, and if we can come up with something easy but reliable, that should be the goal. I suggest SSA’s Master Earnings File, although I know there is a lag in getting the data from the IRS.

1.7 Should the structure for incentive payments be the same across sites, or should it be allowed to differ?

- In general, the incentive structure should be tailored to the specific intervention. But it might make sense to have a generic “the federal government will share 50 percent (or some other percentage) of savings with those who produce the savings” rule to simplify administration of the initiative.
- I believe we should seek variation in order to test different approaches. In fact, this could plausibly be a fundamental research question—what incentive structure is the most successful in encouraging states to increase competitive integrated employment and postsecondary educational opportunities among SSI youth?
- A case can be made for varying the incentives so that we can estimate a response surface. My concern is that we are likely to have a limited number of sites and that variation in impacts across sites could be due to other factors such as differing economic conditions, quality of management and staff, etc. I am definitely in the KISS (“Keep it simple, stupid”) camp here.
- My bias would be to have them the same across all sites, but I could make an argument for either alternative. By making them the same across the board, you have the ability to draw conclusions that are not compromised by “apple vs. oranges” arguments. But by allowing differences in approach, there is the potential for more creativity in design (and perhaps, ultimately, a better mousetrap).
- For the purposes of research and demonstration (not knowing the level of funding or how many demonstrations will be funded), some variation across sites would seem of interest in determining the impact of incentives provided. Should, for example, financial incentives be higher at the front end to encourage early planning and the intensive provision of services (work experiences, assessment, counseling) while the young person is still in school? Financial incentives for individuals and/or families could be considered, as well.
- There will likely be substantial variation across participants and sites. To the extent possible, controlling for these variations and applying similar incentive payments across sites would help with implementation fidelity and evaluation.

1.8 What other issues should SSA consider for PROMISE incentive payments?

- As noted above, a key incentive for participants is maintaining their ability to continue receiving health services if they leave SSI and to return to SSI with relative ease should their transition fail.
- It will be critical, especially if private providers are to be a significant part of this effort (as I think they should be) that there be sufficient early payments to allow a provider without a lot of cash reserves the resources to deliver the services necessary to gain the longer-term outcomes. This need has to be tempered, however, with concern that providers not game the system by collecting the early benchmark payments without ever delivering the long-term interventions that will yield savings to SSA and meaningful employment to the youth.
- There are lessons learned from the TTW program. Because there are a number of unknowns pertaining to the PROMISE program, several questions (likely many questions) need to be addressed. For example, is the PROMISE program viewed as a supplement to other funding intended to promote employment among participants? This would certainly drive public vs. private entity interest in participating. Are the incentives sufficient to be viewed as profitable by service providers? Will the PROMISE program focus on any specific set of participant characteristics or demographics? Meaning, will the program focus on individuals who are most likely to become successfully employed, or is the program going to be open to those who face substantial barriers to employment?
- The group being served needs to be defined up front so that cream skimming doesn't occur. It can't be defined by who actually seeks and receives services from the providers. A specific caseload needs to be defined, and even if only a subset of that caseload gets served, the incentive payments should be based on how the entire caseload fares.
- It isn't clear to me how social impact bonds fit into this. What is the estimated level of savings in SSI benefits for minors?
- If possible, it would seem beneficial to allow sites to weigh in on development of an incentive structure. Not clear if the incentive structure is one of the things you want to evaluate per se and therefore want it to be the same (or differ in specific ways) across the interventions.
- This could kill the project if not done correctly. I think we need to build upon the lessons learned from Project Network and Ticket to Work around structuring the incentive amounts to be adequate in terms of resources.

Criterion 2: Development and Measurement of Partnerships and Interventions

2.1 To what extent should the evaluation track project services as part of a process study?

- Services should be tracked in order to understand what is delivered in order to produce outcomes. Evaluators should consider standardized MIS software (such as Efforts to Outcomes) with uniform measurements and tracking of core services. The grant can mandate this and provide an MIS system through the evaluator.
- This is very important. It is important to verify what services are provided and under what circumstances. This is important for learning if the project design was implemented with fidelity.
- A process analysis should seem to be a critical component of the overall evaluation/research of the PROMISE program. This would start with a clear conceptual framework to guide analysis, identifying service elements or components and short-term and long-term outcomes. The process evaluation/ analysis would be essential in (1) documenting services (and interventions used) each project would implement, (2) evaluating fidelity to the original proposed model over time, (3) identifying implementation barriers as well as successes, (4) monitoring the adequacy of financial incentives from program initiation to implementation, and (5) producing information that can be used in addressing unintended consequences to improve eventual program outcomes and impact.
- I'd think it important to have some kind of process to track project services since they will be the means to the end of meaningful employment. The learnings will all be wrapped up in discovering which services do (or do not) tend to yield the desired outcomes.
- Service types, duration, and quality, as well as education and training, should be examined to the extent possible.
- We need to learn enough to know what program model was delivered or what approach was taken in both the successful and unsuccessful interventions—so that smart decisions can be made about replication and scaling up. We also need to know the total resources that were expended (including perhaps state or private resources) so that a determination can be made of whether the program model is cost-effective in locations where a similar level of private or state resources may not be available.

2.2 What project services should the evaluation track (that is, which are most important)?

- Preemployment intervention with the participant; job development activity and interface with the employer; postemployment supports to employer and participant; interventions in support of vocational growth (increased wages or hours, increased job responsibilities, etc.); use (and cost) of specific accommodation technologies or strategies.
- Project services that appear to be most important include (1) individual work experiences, (2) IEP/transition planning involving multiple agencies and the alignment of individual service plans (e.g., IEP seamlessly aligns with the IPE prior to exiting school), (3) youth development/empowerment (training in self-determination, student-

directed IEP planning, individual choice in decision-making options), (4) system linkages and collaboration (interagency agreements, memorandum of understanding, resource pooling, cost sharing, etc.), (5) family support (financial incentives, providing information that raises family/parent expectations for employment, etc.), (6) postsecondary education participation focused on employment goals, (7) social and health services (special health care transition needs addressed, systematic access to adult health care providers, addressing health care regimens and medications, etc.), (8) waivers (earned income exclusions, student earned income exclusion, plan for achieving self-support, individual development accounts, continuing disability review, etc.), and (9) benefits counseling.

- Given the evidence that minor SSI recipients and their parents have little knowledge of SSI work incentives, it is worth understanding the knowledge of these among controls and treatments, not just use.
- Benefits planning (financial planning); job development and job placement; discovery process for youth; peer family mentoring.
- Job finding, placement, training, follow-along, and secondary and postsecondary education experiences and attainment would seem important.

2.3 What incentives will pilot projects have to report service delivery and costs accurately?

- Unless these becomes cost reimbursement projects (which I would not recommend) the accuracy of cost reporting is really not critical (in fact, I'd question whether it's even necessary). As for service delivery, I think the provider should be able to demonstrate, in advance, their capability to track and document service delivery, it should be made clear that they must have documentation for any services reported, and that the project administrator should have the ability to randomly audit that documentation as he or she feels is necessary. Is it assumed that there will be some kind of regular audits of providers' activities throughout the project regardless?
- This should be a requirement for all pilot projects. Assuming that an oversight evaluation contractor will be used to support the pilot project, the resource components or "ingredients" approach to cost analysis would certainly be a minimum function to support the documentation of service delivery and costs. The cost analysis would be aligned with the process analysis discussed in 2.1. At a minimum, documenting average cost per project participant through outcomes achieved seems necessary for policy and future funding of such efforts. Funds to support the collection and reporting of service delivery and costs should be included in the financial incentives.
- If these are grants, specific monitoring and technical assistance can be put in place, and the grants can be placed on restrictions for failing to accurately and timely report.

2.4 To what extent should the evaluation track project partnerships (the involvement of different agencies and providers) as part of a process study?

- Tracking project partnerships as part of the implementation seems key if the main difference in these interventions is coordination of already existing services. It requires careful thought to come up with service measures of this coordination versus just the use

of separate services. For example, if a program intervention has an increased connection to VR—in the impact evaluation one could see an increased use of VR, but need the implementation study to show there was some new connection to VR.

- This is an essential component of the evaluation. Determining the pattern of relationships necessary to achieve employment outcomes sufficient to reduce SSI payments is critically important. I would recommend focusing on (1) the role and impact of state as well as local interagency agreements between special education and vocational rehabilitation (others as well), (2) the involvement and impact of outside agencies participating in IEP/transition planning teams prior to the student’s exit from high school, (3) the coordination and alignment of individual service plans, and (4) resource-pooling and/or cost-sharing models to support postsecondary education, training, and employment goals.
- I’m not convinced that partnerships and collaborations can really be dictated (despite the fact that it may feel that it would be good if they were happening). Too explicitly require certain partnerships often leads to game playing on the part of the provider in order to meet the funder’s requirements. My sense is that if incentives relative to outcomes are set right, then collaborations will happen when they make sense relative to those outcomes, and if they don’t make sense, then why do them.
- Project partnerships, including describing and evaluating involvement, would also seem important.
- I think this is very important. I think the project should consider mandating partnerships as part of the application process and we should discuss who those mandated partners are.
- This seems necessary to allow replication and scaling up.

2.5 What should a management information system for PROMISE track and for what purposes? What steps can ensure the accuracy and minimize the burden of data collection and entry?

- Training and technical assistance on the MIS is very important. The grant should be structured in a way to make sure that an MIS administrator is in place and responsible for coordination of data input in a reliable and quality way.
- There needs to be a common management information system across projects that includes core elements for each project to ensure consistency in data collection. A common set of measures, including common definitions of services, would be important. Common templates or tracking systems developed on participant demographics/characteristics, service/intervention components, short-term and long-term outcomes, state system impacts, costs, and other measures should be considered. I would assume that a contractor providing oversight for the PROMISE program implementation would address the question of how the data collection and entry could be maintained in an accurate manner and minimize individual project burden on such data collection. Online systems for data entry, contractor data checks, etc., seem reasonable approaches to make the data collection process efficient.

- The MIS should ideally track all services and outcomes. My experience is that an MIS is only useful for the evaluation if it is useful to the site (by producing reports that the sites desire), is provided with sufficient technical assistance to assure that sites implement it correctly, and is monitored for quality.
- Data tracking systems that allow easy tracking of services/outcomes—essentially that provide the most overlap with current program MIS—are likely to be least burdensome and most accurate. Actual streaming of information from the program’s MIS to the tracking system PROMISE sets up should be considered to reduce double entry needs.
- Using extant data collection systems such as the RSA-911, Ticket Research File, and other routinely maintained data systems may help reduce burden as well as provide opportunities for data verification.

2.6 What other issues should SSA consider regarding PROMISE partnerships and the delivery of services?

- Training, technical assistance, and ongoing consultation pertaining to project planning and implementation should be considered.
- We want to think about how to create efforts that will ultimately contribute to (1) the development of policies that support competitive employment in integrated settings as the first and desired outcome for youth and young adults on SSI; (2) the removal of systemic barriers to competitive employment in integrated settings; (3) the implementation of strategies and best practices that improve employment outcomes and decrease reliance on SSI; and (4) enhanced collaborations that can facilitate the transition process from secondary and postsecondary school, or other pre-vocational training settings, to competitive employment in integrated settings.

Criterion 3: Evaluation Design

3.1 What key research questions should the PROMISE evaluation answer?

- What was the participation rate among those eligible for PROMISE? How did participation vary by factors such as age, education, time on SSI, and type of disability? What was the impact of the offer and participation in PROMISE on outcomes of interest such as SSI and other benefits, competitive employment, wage rate, earnings, and health? How did the impact vary by factors such as age, education, time on SSI, and type of disability? What is the net present value of the project from the perspective of participants, the rest of society, and everyone?
- To what extent do incentive payments, services, supports, education, and training increase earnings and family income among SSI transition-age participants over time? How do family supports and services influence earnings and income among SSI transition-age participants over time?
- Which models worked and which ones didn't? I doubt it will be easy to answer the “why” question, but some insights about mechanisms may emerge.
- Three levels of evaluation and analysis should be included: process analysis, financial incentives analysis, and impact analysis:
 - Process Analysis:
 1. What is the local context of the pilot program, in terms of its economic, employment, social service, service delivery capacity, and existing interagency relationships?
 2. What is the nature and quality of the intended intervention? (Is the intervention research based? Has the intervention been widely used within the local pilot program context? Are staff clearly identified and knowledgeable about the intervention to be implemented?)
 3. Were projects implemented with fidelity?
 4. What steps were specifically taken to engage multiple agencies in the planning and implementation of the pilot program? (Is there a multiagency partnership agreement in place?)
 5. How were services and waivers implemented? (What were the outreach/recruitment strategies, intake and assessment, and planning? What services were offered to pilot program participants? What were the intended and unintended outcomes/consequences of delivering services?)
 6. What are the challenges and barriers in implementing the pilot program and its services over time?
 7. What are the costs of providing services leading to positive employment outcomes? (Total program costs and average cost per participant.)
 - Financial Incentives Analysis:
 1. Are the financial incentives significant enough to become profitable for service providers?

2. What are the critical milestones at which point financial incentives will be made available?
 3. What services are made available to program participants, in addition to those provided through the PROMISE pilot program?
 4. To what extent do the financial incentives produce employment outcomes as intended and reduce SSI benefits?
- Impact Analysis:
 1. Did the PROMISE pilot program achieve the intended goals?
 2. Did the PROMISE pilot program significantly reduce dependence on SSA benefits through increased earnings?
 3. What was the impact of the PROMISE pilot program on earnings, employment duration, and beneficiary income?
 4. Did the PROMISE pilot program produce net SSA savings?
 5. What are the costs and benefits of the PROMISE pilot program?
 6. Were the financial incentives significant enough to support the ongoing willingness of providers to remain involved?
 7. Does the PROMISE pilot program yield better employment outcomes for some participants over others?
 8. What are the state-level system impacts, in terms of changes policies, interagency agreements, funding, etc., that resulted from the implementation of the PROMISE pilot program?
 9. What is the family's economic impact, based on their son's or daughter's participation in the PROMISE pilot program?
 10. What are the satisfaction levels of program participants, families, service providers, and employers in relation to the PROMISE pilot program efforts?
 - This is fundamental, and the evaluation should be clear up front on the questions so the best evaluation design can be employed to answer those questions.
 - What are the issues and challenges in designing, implementing, and operating interventions to help youth with disabilities maximize their economic self-sufficiency as they transition from school to work?
 - What approaches are taken to provide, integrate, or add services to promote economic self-sufficiency among youth with disabilities? What are the characteristics of the interventions and the context within which they are provided?
 - What are the characteristics of youth with disabilities targeted by interventions? What are the program functions and treatment methods? What are the strategies for outreach and engagement of these youth, and to what extent do they reach their target population? How do differences in the target populations and their risk factors affect intervention design decisions?
 - What are the net impacts of the interventions on employment and earnings, SSI and Childhood Disability Benefits receipt, educational attainment, and trust fund and general revenues?

- What are the net impacts for different subgroups within the research sample? For example, do intervention impacts vary for youth with disabilities with different characteristics?
- What are the costs and benefits of the program from the perspective of the recipient, the government (federal, state, and local), and society? Do the benefits exceed the costs?

3.2 Should random assignment be implemented in each PROMISE site? Is it essential for a credible evaluation? Would requiring random assignment limit the nature of the innovations to be tested in any significant ways? What ways?

- Yes. If you get different results in sites using different methods, you will never know if the impacts are truly different or it was the methods used.
- Random assignment should be implemented in each PROMISE site. Given the scope and importance of this program, random assignment needs to be used to control confounding variables. I suppose other strategies, such as regression discontinuity designs, could be considered, but the need for them would not be as strong as for randomly assigning subjects to treatment and control groups. Because the types of services and interventions that will be proposed by applicants are not known at this time, it is difficult to assess whether random assignment would limit the nature of the innovations to be tested in any significant way.
- Random assignment would be preferable.
- I think that random assignment should be implemented. This is a fair and equitable way to allocate slots, and also controls for unobservables.
- Whenever possible, random assignment should be implemented. When randomization is not possible, rigorous quasi-experimental strategies should be used. It is hard to be more specific before we see what approaches are being considered. In cases where incentive payments hinge on the outcomes and an argument is being made about cost savings, it is critical that the impact measures be credible. In cases where a rigorous assessment is not possible, I would recommend less reliance on incentive payments.

3.3 If a nonexperimental design is adopted, what is the best approach for selecting a comparison group? That is, what is the best way to develop an understanding of the counterfactual?

- If you have a really large sample, you can try for a regression discontinuity design based on some index of characteristics.
- The answer to this question depends on a number of factors, based on who is served, total numbers of participants expected, etc. Mathematica has extensive experience in developing random and nonrandom group designs. Cost factors enter these decisions, as well. However, to state the obvious, there are nonexperimental comparison approaches that could be considered—interrupted time series analysis, regression discontinuity analysis, etc.

- We should not adopt a nonexperimental design.
- Qualitative methods such as interviews and focus groups may help identify key issues regarding employment experiences as well as what seems to work or not work with project implementation. Perspectives of partners and employers may also provide insight and guide project and/or future interventions.
- It really depends on the intervention. In some cases it may be possible to roll out the intervention gradually throughout a state. In other cases, the baseline outcome may be stable enough to project the counterfactual based on past experience. In still other cases, outcomes in a neighboring state may be a plausible counterfactual.

3.4 Which youth outcomes (e.g., earnings, postsecondary education enrollment, receipt of SSI benefits, use of SSI work incentives) are most important?

- All of those sound right. It is worth considering whether health outcomes, health care access outcomes, and/or a health care utilization measure would be appropriate (there may be impacts on Medicaid and Medicare spending that are significant).
- They are all important—are you really thinking of dropping one or more?
- Earnings, hours of employment, duration of employment, employment benefits, postsecondary education participation, occupational classification, use of SSI work incentives, receipt/reduction of SSI benefits. One strategy would be to limit the number of main or primary outcomes (i.e., earnings and SSI employment reductions over time), and view other outcomes as supporting or supplementary data to the primary analysis.
- The specified outcomes all seem important.
- Competitive, integrated employment is #1; earnings; postsecondary enrollment.

3.5 Should targeted outcomes for the evaluation be permitted to vary across sites?

- No.
- Yes—especially if we allow for sites to have their own innovation. This was the model used in the YTD demonstration and allowed the sites to best address the state and local variation in policy and service provision.
- There should be a set of a priori, primary outcomes where program impacts are expected or desired across all sites. There will, however, be variation in the interventions used by individual pilot sites to achieve specific outcomes. Some variation in outcomes (e.g., access to and completion of postsecondary education) would be useful to document in terms of their influence or impact on primary outcomes, such as earnings and reductions in SSI benefits. As in many research programs, there will be supplementary or unanticipated outcomes that are important to document and report.
- It seems like the basic outcomes—earnings, SSI spending—would be common. But others may differ according to the intervention or the subpopulation being served.

- Perhaps core outcomes could be identified along with targeted outcomes unique to local systems.

3.6 Should evaluation designs be permitted to vary across sites? That is, could one site have random assignment for individuals, another site have random assignment for sites within a state, and a third use a nonexperimental design?

- As a starting point, I would advocate for random assignment for individuals as a primary strategy. Some of this is uncertain at this time, in terms of number of potential sites SSA is interested in funding. It may also come to pass that certain sites selected may, during the course of implementation, experience difficulties in establishing an experimental design and may default to a nonexperimental, nonrandom comparison group design.
- No. This is a bad idea, especially if you do not have many sites.
- What is the time frame for observing outcomes—how long is the whole evaluation? How big is each site—how many participants are likely? This seems key for knowing whether outcomes or designs can be permitted to vary across sites.
- Yes. It is likely that different interventions will require different evaluation strategies.
- Increasing the sample size for random assignment may require consistency across sites.
- No.

3.7 What are the key sources of potential bias in impact estimates about which SSA should be aware, given the adequacy of the design to control for unobserved factors?

- Motivation of participants, expectations of parents, and family support. The best way to control for bias is to employ random assignment.
- One threat is that those who volunteer are likely to be those who expect to gain the most.
- Use of an experimental design would minimize potential bias.
- State/local per capita income, unemployment and poverty rates in addition to sociodemographic, disability, and other factors among individual participants and their families.
- Cream-skimming, self-selection, mean-reversion, age-related trends. It will be easier to identify these once specific interventions are being discussed.

3.8 What number of subjects and/or sites would be large enough to detect meaningful policy impacts? At the project level, are there feasible implementation changes that could substantially improve power?

- Four to six sites, approximately 450 experimental cases, and 350–400 for the control group. A number of factors contribute to statistical power. While I am not completely certain about how these interventions are proposed, or what criteria will be used in

selecting site participants, in general, decreasing sample variability and controlling for confounding variables are key.

- Recommend a power analysis to determine necessary number of sites and participants.
- For binary outcomes, a few hundred is the smallest. I would have to know the earnings variance to do a power calculation on the earnings outcome. Given the large fixed cost in setting up an intervention, there is definitely a minimum scale at which it is just not worth doing an intervention.
- It depends on whether we want to do subgroup analysis among participants (which I recommend). The YTD had a sample of 480 treatment and controls, but this did not allow for meaningful analysis of subgroups.

3.9 What are the key data collection issues for the evaluation, given the outcomes that will be measured? What types of data will be required? How should data collection be designed to minimize its disruption for awardees, SSA, and other agencies?

- The range of qualitative and quantitative data to inform the process analysis will likely be extensive. A common management information system, similar to that used in YTD (Efforts to Outcomes), would allow project staff to track participant-level activities and interactions. Each project would be required to collect and report data on a core set of elements or variables. Consultation with each project would occur to discuss data that are particular to individual projects and to ensure that a comprehensive set of data is collected. Aspects of the data would be relatively broad and include information on the local context/site, the intended interventions, project implementation (i.e., data on fidelity, recruitment and participant descriptions, enrollment in project services, waivers used), service use/impact (participant outcomes and follow-up, impact of waivers, satisfaction, etc.), costs and cost/benefits, and financial incentives analysis. A summary of activities at the close of the PROMISE program would capture additional information and reflections on lessons learned and policy implications.
- The outcomes need to be measurable with administrative data—state UI earnings records or SSA earnings histories, SSI claim data, an appropriate federal or state education database.
- Evaluation should include data on impacts, process, and a cost/benefit analysis. Data should be collected through a third party evaluator in conjunction with the grantees (for process/MIS data).

3.10 What other issues should SSA consider regarding the design of the PROMISE evaluation?

- Lag times in administrative data, compatibility of data sets across ED and SSA, adequately resourcing the evaluation.

Criterion 4: External Validity

4.1 Should grantees be allowed to restrict eligibility to subgroups of SSI youth? If so, what types of restrictions should be allowed?

- I think restricting to subgroups of SSI youth is interesting (certainly by age, possibly by disability type or group). But it really comes down to how many young people we are talking about enrolling. Seems like groups would be small. I wonder what size mandatory participation would lead to? If it is an enhanced service coordination, seems like mandatory would work (why not do it?). Makes it even more important to measure what services are given, though, to understand if the treatment actually got something.
- This is possible, but recommended only if all use the same rule. Candidates are age, expected duration of benefits, and time on SSI.
- My bias would be to noncategorical efforts; however, I wouldn't rule out such restrictions given that a grantee may have expertise in working with one particular subgroup that wouldn't necessarily translate to other subgroups. I don't feel qualified to dictate what restrictions should or shouldn't be allowed at this point—is that a decision that would need to be made on a case by case basis during the proposal review process?
- Yes, if there is a good theoretical or practical reason for doing so.
- Student factors or conditions within the site selected may influence decisions about how youth are included in the sample. It would seem reasonable that the sites should have some say in the subgroups of SSI youth to be served. This may vary in relation to certain age groups (e.g., using interventions focused on youth still in school vs. older students who have recently completed their high school program?). States may choose to focus on specific regions or urban areas of the state in which economic conditions vary, or other variations.
- Random assignment during recruitment and participation. Oversampling may be needed for low-incidence disabilities and other subgroup populations.
- What age range is the focus of PROMISE? It's called "minors in SSI," but I didn't see an age breakdown.
- No. I think that the grantees should include all SSI youth.

4.2 Should the eligibility criteria be allowed to vary across projects?

- No.
- Would varying eligibility criteria across projects create richer learnings based on the relative success of one set of criteria versus another, or are there too many other variables to make this meaningful?
- Again, it would appear reasonable that each PROMISE project should have some latitude in establishing its own criteria for participant eligibility. The intervention/services proposed within individual sites will drive some of this decision

making. Interventions typically address specific characteristics, attributes, and needs of participants, thus influencing who will be included in the pilot program (eligibility).

- Same across sites to the extent possible.
- Yes.
- No.

4.3 Should participation in PROMISE be mandatory or voluntary for SSI youth? If voluntary, how best can the experiences of volunteers, obtained via the recruitment process(es) used in the demonstration, be generalized for SSI youth in other sites or nationally? If mandatory, what should be the consequences for nonparticipation?

- Hard to say without knowing what the specific intervention is. I could imagine successful models in which part of the job of the service delivery organizations is getting youth to participate voluntarily. I could also imagine programs in which participation is a condition of receipt of SSI—though that may require legislative authority.
- This largely depends on whether you want to evaluate a voluntary or mandatory program. If it is likely to be mandatory (although it is not clear what this means in terms of service receipt), I imagine you would get more significant treatment estimates.
- There are at least a couple of factors that argue for a voluntary model:
 - By working with youth who wish to be involved, the potential for positive outcomes is greatly enhanced. If program participation is effectively marketed to youth, the best-case scenario would be to have competition among youth for limited PROMISE slots. Forcing youth to participate is likely to create all kinds of dynamics (overt and covert) that will hinder program success.
 - If participation were to be made mandatory, would there be anywhere near the capacity to provide services for all youth? I would think not.
- Participation in the PROMISE pilot project should be voluntary. Sites that are selected should clearly demonstrate the capacity to generate a large enough number of subjects to allow for the attrition of those individuals and families who opt not to participate in this study. Random assignment will ensure the internal validity of the impact estimates. Regarding generalizability, I am assuming that, because sites are likely to be selected to represent diverse areas or regions of the country, as well as unique characteristics in these areas or regions—urban, suburban, and rural; economically impacted areas; and so on—the generalizability to the nation will be affected (minimized to some extent).
- Results should reflect characteristics and factors among participants interested in working. Examining domains of self-determination and/or motivation may potentially serve as predictors of employment.
- We should also discuss the merits of Intent to Treat vs. Treatment of the Treated. I look forward to the discussion.

4.4 What features of the state and local environments will make it more or less likely for projects to achieve their intended outcomes?

- Public transportation systems; unemployment rates; mix of industry types; typical area wage rates; cost of living; socioeconomic factors that influence attitudes about work.
- Features that would support the project include sites that are large enough to support robust sampling design (approximately 800 youth to be enrolled); sites agree to participate in a random assignment evaluation; sites are recognized for their knowledge and use of research-based interventions; sites that vary in terms of their context economically and demographically; and sites (states) that are fully committed to an interagency partnership to ensure the project's success.
- Please see this document: <http://www.ssa.gov/disabilityresearch/docs/YTD%20Special%20Process%20Report%202-22-2010.pdf>.
- There needs to be strong, enthusiastic leadership at the relevant state agencies. If this has to be forced upon someone, it won't succeed.
- State/local per capita income and unemployment and poverty rates in addition to sociodemographic, disability, and other factors among individual participants and their families. In addition, the funding and vitality of state and local agencies and organizations involved in the project may have impacts.

4.5 What key information should the evaluation produce to support the design of a national program modeled after PROMISE? How should such a program be rolled out?

- The evaluation should have an impact, process, and cost/benefit analysis. It should produce credible impact estimates.
- If there are several promising models in the literature, SSA should try to ensure that they are all tested so that at the end of the day there is enough information to make the decision about going national.
- Given that the ultimate issue is pretty simple (is the cost of supporting a project like this more than offset by reductions in SSI benefit payments?), there are lots of inferences that can be drawn regarding the efficacy of one approach versus another. To draw those inferences one would, obviously, need to collect data relative to the particulars of each approach and match them with outcomes achieved. At the end of the day, however, this will not (or should not, in my view) yield a set of dictates as to how one goes about this work; but rather, a number of suggestions. I would not favor creation of a national program that is overly prescriptive (outside of issues of ethics and legalities) relative to the means of achieving desired results. A national program model should include the capacity for a flexibility in approach that accommodates different organizations, different environments, different skill sets, and perhaps most critical, creativity. In my experience, this work is as much art as science, and art requires some space. To be clear, though, what does need to be clearly defined and presented as nonnegotiable is the desired outcome(s) of the effort.

- Information from the evaluation that would be most meaningful includes (1) a clear description of participant impact data (i.e., earnings, reductions in SSI benefits, other outcomes), (2) complete descriptions of the sites and the implementation steps used (i.e., advisory councils, state support, financial incentives, challenges and barriers encountered and overcome), (3) detailed descriptions of interventions used (fidelity criteria, evaluations of impact, evidence base of the intervention, etc.), (4) waivers used in relation to achieving participant outcomes, and (5) program costs.

4.6 What should the evaluation examine to inform SSA about the feasibility and appropriateness of the incentives for a national program?

- Bottom line seems pretty simple: Is the cost of supporting a project like this more than offset by reductions in SSI benefit payments? If not, could the incentives be tweaked to create a more favorable return on investment?
- A benefit/cost analysis of the PROMISE pilot project should be conducted. The findings from the benefit/cost analysis would inform SSA and help them decide whether or not to expand the program. Information could also be obtained qualitatively to reflect on the lessons learned in relation to the individual service provider experiences with the financial incentives instituted.
- It should critically examine work incentives and any special waivers employed.

4.7 What other issues should SSA consider regarding the external validity of the PROMISE evaluation?

- How to best have external validity given that I assume we do not have the funding to create a design that will employ a nationally representative sample?
- If the PROMISE project demonstrates a viable option in promoting positive changes in the outcomes of youth with disabilities who receive SSI, then a formal scaling-up plan would need to be undertaken by SSA. There would also need to be an ongoing process of information development and dissemination, training and technical assistance, and other forms of consultation to support states and local communities to fully implement this type of program.

OTHER ISSUES

What issues other than those covered in this document should SSA consider for PROMISE?

- Suppose that one approach will be Pay for Success (social impact bond) contracts with private service providers (or private intermediaries who assemble the teams of private service providers and raise private capital to finance operating costs in advance of the performance-related payments being received). An interesting question is whether SSA should get directly involved in setting up the pay for success contracts or whether it should simply offer states 50% (or some other fraction) of SSI savings if states set up social impact bonds that reduce SSI spending and achieve good self-sufficiency outcomes. On the one hand, it would be operationally challenging for SSA to negotiate and oversee these contracts directly. On the other hand, states may not take the initiative to set these up on their own without significant encouragement from SSA—since the savings will accrue primarily to the federal government. States may benefit if their Medicaid costs go down. At any rate, it is important to think about how to structure this in such a way that states will choose to participate.
- How does this project relate to the Youth Transition Demonstration evaluation? My understanding is that project is targeting roughly the same group with a focus on employment/employment services. I'd like to hear more about the connections and what SSA's plan for differentiating this initiative is.
- I think we should all be versed in the lessons learned from YTD: <http://www.ssa.gov/disabilityresearch/docs/YTD%20Special%20Process%20Report%202-22-2010.pdf>.

APPENDIX B

PROMISE TECHNICAL ADVISORY PANEL MEETING NOTES

PROMISE TECHNICAL ADVISORY PANEL MEETING NOTES

DECEMBER 6, 2011

Opening Remarks by SSA

The aim of the PROMISE initiative is to improve outcomes for minors in SSI in the areas of employment, education, and health. The intervention will build from the lessons learned in the Youth Transition Demonstration but will differ in several key ways. First, PROMISE will focus on the family unit, recognizing that improvements in the lives of minors in SSI will mean improvements for the entire household. Second, PROMISE will emphasize improved coordination among service providers to better and more efficiently serve the population. Third, PROMISE will incorporate incentive payments to encourage outcomes that will lead to SSI cost savings.

Currently, funding for the PROMISE intervention and evaluation are estimated at \$30 million and \$10 million, respectively. However, the ultimate funding amount could be less or nothing at all. The goal of the TAP is to identify evaluation priorities, given the potential for short-term funding, and develop criteria for evaluating grant proposals.

Opening Remarks by Department of Education

PROMISE will involve collaboration and build on the power of partnerships. It will show how to leverage resources at the local, state, and federal levels. If PROMISE is not funded, then the planning for PROMISE could still lay the groundwork for changing the system to promote SSI youth outcomes.

Topic 1: Interaction Between PROMISE Incentive Payments and the Local Interventions

Incentive payments (IP) do not need to be based on actual SSI savings for the demonstration. IP will be paid through the \$10 million allocated to SSA for the evaluation. Approximately half will be spent on evaluation efforts, leaving approximately \$5 million for IP. In addition, grantees (states or providers, not specified currently) could choose to use some of their funding from ED as incentive payments to families or organizations.

Many interventions are implemented without much confidence that they will achieve impacts. PROMISE should do what it can to promote a larger, more intensive intervention, since SSI savings could potentially be \$20,000 or more.

If states are the grantees and received the incentive payments, they could propose innovative methods for blending and braiding funding among agencies. State-level interventions would likely be more scalable than a provider-level intervention. However, providing the IP at the state level could mean that less money will trickle down to the service providers, and state politics could interfere, encouraging some odd behavior, as exemplified by the Workforce Investment Act or Job Training Partnership Act initiatives. Incentives to states may also be too distant for providers who are working toward youth outcomes.

The PROMISE demonstration will have limited opportunities for outcomes during its five-year duration, particularly on impacts. Linking IP to SSI savings may not be feasible for PROMISE

because real savings may not be realized in the time frame. Additionally, incentivizing short-term savings could be detrimental (keeping people off rolls, diversion from education/training, and so on). Given five years for the evaluation, intermediary outcomes (milestones that likely lead to SSI savings) may be the only possible success measures available and, thus, where the IP need to be linked.

The incentive payments could be provided in stages. Initial payments could be offered for specific intermediate outcomes (such as each treatment group youth successfully completing high school). A secondary payment could reflect impacts (for example, the difference in high school graduation relative to the control group) that are prespecified and suggest meaningful differences. A final payment could be based on long-term impacts (SSI benefit reductions). These approaches vary in the time needed for data collection and analysis.

Besides IP encouraging certain outcomes or SSI savings, they also serve as a new revenue stream and a method of service sustainability. After the initial PROMISE grant depletes, the project could theoretically continue running on IP.

Families of SSI minors are an important facet of PROMISE. They could potentially receive IP (for example, AAA Take Charge provides a portion of Ticket payments to beneficiaries), but the project should avoid creating motivations based solely on money. This is true for employers as well as families: subsidized employment often does not last beyond the subsidy, and employees drop out when the subsidies stop. It may be best to encourage a family to increase their employment and earnings by providing assurance that they will not lose SSI support (health care, cash assistance) or they will be able to reclaim these supports quickly if they need to (that is, through benefits counseling). Benefits counseling will be an important aspect of the interventions to teach youth and families about issues such as work incentives and health coverage options. Communicating about the program and its incentive model to SSI families will be important to building trust and ultimately encouraging change. Working with families and deciding their role in IP will require innovative ideas and could be left to the discretion of the grantees.

SSI cost savings is not a good short-term measure to use in awarding incentive payments. Encouraging SSI cost savings for families with youth could have detrimental effects by disincentivizing school or unpaid work experiences that could pay off in the long term. Intermediate outcomes that are leading indicators of SSI cost savings and can be tied to incentive payments include high school achievement, postsecondary training, and work experience in high school. Grantees could suggest the intermediate outcomes that are the right measures for incentive payments (and so identify what they feel sure that they can achieve). However, the language in the PROMISE proposal specifically states that the incentive payments should be tied to SSI savings, and thus the \$10 million allocated to SSA may not be used to reward gains in intermediate outcomes. The \$30 million for ED does not have such restrictions and could potentially provide funding for intermediate IP. However, any ED funds used for IP would not be available for direct or up-front service funding.

The strict language requirement in the PROMISE proposal guarantees that the ultimate goal of SSI savings is not lost, which could happen by incentivizing education and early work experience outcomes. In theory, IP reward decreased dependency on SSI receipt generated from services that would have been difficult to fund without blended money. The IP could continue to fund the new, collaborative system. The concept of linking IP to SSI savings stems from the social impact bond model, where real investment returns are expected. There is a risk that rewarding intermediate

outcome gains with IP will not lead to future returns and will not support successful, collaborative systems.

Topic 2: Development and Measurement of Partnerships and Interventions

The process study should provide a detailed account of a site’s approach to improving SSI youth outcomes that will lead to SSI cost savings. A good process study can show why certain interventions worked better than others and illuminate the black box between inputs and outcomes. The study should record all services, especially the intensity of the services (average cost and duration) and the fidelity of the project implementation to the proposal. These are important to record because they can show how the program had an impact and can hold a grantee accountable to its proposed intervention. Only well-funded and intense services with good success measures are likely to produce significant impacts. Given that each project will have different interventions, it may be helpful to have the grantee suggest some of the process measures.

The process study should include measurement of the partnerships that arise and capture the system change efforts by the grantee. What are the providers collaborating on? How are they combining their funding? One question is deciding how to identify partnerships, not just having agencies “agree to agree.” Many agencies already have partnerships; however, it is not clear that services have increased because of those partnerships. Coenrollment of families into services and the dollar amount of services (in-kind and actual) could be important to measure. There is less interest in how agencies “play” together than how they leverage funds and coordinate services together (for example, resource pooling, cost sharing, and policy alignment). Another way of looking at partnerships is whether families receive services from multiple providers (increased their use of existing services), which could reduce reliance on SSI.

The RFP could suggest some successful intervention approaches from past studies (collaboration, family benefits counseling) that the applicant could employ. It will be important to avoid steering the applicants toward certain methods or taking a checklist approach to the application, thereby reducing innovation. A second option is to present clear outcome measures to evaluate in the RFP and allow the applicants to specify the inputs (and how to measure them) for producing the desired results. The grant application would require applicants to submit a logic model that specifies how the inputs can produce the desired outcomes and provide evidence of the theory, and to suggest data and measures to collect as evidence of their model. The advantage to this approach is that it allows proposals flexibility and creativity; ED staff should be able to identify appropriate proposals without having to prescribe the actual interventions. With either approach to writing the RFP, it will be important to provide some common definitions of key outcomes. Grants would be awarded to the applicants that have the most evidence-based logic model along with a rigorous MIS.

One minimal approach to the process evaluation would be to randomly assign cases, use administrative data for outcome measurement, and conduct site visits to document services. This avoids survey costs and allows more money to be available for services.

While PROMISE should foster innovation, if grants are awarded to states, and then states subcontract to other agencies/providers for services, with each providing different kinds of services or using different models, there may be a risk for the evaluation in finding results. What may be important, though, is the incentive: if all are working toward the same outcomes, it doesn’t matter what they did to achieve the goal. Process analysis can show what they did.

Given the involvement of minors and families, privacy, choice, and confidentiality will have to be important parts of the demonstration and research designs. These also have implications for data collection, particularly in accessing state and federal databases.

MIS Reporting

Having a common data collection system across sites through a central MIS could be important for understanding site interventions. The MIS developer would need to have a firm understanding of the evaluation and the data analysis needs in order to design the system for efficiency and usefulness. Ideally, the people using the data for evaluation purposes would also be involved in MIS development.

Several approaches can promote the quality of data entered into the MIS. First, the grantees can be asked to provide input about how to design the MIS. Second, the grantee should be able to use the MIS for its own benefit so that it is more likely to invest the resources to ensure quality data. Technical assistance and customizing the systems will help toward this goal. Third, the amount of data should be minimized to avoid burdening the provider with data collection tasks that are not vital. If possible, avoid double data entry (into the grantee's and PROMISE's MISs), and maintain a good amount of consistency across sites for comparison purposes.

A part of the evaluation criteria should be a grantee's capacity to monitor and track their participants. As part of the application process, applicants should be asked to demonstrate their current MIS by generating reports based on past projects and clients and describing their quality control and audit processes. Another indicator of MIS capacity is access to dedicated MIS staff.

An advantage of using an MIS is that it can inform the process analysis by enriching the understanding of the intensity and types of services delivered. This could be important for PROMISE, particularly if the approach is to not prescribe the services to be delivered. A disadvantage of the MIS is that it will not contain information about the control group youth; a survey would need to collect service information to look at differences between treatment and control group youth. Additionally, it may not contain the qualitative data needed to understand why some interventions worked better than others.

Is an MIS Necessary?

While important, a central MIS may not be feasible for the evaluation budget. One alternative to an MIS would be to access data from existing federal or state systems. SSA administrative data can provide earnings and benefits. RSA-911 data can provide vocational service information for those receiving VR services. ED has information on education, and states are working to develop longitudinal databases that include postsecondary data (with some states further along than others). Disadvantages to using existing data systems are data-sharing conflicts and security issues involved in a multiagency effort, and the systems may not include all of the information needed for the evaluation.

A second alternative would be to gather information from the grantee/providers, using either the MISs that they already use or by collecting periodic reports on services provided. This information could inform the process analysis without the need for further monitoring. The evaluator should have a clear understanding of the definitions of the data that are being collected to ensure uniformity of data across sites.

A third alternative would be to measure service delivery through a participant survey. However, what participants report they experienced may not match up with MIS data.

A fourth alternative would involve well-conducted site visits and focus groups (including with control group youth), which could provide the process information needed to understand the on-the-ground intervention at each site.

Topic 3: Evaluation Design

PROMISE should incorporate evaluation design into the intervention design. YTD introduced the evaluation after the sites had begun services and random assignment was no longer possible for some sites. If an evaluator is appointed at the same time as the project officer, they are able to collaborate and produce a more seamless intervention and evaluation design. In addition, grantees have an understanding of the evaluation and what it will require. As a result, they can plan their implementation, participant selection, and data monitoring accordingly.

Random assignment of all youth in SSI or in a certain age range would be a large endeavor. Judging from past studies such as TIW and YTD, the participation rate for the treatment group could range from 2 percent to 20 percent, though more likely between 5 percent and 10 percent. Such a design would result in the evaluation following many treatment participants who did not receive the treatment. The treatment effect would be diluted in a group where 90 to 95 percent is not receiving the intervention (although statistical analysis could still provide sound impact results with the intent-to-treat model). Additionally, a large sample and finite evaluation funding would preclude a full follow-up survey. Primary outcomes would rely on administrative data, and a follow-up survey with a stratified design would survey treatment and control youth, giving more weight to selection of participants. The follow-up survey would collect data on secondary outcomes.

Alternatively, PROMISE could select its sample from SSI minors and families that volunteer for the intervention. This model is similar to YTD, which aggressively recruited participants. The sample would include approximately 15 to 30 percent of SSI youth. After volunteering, a minor would be randomly assigned to the treatment or control group. A full follow-up survey would be feasible with this sample. It would also be possible to have a second control group of minors drawn from those who did not volunteer for PROMISE. A criticism of the volunteer sample approach is that it would not capture the impact to nonvolunteers (or less motivated individuals). The intervention could potentially have a strong impact on this population.

The main advantages of using the full SSI population for the sample are that the evaluation could determine the participation rate for the intervention and understand why some refused to participate. The results would also be better predictors of the impact if the PROMISE intervention were to be expanded nationwide. The volunteer approach may compromise external validity and be less informative of the impact PROMISE could have on all SSI youth, but its results would better inform on the potential impacts for program participants. The smaller sample size would allow for a more complete survey approach and a treatment population with a higher participation rate.

A third evaluation design approach would be to choose a subpopulation (by age range or disability type, for example) that would likely benefit from work-related activities and randomly assign each of those minors to treatment or control groups. Some external validity would be maintained while minimizing the cost of data collection.

Several demonstration populations were mentioned:

1. The PROMISE target population could include those youth who do not receive positive age-18 redetermination decisions. Such a model would be more like the UK prison recidivism model that is using social impact bonds for funding.
2. Proposals may focus on SSI subgroups, such as those with mental illness or behavior problems.
3. PROMISE could be structured as a diversion program for SSI applicants (such as those who apply at age 18).

Similar to other programs, PROMISE will involve voluntary participation. It would be difficult to mandate work or work-related activities for treatment group members.

The PROMISE intervention may involve system change efforts, including increased collaboration and cooperation among agencies and providers. If so, control group members could be contaminated by these efforts. It could be possible, if states were the grantees, to randomize sites within a state, if the systems change efforts were not statewide.

The purpose of PROMISE is to fund places that have good ideas to “reach for the stars,” not to get areas with poor outcomes to achieve average outcomes. To promote the chances of detecting impacts, ED could select grantees in service-poor environments, but they must have a vision of how to take their service environment to the next level.

“Minor” implies youth ages 14 to 17. PROMISE will need to involve youth and their families, as parents will need to sign consent forms, but family involvement should go beyond that. Determining the age of the SSI youth in the intervention will be crucial to its design and success. An intervention for 14- or 15-year-olds might target family employment, while 16- or 17-year-olds might benefit from an intervention promoting work-related skills and activities. Measures of family impact will be harder to detect as there are more-complicated factors (such as intersections with TANF and child services). It may be unfair to ask grantees to show short-term impacts on families since changing family behavior is a slow and difficult process. The RFP could potentially proscribe some services (such as family benefits counseling), but leave other services, including family services, up to the discretion of the applicant.

Central Elements of the Evaluation

The evaluation should consist of a process study, impact analysis, and a cost-benefit analysis. Data should include state contexts, school attendance and graduation, employment and internships for youth, family employment and earnings, and other services and benefits received. Subpopulation analysis could be included, depending on sample. The evaluation could use only a follow-up survey, not a baseline survey, to limit costs.

An analysis including family unit outcomes is essential because the intervention is looking to reduce SSI and improve employment without disrupting the family life. There is little evidence pertaining to SSI youth and family interventions. Selecting participants based on family characteristics may be possible, but the project will have to consider how to include SSI youth who are in non-family households (such as foster homes).

One concern is having interventions vary in their outcomes. For example, projects focused on youth employment would have different outcomes than projects focused on family employment. Having different outcomes across sites would increase the cost of evaluation.

SSI savings will likely not be achieved for youth during the demonstration except through household changes in employment and earnings. A five-year intervention that starts with youth ages 14 to 17 may be able to affect intermediary outcomes such as school attendance, school achievement (high school diploma, not certificate) or work experience that could eventually lead to SSI reductions in later life. An emphasis on education or health outcomes presents a risk that the project will have less focus on employment and therefore will not achieve any cost savings through SSI.

PROMISE does not have to be cost neutral with its incentive payments during the demonstration. Such a standard would be too high given the short time period. However, the evaluation should produce evidence as to whether the model can sustain itself in the long term (7 to 10 years). SSA and other agencies can follow PROMISE youth in administrative data in the long term (10 years); if that aspect of the evaluation demonstrates cost savings, then that savings could be front loaded to fund the program.

Topic Four: External Validity

For the PROMISE demonstration to produce information to support the design of a national program modeled after PROMISE, the evaluation will need to provide evidence of program impact on intermediate and long-term outcomes—specifically, that youth left the SSI roles. It should offer a model for impacting SSI youth and their families' future SSI dependency (and making them better off overall financially).

The evaluation should identify federal mandates or policies that, when softened or waived, open up collaborative possibilities at the state and local levels for serving SSI youth better. PROMISE can demonstrate how four separate government agencies can cooperate and share data to achieve more success for the people they are serving.

More realistically, PROMISE results would likely not lead directly to a national initiative, but instead build knowledge for other research and demonstration projects. Thus, it will be important for PROMISE to evaluate grant applications on their ability to be replicated and scaled up to other areas. Additionally, grantees may produce significant outcomes for a specific population only, so further exploration will be needed on how to expand the services' impact to a wider range of youth.

PROMISE will need to follow youth through administrative data for long-term primary outcomes: employment, earnings and benefit receipt. PROMISE participants could feed into other work demonstrations such as TTW. This would allow PROMISE to continue tracking participants and give longer-term job support to the SSI youth.

If PROMISE did manage to show impacts on intermediary outcomes such as high school graduation or postsecondary training, that evidence would be enough for ED to consider providing more funding for the future.

APPENDIX C

PROMISE TAP MEMBER POSTMEETING COMMENTS

PROMISE POSTMEETING COMMENTS FROM TAP MEMBER A

1. Interaction between PROMISE incentive payments and the local interventions. *How should SSA structure the incentive payments to promote the intended efforts by service providers and outcomes for participants?*

This topic deals with the important issue that providing the right incentives can go a long way toward promoting the intended outcomes of the PROMISE demonstration. As was brought out during the discussion, providing the *wrong* incentives is far worse than providing no incentives—they can actually lead to worse outcomes.

As was noted in the discussion, incentives can be provided to states, providers, and SSI families. An important first consideration is which of these parties already has its incentives aligned with the federal government and would like to reduce SSI payments. One issue that did not arise in the meeting is that many states have their own SSI programs that supplement the federal SSI program. Depending on the structure of the program, such states may already have an incentive to reduce future SSI payments and thus need fewer incentives in the demonstration.

A point only tangentially discussed at the meeting is that there are existing programs that provide incentives to one or more of these groups. I do not have detailed knowledge about these programs, but we should see what lessons can be learned. Reference was made to the YTD demonstration, but no one pointed to any specific lessons for this project, which was puzzling to me. I believe the greatest experience we have is on providing incentives to states and providers under the Ticket to Work program. I have not kept up with this literature, but I believe that in the early days of the program, the incentives did not work well and provided incentives for vocational rehabilitation agencies to capture the incentive payments. One major lesson from Ticket to Work, as I recall, is that providers are reluctant to spend a large amount of their own resources with the hope of a payoff in the distant future. Thus, the dilemma is that even though we wish to achieve long-term savings in the SSI program, we need to balance that with the need to assure providers that they have a good opportunity to recoup their costs in a reasonably short period. This would seem to call for payments based on short-term predictors of future success and/or expenditures by providers, but those strategies have other problems, which are discussed below.

Incentives to participating families appear to be a necessary part of any incentive package. Families are rightfully worried about leaving SSI and losing health insurance and (to a lesser extent) cash benefits. SSA is currently implementing the Benefit Offset National Demonstration (BOND) to test incentive payments to adults on SSI. Although I believe it is too early for results, I believe the demonstration went through a large number of designs with many variations being considered, including varying the combinations of benefit incentives and services. It would certainly be worth reviewing all the designs that were considered for BOND and what the rationales were for various alternatives.

As I mentioned above, having the wrong incentive can easily be worse than having no incentive. The PROMISE demonstration is tailor-made for possibly having the wrong incentives because we need to offer incentives in the short run that may change long-term outcomes. For example, although the ultimate goal is to increase employment and reduce SSI payments, providing incentives to increase employment in the short run could cause participating youth to work instead of increasing their education and training. There is a substantial literature on the effects of work

while in school on later outcomes, so that literature should be reviewed. In any event, it seems likely that a good strategy would provide varying incentives for youth who have not graduated from high school and those who have and the incentives might encourage particular types of investment in education and training, and not necessarily reward full-time work more than part-time work.

This brief note is not the place to summarize what has been learned from the performance incentives used for the Workforce Investment Act (WIA) and the Job Training Partnership Act (JTPA), but the difficulty of aligning state and federal interests is clear. Originally, the JTPA program rewarded status at program termination, and programs responded by emphasizing job placement over training and education. Over time, the program has used somewhat longer follow-up periods, which provides a better incentive to provide lasting gains. The fact remains, however, that short-term outcomes, at least in training programs, do a relatively poor job of predicting long-term outcomes.¹

One issue that was not discussed in detail at the meeting is that people with disabilities are often reluctant to seek and continue full-time employment because qualifying for SSI is arduous and time consuming, and they do not want to risk losing their fought-for benefits. Thus, it is important for the incentive structure to make it relatively easy for beneficiaries to return to their prior status if their effort at work fails.

My main lessons on this issue are the following:

- Although incentives can be useful, it is impractical to offer incentives based on long-term outcomes, and short-term outcomes may be a poor or even inappropriate measure for the short run;
- It is more important to offer incentives to families than to government agencies or providers; and
- Given that the desired outcomes are beyond the time frame of the demonstration and that providing incentives for short-term indicators does not necessarily provide the right incentives for the long-term outcomes desired, I would recommend keeping the incentives relatively small.

2. Development and measurement of partnerships and interventions. *What incentives will service providers and state and local agencies have to work together to efficiently deliver the intended services? How should interagency cooperation and the delivery of services be measured?*

I was somewhat puzzled by three aspects of the write-up for this issue. First, the write-up appeared to stress minimizing the process study; second, the write-up downplayed the utility of a good management information system; and third, the write-up did not discuss tracking control/comparison group services. On all three of these issues, I believe the write-up was incorrect.

¹ See Burt S. Barnow. “The Relationship Between Performance Management and Program Impact: A Case Study of the Job Training Partnership Act.” *Journal of Policy Analysis and Management*, vol. 19, no. 1, 1999, pp. 118–141; and James Heckman, Carolyn Heinrich, and Jeffrey Smith. “The Performance of Performance Standards.” *Journal of Human Resources*, vol. 37, no. 4, 2002, pp. 778–811.

The write-up states that “There is less interest in how agencies ‘play’ together than how they leverage funds and coordinate services together.” Although approaches like braiding funds are currently popular, I do not think this precludes how well the agencies work together and how well aligned their missions are.

The discussion of minimizing the process study seems to ignore the fact that getting people to successfully leave SSI is complicated and we do not have a lot of documented models of success. If one or more of the sites is successful, we would certainly like to know why, and simply gathering data on services and outcomes may not be enough to disentangle complex effects. In any event, I do not think that process studies are that expensive that we need to minimize them.

The write-up questions whether an MIS is necessary. I do not recall this being the consensus at the TAP meeting, and even if it was, I do not agree. Without a uniform MIS, the evaluation will be unable to determine if characteristics, services, and outcomes are measured consistently across sites. The only way to assure that adequate data are collected is to specify the items to be collected, provide clear definitions, provide technical assistance to sites, and monitor use of the MIS. I have tried to evaluate programs where the MIS was not well developed and/or implemented, and usually there were so many problems with the data that they were essentially useless. The write-up suggests some alternatives, and I would advise against all of them:

- Using administrative data is satisfactory only if states all collect the same information, use the same definitions, have a low incidence of missing data, and the data are quality controlled. It is unwise to assume that these conditions are met.
- Using extant MIS is subject to the same concerns raised in the prior bullet.
- Participant survey data are often useful, but they are problematic for many reasons. In the simplest case of income, for example, the JTPA evaluation found large discrepancies between administrative and survey data. With regard to services and program participation, participants often do not know the program they are in, nor do they know how to categorize the services received.
- Site visits and focus groups are useful, but they are not a substitute for a MIS.

The write-up never discusses measuring services received by the control or comparison group. Collecting this information is critical for several reasons. First, it is important to document what alternative sources of services are available for SSI children who are not in the treatment group. Second, without understanding the services received by the control or comparison group, it is impossible to interpret the impact evaluation findings because we do not know what the treatment is being compared to.

3. Evaluation design. *Given the constraints implied by the PROMISE framework, what experimental or non-experimental evaluation design(s) would produce internally valid impact estimates? What other features should the design(s) incorporate?*

This issue focuses on who should be eligible for random assignment. One important issue is whether all those eligible should be randomly assigned or if only volunteers should be randomly assigned. A second issue is whether all families with SSI youth should be subject to random assignment or if certain groups should be selected.

There are arguments for both the strategies of randomly assigning all eligible or only assigning volunteers. One advantage of randomly assigning all those eligible is that it is the only strategy that permits estimation of the participation rate. However, the proportion of eligibles receiving the treatment is likely to be very small, so nonexperimental methods will have to be used to estimate the impact of the treatment on the treated. There could also be operational problems in assigning a large population to the treatment group and then dealing with the fact that only a small proportion can be served. Another potential problem is that policy officials may confuse intent-to-treat impact estimates with treatment on the treated estimates. If random assignment is only used among those interested in the treatment, the problems mentioned above are reduced. One could estimate participation rates as the proportion of those eligible who choose to volunteer for the program. Note that both approaches provide estimates of participation or interest rates as well as estimates of the impact of the treatment on the treated. I have a mild preference for conducting random assignment among volunteers as you would not be required to collect a massive amount of data on people who have no interest in and do not receive the treatment, and this approach is less likely to lead to wrong inferences by people who see the intent-to-treat results.

The write-up also deals with the possibility of restricting the project to youth and families with particular characteristics. Unless there is no interest in results for youth and families with particular characteristics, this approach strikes me as ill advised, primarily because it greatly reduces external validity with no clear payoff.

The write-up also discusses the possibility of eliminating a baseline survey to save money. Although this strategy would permit a larger sample, it does not seem like a wise strategy to me for several reasons. First, without the baseline survey, it will be impossible to know if the treatment and control groups are well matched and to control for any differences through the use of regression analysis to mitigate the potential bias due to sampling error. Second, data from the baseline survey can be used in the impact regressions and reduce the minimum detectable effect (MDE) for the impact study. Thus, by having a baseline survey, we reduce the size of the sample needed.

4. External validity. *To what extent, and how best, can results from multiple projects, perhaps with differing interventions, conducted in various states or localities selected through a process that might not yield areas representative of the country as a whole be generalized to support the development of national policy?*

The write-up for this section raises a serious issue, but I would not characterize the issue as external validity. The key point raised in the write-up is that given the structure of the project and likely available funding, an evaluation will not be able to determine the effect of the strategies used to generate savings to the SSI program. I agree with the assessment, but I would characterize it as an internal validity problem, not an external validity problem. When viewed in this context, the solution seems obvious—include sufficient funding to conduct at least one long-term follow-up using only administrative data. This is essentially the strategy used in the evaluation of the Job Corps by Mathematica Policy Research. Given the experience with the Job Corps evaluation, where the long-

term impact evaluation reversed the results of the intermediate-term evaluation, it is of paramount importance that a long-term impact evaluation be built in to the evaluation plan.²

² In addition to the Job Corps evaluation example noted above, Mathematica’s evaluation of the WIA voucher experiment experienced a change in the findings regarding which voucher strategy produces the greatest earnings impact.

PROMISE POSTMEETING COMMENTS FROM TAP MEMBER B

1. Interaction between PROMISE incentive payments and the local interventions. *How should SSA structure the incentive payments to promote the intended efforts by service providers and outcomes for participants?*

Following our discussions last week I am even less in favor of most of the incentive payments (IP) suggested by PROMISE, especially those for intermediate (pre-SSI/DI savings) achievement, than before. Key reasons are:

- States, service providers and youth and their families need to be enthusiastically invested in this initiative because they recognize intrinsic value for them in its ultimate success. For states this value may be tied to an ultimate objective of a more productive citizenry and a reduction in public assistance rolls; for service providers an opportunity to better meet their organizational mission, enhance their reputation and gain additional support for provision of their services; for youth and their families an assurance that they will realize an enhanced quality of life through productive and fulfilling employment.
- If IP are provided for various behaviors or intermediate goal attainment, all parties (states, providers, youth and their families) will inevitably (consciously or unconsciously) be managing their efforts to achievement of the short term benchmarks by which IP are attained, and will be at risk of losing sight of (or interest in) the ultimate goal.
- Incentivizing some behaviors/interventions/outcomes and not others is very likely to bias the actions and interventions of all parties to the project which, it seems to me, would render the evaluation of project results somewhat less useful.
- When IP are an important element in encouraging behavioral change, there is significant risk that when the IP end (as they inevitably will) then so will the behavior.

It's much more important that PROMISE assure that all necessary services and interventions are adequately funded and that a robust evaluation of the initiative be done that assures significant learnings. It is in those areas that I would focus project resources.

Recognizing that a reason for IP is to steer project grantees and participants toward interventions and intermediate goals that the funder feels are critical to ultimate success, it is important to note that there are, in the absence of IP, still some significant levers available toward that end—in particular, those might include the development of the RFP itself, the review of proposals and selection of grantees, and an annual determination as to whether a grantee deserves continued funding.

As an aside . . . it's important to remember that by far the most at-risk participants in this initiative will be the youth and their family who, as is recognized be all, may be dependent on their SSI/DI benefits as a critical survival mechanism. It will be vital that PROMISE can assure them that their participation in this effort holds substantive opportunity for a brighter future—that they will profit from this participation, while being protected against any reduction in their circumstance on the way to that brighter future. Not to be able to deliver on this has both practical and ethical implications.

Finally, I have no problem with (in fact would favor) the idea of IP tied directly to reductions in SSI/DI payments as youth (and/or their families) come off the rolls or are eligible for reduced benefits as a result of their employment status. Those payments should go directly to the service providers and may encourage them to maintain support of SSI/DI recipients beyond the life of PROMISE—support that may be critical to their continued employment success.

2. Development and measurement of partnerships and interventions. *What incentives will service providers and state and local agencies have to work together to efficiently deliver the intended services? How should interagency cooperation and the delivery of services be measured?*

The primary incentive to encourage partnerships will be the conviction on the part of various potential partners that such collaboration will help them deliver outcomes that are critical to their mission. To the degree that they recognize an opportunity to enhance the work that they are charged with through cooperation with other agencies or providers, they will embrace cooperation (or at least try to make it work). But if they don't recognize that potential, they will be at least ambivalent toward such efforts (and may even be hostile).

What I think doesn't typically work is a mandate for such cooperation from the funder. My sense is that such mandates encourage grantees to meet the letter of the law relative to partnering, but may not have much effect on what actually happens in the trenches. There is no question that collaboration can be a powerful force relative to the effectiveness of an effort like this, but it's important to recognize that there are also a lot of dynamics that get in the way of them happening. Among them:

- Distrust
- Turf issues (who has control/who'll get credit)
- Competition for limited resources (funding in particular)
- Differences in organizational cultures and management styles
- Conflicting goals or objectives (perceived or real)

Perhaps the best approach is to encourage a collaborative approach in the RFP, starting with the development of the proposal and then try to judge, during the review process, how successful each proposing group was in that effort. If they collaborated effectively in writing the proposal, there's a likelihood they'll collaborate effectively in its implementation.

As for measurement of these partnerships and the services delivered by them, it seems important to develop an MIS that is consistent across all sites and all partners. It should be developed by the same organization that's going to be analyzing the data for the project, and perhaps that development shouldn't be finalized until grantees have been chosen so that it matches what the grantees say they'll be doing. The MIS needs to effectively capture all critical information relative both to interventions being delivered (many of which will be qualitative in nature) and outcomes achieved, but be simple enough that it doesn't create a tremendous amount of extra labor on the input side, and lends itself to analysis on the output side. I would think that the organization primarily responsible for delivering an intervention should do that input while inputs relative to outcomes should be entered by a lead agency to assure consistency. Perhaps inferences could be

drawn relative to how effectively collaboration is happening from what interventions are being input by whom.

3. Evaluation design. *Given the constraints implied by the PROMISE framework, what experimental or non-experimental evaluation design(s) would produce internally valid impact estimates? What other features should the design(s) incorporate?*

I offer my thoughts in this area with the caveat that I am not a statistician or a researcher and have no training in either of those areas. Nonetheless, a couple of thoughts for what they're worth:

- I agree fully that evaluation design should happen concurrent with intervention design.
- My inclination would be to use random assignment of youth and families who volunteer to be involved. The smaller sample size would allow for a more robust study of the impacts of the interventions, and conclusions drawn could potentially be stronger because the primary nonintervention-related variable—whether or not the youth and their family are committed to a positive employment outcome—would be removed.
- An important peripheral evaluation objective might be determination of how best to encourage/convince SSI/DI recipients to volunteer for an initiative like this. If commitment to positive employment outcomes are, as I believe, one of the strongest predictors of success, then it would be useful to know what would create that kind of commitment.
- I would not whittle the sample down any further into subgroups like disability type, age, redetermination status, etc. It seems to me that doing so would greatly impede the evaluation's ability to suggest broad, sweeping conclusions while eliminating the opportunity to examine whether there is any significant difference in the effect of interventions on these various subgroups.

4. External validity. *To what extent, and how best, can results from multiple projects, perhaps with differing interventions, conducted in various states or localities selected through a process that might not yield areas representative of the country as a whole be generalized to support the development of national policy?*

My guess is that PROMISE results will not be able to be generalized in a significant enough way to be the basis for national policy for all of the reasons stated in the question and at least one more: given our discussion last week, there are not going to be enough projects funded to yield results that can be generalized. Further, this effort to prepare minors across the country to move successfully beyond SSI/DI benefits in no way lends itself to a neat package of prescribed interventions. What works in LA will almost certainly be very different than what works in Poughkeepsie, just as what works for Jose will be very different from what works for Jeannette. And that may be the most important reminder of all—that success will be predicated on customized interventions appropriate to the individual(s) and their circumstance.

On the other hand, PROMISE projects should allow an examination of a variety of approaches to determine relative efficacy, which could yield, if not a set of cookie cutters, at least a selection of suggested recipes for the consideration of service providers in the future.

Further, I think there *are* a couple of related questions that are paramount to having success with this effort, that could be examined in the context of PROMISE projects. Those questions are:

- How do you effectively develop real enthusiasm in the SSI/DI recipient to work toward a successful employment outcome and, ultimately, a reduction in (or elimination of) their reliance on SSI/DI benefits?
- What adjustments or waivers in benefit parameters are most effective at making individuals and/or families feel assured that in taking the road to employment their risk is minimal and the opportunity significant?

Finally, I think it's very important that participants in this project be tracked over a much longer period than the life of PROMISE—relative to their earnings and the level of benefits they draw into the future. A meaningful cost-benefit analysis will only be possible with the perspective of time—significantly more time than five years. Ultimately, support for initiatives like these need to be based on demonstrated ROI . . . and a longer examination period is going to be necessary in order that ROI become demonstrable here.

5. Other issues. *What issues other than those identified above should SSA consider for PROMISE?*

My greatest concern for this initiative, as I voiced last week, is that it is trying to do too much, which could yield a couple of less than positive results:

- By tackling too many challenges all at once, none of them are really addressed in a meaningful way. One of my mantras is, “If you try to do all things for everyone you may wind up doing nothing much for anyone,” and I think that could be a risk here.
- If too much is undertaken with too many players, interventions and outcomes may wind up being so intertwined and complex that it is difficult to confidently draw *any* conclusions from them.

I absolutely agree that family issues can have a tremendous impact on youth relative to everything this initiative is about, and I support, in principle, the idea of helping the family on the road to productive employment and reduction in dependence on benefits. In reality, though, as soon as one is focused on helping not only the youth but also the family, life gets much more complicated. In a context where resources to provide help are very limited, I think the potential to create significant progress and, maybe, change the paradigm for the next generation of families lies in focusing on the youth.

PROMISE POSTMEETING COMMENTS FROM TAP MEMBER C

1. Interaction between PROMISE incentive payments and the local interventions. *How should SSA structure the incentive payments to promote the intended efforts by service providers and outcomes for participants?*

I would agree that linking incentive payments (IPs) to SSA savings will not be feasible for the PROMISE demonstration, because the ultimate outcomes of such savings are longer term. This does necessitate, then, that states (or SSA, in the context of the RFP itself) identify specific intermediate outcomes or milestones that should be achieved for these youth.

As part of the discussion on December 6th, questions were raised as to whether or not the incentives should be positioned at the state level or closer to service providers who work directly with SSI recipients. An approach that would incorporate both state-level and local service provider IP would be useful. State applicants would need to describe how additional flexibility to use funds in a coordinated way across agencies would result in improved outcomes. State applicants should also produce partnership agreements or MOUs (Memorandums of Understanding/Agreement) to articulate this level of coordination and how flexibility in the use of funds would support better intermediate and potentially longer-range outcomes. This may not involve IP benefits directly to states, but rather would engage them in demonstrating how increased levels of coordination and collaboration across agencies would support the PROMISE demonstration, as well as longer-term policies and strategies for achieving the desired outcomes. It is important to be watchful of states that propose to retain a significant amount of IP at the state agency level. It is critical that the IPs stay focused at the service provider level, aligned or tied to specific intermediate outcomes that will have an impact on longer-range goals. It is also important that IPs are triggered by and granted to service providers early enough to permit them to gain enough forward momentum to adequately serve youth. Perhaps something can be learned from the Ticket to Work Program and regulations that have improved the payments for both milestones and outcomes.

Families and individuals with disabilities themselves were also identified as potential recipients for IPs. An important part of that discussion did, however, focus more on the issue of attempting to remove disincentives (premature loss of SSI support, health care, cash assistance, etc.), rather than using direct cash incentive payments to families and/or youth (although if costs are incurred by families, some consideration should be given to addressing this issue with IPs). It is also important that the SSA RFP require states and their service providers to identify how parents and youth will be informed about the program and its incentive model. Working with families as early as possible in discussions regarding their child's SSI and longer-term employment goal is critical.

Intermediate outcomes need to be carefully identified and defined by applicants/grantees. These would be intermediate outcomes that have a high probability or likelihood of achieving longer-term employment and SSI reduction outcomes. It was suggested during the discussion, and I would strongly agree, that outcomes such as school achievement (focusing on reading and math literacy), school-based work experiences, high-school completion (including the receipt of a regular or standard high-school diploma, rather than a certificate), and postsecondary education (access to, successful participation in, and completion of a postsecondary education program with an earned credential) are all meaningful to future employment outcomes. SSA could make the argument that these types of intermediate outcomes should already be addressed at the secondary and postsecondary education levels, and need not be further financed with additional financial

incentives. In practice, however, this is not occurring, due to lack of adequate service coordination that supports individuals with more-intensive levels of support needs to achieve these intermediate goals and outcomes.

2. Development and measurement of partnerships and interventions. *What incentives will service providers and state and local agencies have to work together to efficiently deliver the intended services? How should interagency cooperation and the delivery of services be measured?*

It is important incentives be tied to what we know “works,” based on research evidence, to the extent possible. We concluded a study in 2007 of state vocational rehabilitation agency policies and practices concerning transition-aged youth. The following strategies for enhancing the development of partnerships and interventions were found:

- Two organizational characteristics appeared to consistently support and promote the transition to employment for youth with disabilities—visible and strong state leadership and strategic assignment of vocational rehabilitation personnel to schools. Based on site visits, parents, youth with disabilities, and school personnel confirm the significance of the presence, availability, and support of vocational rehabilitation personnel in discussing transition goals, such as postsecondary education and employment at the school site.
- Our evaluation of the state vocational rehabilitation agencies confirmed that the utility of interagency agreements increases in proportion to the specificity of the agreement, the infrastructures created, and the degree to which the implementation of the agreement is monitored. Simply “agreeing to agree” is insufficient. Underscored also was the importance of the local interagency agreement, that articulated the level of collaboration between the vocational rehabilitation counselor, the school personnel, and service providers that would support the development of a student’s transition plan and provide the supports during the transition from school.
- While the most frequently provided vocational rehabilitation services to transition-aged youth with disabilities occur near or shortly after school exit, vocational rehabilitation agencies valued early engagement with eligible youth and coordinated IEP/IPE planning. Moreover, there were examples of agencies, such as those participating in our site visits, which provided eligible students with community-based vocational assessment, training, and paid work experiences while they are in school. This was all done on a collaborative arrangement and partnership with special education.
- We found significant variance in expenditures and average costs associated with vocational rehabilitation services provided to transition age youth across state agencies. The majority of the vocational rehabilitation agencies reported that current funding levels were insufficient to support required transition activities (no surprise). Even so, a few agencies use cost-shared positions with educational agencies to expand transition services. Alabama and Colorado demonstrated that jointly funded positions can and do benefit participating vocational rehabilitation and education agencies, as well as transition-aged youth and their families. Similarly, Vermont uses cooperative planning and service delivery with multiple partners to expand its availability of transition services and resources.

The critical point about these findings is that they specifically point to the importance of implementing and then measuring how agencies are coordinating service plans, leveraging resources, and mutually supporting families and youth with disabilities in achieving these varied outcomes. Documenting and measuring the mix of service intensity, duration, and cost, irrespective of who provides them, is essential.

There appears to be consensus, based on our discussions, that the RFP should promote state innovation and creativity in achieving desired results and outcomes. In promoting innovation, however, the challenge (as we all know) is that of sufficiently documenting the models, services, and practices provided and their relationship to outcomes. Requiring the submission of a logic model that provides evidence of a program theory, and a clear identification of the inputs into the model is essential.

The issue of whether or not an MIS reporting system should be developed is clearly challenging. I would agree with the statement in your notes (p. 4) that grants will be awarded to the applicants with the most evidence-based logic model, along with having a rigorous MIS. It may be cost-prohibitive, given the amount of funds available, to develop a comprehensive MIS for the demonstration program. I don't think it is useful to rely on existing administrative data from federal or state systems (RSA-911, etc.), due to data-sharing problems, reliability of data, and security issues (use of a common identifier for students/clients). I think that working directly with a grantee and using their existing MISs, and modifying them to ensure that critical measures are included in their agency data collection, may work. The evaluator would need to work very directly with the grantee (and across grantees) in modifying these agency-specific data collection systems.

3. Evaluation design. *Given the constraints implied by the PROMISE framework, what experimental or non-experimental evaluation design(s) would produce internally valid impact estimates? What other features should the design(s) incorporate?*

SSA's allocation of resources to the evaluation design and implementation will ultimately influence the type, level, and rigor of the evaluation. I strongly agree that appointing an evaluator at the same time as the Project Officer is important and will overcome some of the difficulties that were experienced in the YTD program. The group discussion supported the use of an experimental design and leaned toward the voluntary participation of families and youth with disabilities. An evaluation design similar to YTD, consisting of voluntary participation and aggressive recruitment seems feasible. I do believe that it is essential that a full follow-up survey be incorporated into the design. I also agree that it would be possible to have a second control group of youth, drawn from those who did not volunteer for the PROMISE intervention, to determine why they did not.

The option of choosing subpopulations by age range or disability type is problematic. Recruiting sufficient numbers of youth in certain subcategories, such as mental illness or others would intensify recruitment costs and local service provider efforts. Unless SSA is proposing that a certain subgroup of youth are more or less important to include in the demonstration, I don't think I would advocate for this approach (as interesting as this would be). In regard to the question of whether grantees should be selected to represent service-poor environments, I think this limits the opportunity to understand impacts across sites with varying conditions.

Ideally, if this were a 7–10 year demonstration project, I would advocate for engaging 14- to 15-year-olds and their families as participants. Early intervention is critical to informing and educating

families, developing IEP/transition plans that reflect the types of in-school experiences necessary to promote gainful adult employment, and engaging outside agencies in supporting these plans. However, I would advocate for working with 16- to 17-year-olds, at a minimum. I also agree that it would be unfair to ask grantees to show short-term impacts on families, since, as noted, changing family behavior is a slow and difficult process.

I would agree that the central elements of the evaluation must consist of a process analysis, impact analysis and cost-benefit analysis. Descriptive data concerning each state's context, economic conditions, mix of school and post-school services provided, family employment and earnings, and other services and benefits received also need to be well documented. The suggestion that the evaluation use only a follow-up survey and not a baseline survey, to limit costs, seems reasonable. I agree that there should be an embedded study to document the social and financial impact of the project on family life.

4. External validity. *To what extent, and how best, can results from multiple projects, perhaps with differing interventions, conducted in various states or localities selected through a process that might not yield areas representative of the country as a whole be generalized to support the development of national policy?*

From a methodological perspective, I am uncertain as to how to achieve external validity, if SSA does not fund the study at a level to include a national representative sample. The PROMISE evaluation could, however, certainly determine the potential for specific models and interventions to be replicated and scaled up in other states and/or communities within a state. Information from this type of evaluation that would be most meaningful include (1) complete descriptions of the state policies and practices used to positively impact intermediate and potentially longer-term employment outcomes (state policy changes, use of state and local interagency agreements, advisory councils, resource braiding and blending strategies used, challenges and barriers encountered and overcome, etc.); (2) a clear description of participant impact data (high school completion, postsecondary education participation, job acquisition, earnings, etc.); (3) detailed descriptions of the interventions and services used, including information on fidelity criteria; and (4) strategies used in achieving participant outcomes.

Despite the production of this good information on policies, practices, and service interventions that could potentially produce long-term employment outcomes for youth with disabilities, the question for me remains what SSA really envisions as the incentives for state participation. Since the net savings will primarily benefit the federal government, it is of interest to get a better sense of what SSA is thinking and the real incentives for state to participate. In that the PROMISE program is not testing a national model, but rather testing an array of service strategies and interventions that influence a common set of agreed-upon outcomes, there really isn't justification for a national policy per se.

PROMISE POSTMEETING COMMENTS FROM TAP MEMBER D

Member provided general comments about the meeting and the project; responses were not assigned to specific PROMISE criteria.

I believe the meeting was productive and that we had a sound discussion on many issues. However, I'd like to take the time to highlight now some additional thoughts I have which I think are important to the discussion.

We need to be cognizant about what is DIFFERENT about PROMISE compared to YTD for the following reasons:

- YTD is not yet completed and we don't know the results
- The language provided by congress specifically asks to work with families

Given this, I suggest that we think about the fact that SSI families are often families that have multigenerational poverty issues, and that we explore (or allow states to explore in the application process) interventions that also include the parents—both for their own employment and the employment of the youth. This argues for bringing in the TANF thinkers, including leadership and expertise from ACF in the Office of Planning Research and Evaluation—Naomi Goldstein and Nancye Campbell are good contacts—and OFA—Earl Johnson, director.

I also think we need to be putting more focus on the disability piece of the family equation—what is new and different or low income families who have a youth with a disability on SSI—what do families need in order to address the disability of the youth to successfully lead to employment? How do we improve outcomes based on the expectation that SSI youth can be employed? What interventions do families need in order to counteract the negative motivations around loss of the SSI check and to improve expectations for competitive integrated employment for youth?

Regarding building a demonstration based on a representative sample vs. a targeted population, upon further reflection I believe that perhaps given the amount of funding and the tight time period it would be appropriate for SSA/HHS/ED to think about a targeted population of SSI youth—instead of a representative sample. Different disabilities may demand specific strategies, and I think that being strategic in the ability to target groups (or provide guidance for states to target specific groups in their application process), especially given the five-year time period to demonstrate results, can be a helpful perimeter to establish. Applicants should be encouraged to provide evidence-based practices in their technical proposals.

PROMISE POSTMEETING COMMENTS FROM TAP MEMBER E

1. Interaction between PROMISE incentive payments and the local interventions. *How should SSA structure the incentive payments to promote the intended efforts by service providers and outcomes for participants?*

I think there was still a bit of confusion (at least on my part) at the end of the meeting about how closely the actual incentive payments needed to be tied to SSI benefit savings and the time frame over which the savings would be considered. I strongly encourage a broad view on both of these points. Since the age range wasn't clearly specified but certainly seemed to be intended to be minors (under age 18), it stands to reason that increased employment of the youth, which could potentially lead to decreased SSI payments, is only likely to occur over a matter of years. However, the savings over a lifetime could be dramatic. Taking a long view on savings would enable a larger amount of money to be used for incentive payments. In the same way, I think SSA should take a broad view of tying the incentive payments to specific outcomes. Requiring a short-term direct link (e.g., the activity reduces SSI benefits directly) suggests that only employment can really serve as an outcome. But I think it important to encourage other outcomes, in particular high school graduation. Making the case that an activity can lead to increased employment and decreased SSI benefits down the road so is worth investing in is important.

Some interesting issues arose around at what level the incentives should be given: to the individual, the state, the program providing services. I don't think the structure can be decided without knowing more about the individual demonstration programs, the outcomes, and the overall program goals. Incentive payments to individuals would be most direct, but seems very tricky when talking about minors—would have to go to families. Incentive payments to families for youth graduating high school or finishing a training program (etc.) could have influence, but incentive payments to families to have youth work seems more problematic. Incentives at the provider level seem good if the outcome on which the payment is based is clearly definable and achieved not too far away (in time) from the needed intervention. I was intrigued with the suggestion that providers holding tickets in TTW might be willing to wait for payment for some number of years. It seems to me that timing substantially reduces the incentive effect.

2. Development and measurement of partnerships and interventions. *What incentives will service providers and state and local agencies have to work together to efficiently deliver the intended services? How should interagency cooperation and the delivery of services be measured?*

This has to depend on the intervention—whether partnerships are required, which ones, and how they will look. One could require certain “players” be involved and assess whether that is meaningful instead of just lip service. A good implementation analysis will tell you about this. The incentives should be around outcomes, not around partnerships. I also think outcomes not delivery of service should be rigorously measured. Implementation analysis should give a good picture of what the services/program is.

3. Evaluation design. *Given the constraints implied by the PROMISE framework, what experimental or non-experimental evaluation design(s) would produce internally valid impact estimates? What other features should the design(s) incorporate?*

As in many of these projects with multiple potential sites, the desire to foster program innovation and allow for sites to do whatever is best and most innovative is at odds with the need for a sizeable sample that would depend on pooling across sites in a meaningful way. If each site is a mini-separate evaluation, what do we learn with essentially one data point on a program intervention impact? Something in between (allowing variation in models but at least with similar outcome measures?) would give a better sense. For example, “here are three sites that tried to improve graduation rates and here is the impact.” This discussion is really a program design issue rather than an evaluation design issue. I also heard the cost concerns of having salient outcomes be based on administrative data rather than surveys.

To meet the needs of a long time-window for measuring SSI savings (discussed above), I liked the suggestion of a design that allows for longer-term outcome measurement. This design would measure “interim” outcomes (which could be training or educational attainment, etc.) during the several years of the actual project, and then, using links to admin data on employment and SSI receipt, measure these outcomes to determine impact for out-years. Of course this has the disadvantage of not getting the “punch line” during the project window. I think if the interim outcomes are factors that are of a lot of concern and readily accepted as important in and of themselves, this could be interesting.

4. External validity. *To what extent, and how best, can results from multiple projects, perhaps with differing interventions, conducted in various states or localities selected through a process that might not yield areas representative of the country as a whole be generalized to support the development of national policy?*

If the programs are really different, I think an impact evaluation of two to three sites gives us rigorous case studies. That is, a finding of impact in one state/site is one data point on a program that works. It would be more persuasive if there is some amount of similarity of purpose/outcomes/intervention across the sites. Even with only a few sites, it would be nice if there could be a test of whether a big intervention (a sizeable amount of money) could have a big impact.

5. Other issues. *What issues other than those identified above should SSA consider for PROMISE?*

One issue generally is what the practical impact on evaluation and design is of the idea of having the family as a target as well as the youth. This is an interesting idea that was brought up as one goal of the program, and families are very important parts of the path to independence for minor youth. But it did not seem to be well thought out in the context of this demonstration. Does this mean that a program could exclusively target raising parental income to reduce SSI benefit receipt (eligibility) of the youth? This seems fundamentally different than preparing the youth for a future of work on his/her own—but both could lead to reduction in SSI benefits. This is not only an important conceptual issue for what the program wants to be; it would have real implications for evaluation design—even what is the unit of analysis. I suspect a better way to implement this idea is to focus on the youth, but provide explicit services/information, etc., to the parent—in other words, take the parent’s role into account.

I still do not understand how the age-18 redetermination plays into this project. From what we know, about one-third of these minors will have a reduction in SSI benefits after 18 (or 22) simply from the redetermination, not from anything the program will have done. A random assignment design will take this into account (i.e., the controls will also have this decline). But it still seems worth considering, especially if you randomize on those who volunteer and volunteers are drawn more heavily (as I would suspect) from those who are more likely to lose benefits (they are in better health on average, for instance). Does it mean something for what the effective sample size is—the sample that can have a decline in SSI benefits due to the program?

PROMISE POSTMEETING COMMENTS FROM TAP MEMBER F

1. Interaction between PROMISE incentive payments and the local interventions. *How should SSA structure the incentive payments to promote the intended efforts by service providers and outcomes for participants?*

1.2 Who should be eligible to receive incentives—private providers, state agencies, and/or participants?

- Incentive payments may be most effective when provided to participants and their families. Improved coordination among VR agencies may result in enhanced reporting for SSA reimbursements. Similarly, state and local agencies supporting youth with disabilities and their families may benefit from some incentive support, though, again, incentives paid to SSI participants directly may most encourage involvement in employment activities.
- Considering findings from Hemmeter, Kauff, and Wittenburg (2009), the initiative may need to establish some sort of promise or agreement that their participation will not result in loss of SSI or Medicaid eligibility. That is, the odds of being on SSI at age 19 for child SSI recipients who had worked were 42 percent lower than for recipients without employment experience. Furthermore, for those who did not move into the adult SSI program, only about 60 percent were employed at age nineteen (Ibid). These authors also concluded in a 2008 paper that, “while some youth appear to be successfully transitioning from child SSI benefits to adult benefits or other activities, some consideration might be given to extending supports for those who appear to have limited prospects for long-term self-sufficiency, particularly those with mental impairments” (p. 5). Steps should be taken to ensure that participation in PROMISE activities do not take away vital supports for a vulnerable population at risk of negative consequences if removed from the SSI program prior to receiving adequate preparation for work and independent living.

1.7 Should the structure for incentive payments be the same across sites, or should it be allowed to differ?

- There will likely be substantial variation across participants and sites. To the extent possible, controlling for these variations and applying similar incentive payments across sites would help with implementation fidelity and evaluation.

2. Development and measurement of partnerships and interventions. *What incentives will service providers and state and local agencies have to work together to efficiently deliver the intended services? How should interagency cooperation and the delivery of services be measured?*

2.1 To what extent should the evaluation track project services as part of a process study?

- Service types, duration, and quality, as well as education and training, should be examined to the extent possible.

2.2 What project services should the evaluation track (that is, which are most important)?

- Job finding, placement, training, follow-along, and secondary and postsecondary education experiences and attainment would seem important.

2.4 To what extent should the evaluation track project partnerships (the involvement of different agencies and providers) as part of a process study?

- Project partnerships, including describing and evaluating involvement, would also seem important.

2.5 What should a management information system for PROMISE track and for what purposes? What steps can ensure the accuracy and minimize the burden of data collection and entry?

- Using extant data collection systems such as the RSA-911, TRF, and other routinely maintained data systems may help reduce burden as well as provide opportunities for data verification. Agreements for linked data between federal agencies already exist, provide reliable information, and are cost-effective as compared to developing and implementing new survey designs.

3. Evaluation design. *Given the constraints implied by the PROMISE framework, what experimental or non-experimental evaluation design(s) would produce internally valid impact estimates? What other features should the design(s) incorporate?*

3.1 What key research questions should the PROMISE evaluation answer?

- To what extent do incentive payments, services, supports, education, and training increase earnings and family income among SSI transition-age participants over time? How do family supports and services influence earnings and income among SSI transition-age participants over time?

3.2 Should random assignment be implemented in each PROMISE site? Is it essential for a credible evaluation? Would requiring random assignment limit the nature of the innovations to be tested in any significant ways? What ways?

- Random assignment would be preferable.

3.3 If a nonexperimental design is adopted, what is the best approach for selecting a comparison group? That is, what is the best way to develop an understanding of the counterfactual?

- Qualitative methods such as interviews and focus groups may help identify key issues regarding employment experiences as well as what seems to work or not work with project implementation. Perspectives of partners and employers may also provide insight and guide project and/or future interventions.

3.4 Which youth outcomes (e.g., earnings, post-secondary education enrollment, receipt of SSI benefits, use of SSI work incentives) are most important?

- The specified outcomes all seem important.

3.5 Should targeted outcomes for the evaluation be permitted to vary across sites?

- Perhaps core outcomes could be identified along with targeted outcomes unique to local systems.

3.6 Should evaluation designs be permitted to vary across sites? That is, could one site have random assignment for individuals, another site have random assignment for sites within a state, and a third use a nonexperimental design?

- Increasing the sample size for random assignment may require consistency across sites.

3.7 What are the key sources of potential bias in impact estimates about which SSA should be aware, given the adequacy of the design to control for unobserved factors?

- State/local per capita income, unemployment and poverty rates in addition to sociodemographic, disability, and other factors among individual participants and their families.

3.8 What number of subjects and/or sites would be large enough to detect meaningful policy impacts? At the project level, are there feasible implementation changes that could substantially improve power?

- Recommend a power analysis to determine necessary number of sites and participants.

4. External validity. *To what extent, and how best, can results from multiple projects, perhaps with differing interventions, conducted in various states or localities selected through a process that might not yield areas representative of the country as a whole be generalized to support the development of national policy?*

4.1 Should grantees be allowed to restrict eligibility to subgroups of SSI youth? If so, what types of restrictions should be allowed?

- Random assignment during recruitment and participation. Oversampling may be needed for low-incidence disabilities and other subgroup populations.

4.2 Should the eligibility criteria be allowed to vary across projects?

- Same across sites to the extent possible.

4.3 Should participation in PROMISE be mandatory or voluntary for SSI youth? If voluntary, how best can the experiences of volunteers, obtained via the recruitment process(es) used in the demonstration, be generalized for SSI youth in other sites or nationally? If mandatory, what should be the consequences for nonparticipation?

- Results should reflect characteristics and factors among participants interested in working. Examining domains of self-determination and/or motivation may potentially serve as predictors of employment.

4.4 What features of the state and local environments will make it more or less likely for projects to achieve their intended outcomes?

- See 3.7. In addition, the funding and vitality of state and local agencies and organizations involved in the project may have impacts.

PROMISE POSTMEETING COMMENTS FROM TAP MEMBER G

Member provided general comments about the meeting and the project; responses were not assigned to specific PROMISE criteria.

I think the most useful thing I can do for more comments is to sketch out how an overall PROMISE initiative might work budget-wise. So here goes:

Suppose Congress appropriates \$30 million a year for five years for PROMISE (\$30 million was the amount requested in the President's budget for the Education Department to use on this project for grants to states).

This could fund

- Grants to two states of \$10 million a year each to implement full-blown pay-for-success contracts with service providers. The intention would be to continue to fund the chosen states for all five years. These would be contracts where services are delivered up front and payments are made by the states to service providers two or three years later, based on the extent to which performance objectives are achieved. Service providers would likely raise capital from private investors to fund operating costs in advance of receiving the performance-based payments. If we assume service costs of about \$5000 per participant, then each state could write pay-for-success contracts to obtain services for about 2000 individuals per year.
- A \$10 million pot to make smaller, more traditional grants to states that come up with plans to make better use of existing funding streams. Presumably these grants would be combined with waivers to let states make more flexible use of existing funding streams if they have well thought out ways to do so. What we would hope to receive from states would be plans to do better case management and tying together of existing services, better staffing and coordination within state government, and perhaps some innovative programming and added services if there are clearly identified service gaps.

MATHEMATICA
Policy Research

www.mathematica-mpr.com

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC

Mathematica® is a registered trademark of Mathematica Policy Research