Lessons from the Field

Megan J. McCormick, Kelsey Chesnut, and M.C. Bradley
Mathematica

April 2023
OPRE Report # 2023-121

Innovation and Support for Assessing Summative Evaluation Readiness: Lessons Learned

The Children's Bureau, within the Administration for Children and Families (U.S. Department of Health and Human Services), is funding a multiphase grant program to build the evidence base on what works to prevent homelessness among youth and young adults who have been involved in the child welfare system. This program is called Youth At-Risk of Homelessness (YARH). YARH focuses on three populations: (1) adolescents who enter foster care from ages 14 to 17, (2) young adults aging out of foster care, and (3) homeless youth and young adults up to age 21 with foster care histories.

Eighteen organizations received funding for the first phase (YARH-1), a two-year planning grant (2013–2015). Grantees used the planning period to conduct data analyses to help them understand their local population and develop a comprehensive service model to improve youth outcomes related to housing, education and employment, social-emotional well-being, and permanent connections. Six of those organizations received funding to refine and test their comprehensive service models during the second phase (YARH-2), a four-year initial implementation grant (2015–2019). During the third phase (YARH-3, 2019–2028), Mathematica is continuing to support the YARH-2 grantees (also known as sites) in building and disseminating evidence related to grantees' comprehensive service models. In addition, Mathematica designed and is implementing a federally-led summative evaluation of Colorado's Pathways to Success comprehensive service model.

This brief describes the multiphase evidence-building process YARH undertook to select the Colorado Pathways to Success comprehensive service model for the YARH-3 summative evaluation and to support YARH-2 grantees not yet ready for summative evaluation in identifying lessons learned and possible next evidence-building steps. This brief should be of interest to evaluators, funders, child welfare agencies, and program developers and implementers.

For more information on YARH, please see https://www.acf.hhs.gov/opre/project/building-capacity-evaluate-interventions-youth/young-adults-child-welfare-involvement

Throughout YARH, the Children's Bureau (CB) and Mathematica have supported grantees in developing and fine-tuning interventions and thinking about what kind of evidence they could produce about the effectiveness of their interventions. This work is culminating in a federally-led summative evaluation of Colorado's Pathways to Success comprehensive service model, the result of a multistep process Mathematica led to identify and recommend a YARH site prepared for rigorous evaluation. YARH has also provided continued evidence-building and dissemination support to grantees who were not prepared for a summative evaluation.

This brief highlights YARH's multiphase evidence-building process, and several lessons learned such as the time and purposeful planning needed to build evidence, the need to balance providing services with assessing grantees' readiness for a rigorous summative evaluation, and the need to be transparent about how grantees should demonstrate readiness for a summative evaluation.

History of YARH

YARH seeks to guide grantees along an evidence-building continuum (Exhibit 1). In the first phase of the grant program (2013 to 2015, known as YARH-1), 18 grantees received two-year planning grants to understand the characteristics of the three primary populations for YARH (Exhibit 2), develop partnerships with the child welfare system and teaming

structures, and begin designing comprehensive service models to prevent homelessness among youth and young adults who have been involved in the child welfare system.

Exhibit 1. Evidence-building path in YARH
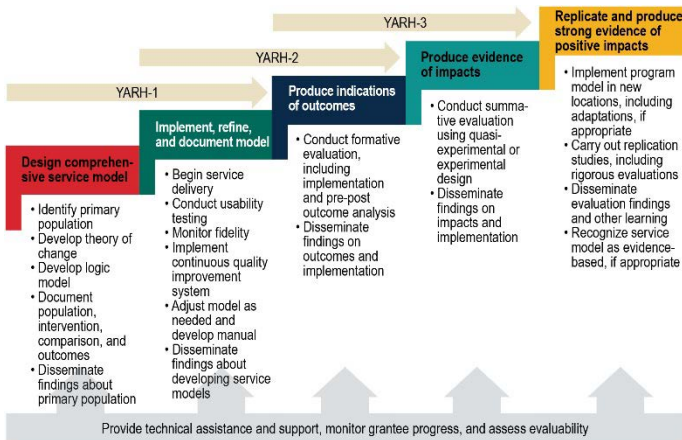
Adapted from: Langford, B.H., M. Flynn-Khan, and B. S. Lyght. *Investing in Evidence-Based Approaches for Youth Transitioning Out of Foster Care.* n.d. Available at http://www.ytfg.org/wp-content/uploads/2015/02/FCWG_Funder_Guide_Investing.pdf

Exhibit 2. Primary YARH populations
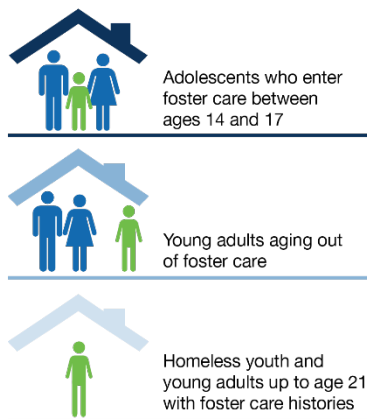
Grantees made varying amounts of progress toward defining a model during the YARH-1 planning grant. The [YARH-1 process study](#) documents the activities and progress of grantees during the planning period. Several factors affected grantees' progress, such as access to quality data, the extent to which grantees made progress on data analysis, and the structure of the planning team (Stagner et al. 2017).

As YARH-1 concluded, CB invited the 18 YARH-1 grantees to apply for YARH-2 funding, which supported initial implementation and refinement of the proposed model. Ultimately, CB awarded implementation grants to six YARH-1 grantees using criteria described in the funding opportunity announcement to evaluate applications (Administration for Children and Families 2015). The six YARH-2 grantees included: Alameda County, California; Colorado; Lighthouse Youth & Family Services (Cincinnati, Ohio); New Jersey; University of Maryland; and Westchester County, New York (Exhibit 3).

Exhibit 3. Location of the YARH grantees
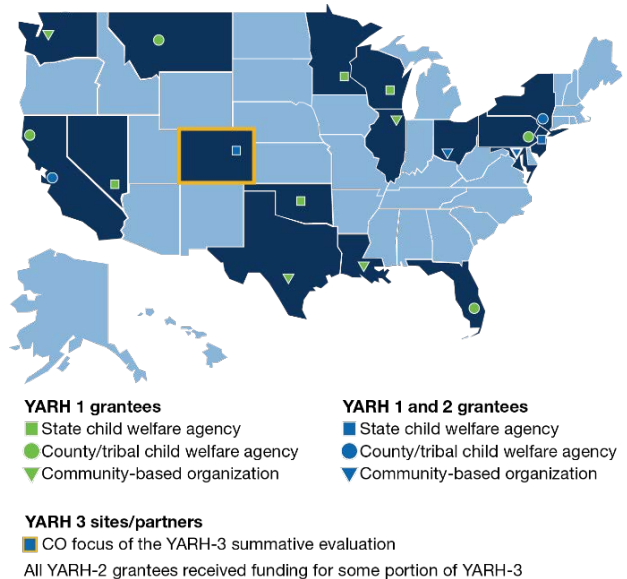
During YARH-2, grantees further specified their comprehensive service models, began delivering services, completed usability testing of key components of the service models, and conducted formative evaluations to assess program implementation and early outcomes for youth served. Findings from the [YARH-2 process study](#) identified recommendations for implementing models intended to reduce homelessness among youth and young adults with foster care histories and demonstrated that youth engagement with services is a key challenge and an important factor in successfully implementing the models (Keith et al. 2020).

The Office of Planning, Research, and Evaluation (OPRE) contracted with Mathematica in YARH-1 and YARH-2 to provide evaluation technical assistance (TA) to grantees, support them in articulating and refining the design of their service models, assess the evaluability of each service model, and disseminate the knowledge developed. Throughout YARH-1, Mathematica provided group evaluation TA to grantees, including webinars, conference presentations, and peer-learning conference calls. During YARH-2, each grantee worked with two dedicated TA liaisons who held monthly calls to support grantees in completing templates of the stages of work needed to prepare their comprehensive service models for a summative evaluation. For more information on evaluation TA provided to grantees in YARH-1 and YARH-2, see the [Strengthening Grantee Capacity Through Evaluation Technical Assistance](#) issue brief.

Current context: YARH-3

In 2019, ACF again contracted with Mathematica for the third phase of YARH (2019 – 2028, known as YARH-3), which provides information to the field on how to better serve youth and young adults through a rigorous summative evaluation. YARH-3 incorporates assessments of grantees' readiness for summative evaluation; a federally-led evaluation of one comprehensive service model (conducted in partnership with the state of Colorado, a YARH-1 and YARH-2 grantee), including an implementation study and an impact study; and ongoing dissemination of knowledge gained through project activities.

Exhibit 4. YARH key outcomes

For the YARH-3 summative evaluation, the Administration for Children and Families (ACF) intended to select at least one YARH-2 intervention that was likely to produce useful evidence about preventing homelessness and improving key outcomes (Exhibit 4) among youth and young adults who have been involved in the child welfare system. To support selection of the summative evaluation intervention, ACF and Mathematica convened two evidence-building meetings in early 2020 with consultants from the field.

ACF and Mathematica also convened an evidence-building meeting with consultants in late 2020 to provide feedback to and identify opportunities for the YARH sites not participating in the summative evaluation to continue their evidence-building work.

Big E Meetings: Selecting the YARH-3 summative evaluation site

Big E: Evidence derived from rigorous quasi-experimental or experimental studies.

Little e: Evidence derived from studies that are not quasi-experimental or experimental.

During the first year of YARH-3, ACF and Mathematica engaged consultants from the field in two evidence-building meetings that focused on "Big E"—that is, evidence derived

from rigorous quasi-experimental or experimental studies. These meetings served two purposes:

1. Provide feedback to the YARH sites preparing for the YARH-3 summative evaluation on ways to strengthen their readiness.

2. Make a recommendation to ACF leaders about which YARH site(s) were best positioned to move forward as part of the YARH-3 summative evaluation

Through a series of evaluability assessments and meetings with grantees at the end of YARH-2, Mathematica recommended which of the YARH-2 grantees had the potential to produce Big E, rigorous quasi-experimental or experimental evidence, the goal of the YARH-3 summative evaluation. These sites shared their interventions and evidence on their readiness for participating in a summative evaluation during Big E meetings with ACF, Mathematica, and consultants.

During the Big E meetings, Mathematica, ACF, and 10 consultants examined the interventions, refined potential study designs for participation in the rigorous summative evaluation, and evaluated which site(s) were well positioned to move forward with the YARH-3 summative evaluation. One of two key questions guided each Big E meeting:

1. *Could* the YARH site be a part of a summative evaluation?

2. *Should* the YARH site be part of a summative evaluation?

Big E Meeting 1

In January 2020, Mathematica hosted the first Big E meeting,– focusing on the "could" question which explores the technical or mechanical aspects of being in a summative evaluation. These issues guided meeting discussions:

- Definition of intervention
- Implementation fidelity
- Initial findings from YARH-2 formative evaluations
- Organization capacity
- The potential for the study to produce a valid test of program effectiveness

Before the first meeting, Mathematica sent participants the meeting agendas, critical questions to consider (Exhibit 5), and a package of documents for each site, including the readiness assessment, intervention manual, lessons from the formative evaluation, letters expressing interest in participating in a federally-led summative evaluation, and an initial summative evaluation design from Mathematica.

The meeting began with a broad discussion about preventing homelessness among youth and young adults with foster care histories. Participants reflected on what the field knows about (1) serving transition-age youth and preventing homelessness in this population, (2) outcomes (both desirable and undesirable), (3) what types of evidence programs should seek to create, and (4) where the most important gaps are in current evidence.

Following this discussion, consultants engaged in site-specific sessions. Only the site-specific team was present for the site-specific sessions. Each site presented on its comprehensive service model, and then Mathematica presented potential summative evaluation design options to evaluate each intervention.

Meeting participants discussed the strengths and areas of concern for each site's intervention and study design. The meeting concluded with a wrap-up discussion including all site teams in which the group discussed common themes across the sites, potential contributions to the field, and concerns about the summative evaluation. Following the meeting, Mathematica synthesized feedback from consultants to propose a list of action items for each site to focus on during their continued preparations to produce Big E as part of the YARH-3 summative evaluation.

Several key themes and next steps emerged from the first Big E meeting discussions.

1. **Formative evaluation.** Initial findings from sites indicated the usefulness of their interventions to the field. Site teams produced evidence that their models demonstrated positive changes in key outcomes for youth and young adults. Further exploration of data from YARH-2 formative evaluations could help sites gain additional insight into topics such as enrollment and engagement. Sites should explore variation within the data—for example, whether there are differences in fidelity by service delivery location or differences in enrollment by youth and young adult characteristics.

2. **Intervention fidelity**. Sites identified managing and maintaining consistent implementation of the intervention as a common challenge, especially as the programs expand to include more providers and serve more youth and young adults.

3. **Population of interest.** There are inherent tensions between the necessary sample size required to detect statistically significant impacts and the purpose of YARH: serving youth and young adults who are most at-risk, a hard-to-reach population. Much of the discussion in site sessions focused on screening and enrolling an adequate number of participants, in addition to potential concerns resulting from nonresponse bias or dropout.

4. **Engagement.** Youth and young adult engagement is a critical component of sites' interventions. However, sites need to refine definitions of youth engagement to promote consistent measurement of engagement across intervention and comparison conditions.

5. **Understanding the counterfactual.** Understanding the counterfactual is important for understanding the contrast in services in a summative evaluation. Consultants recommended that sites explore what the counterfactual might look like, what typical case management would look like, and what other programs offer case management in their respective service areas.

6. **Contributions to the field.** Sites demonstrated potential for contributing novel insights to the child welfare field. Sites had already begun to learn more about predictors of homelessness in their service areas. Interventions specifically engage adolescents and young adults (rather than youth of any age), are built around intensive case management models, and reflect the need for youth-driven services. This is important because the child welfare field struggles with how to meet the needs of the older youth and young adults it serves while using tools, like intensive case management, that are familiar to the field.

Big E Meeting 2

In March 2020, Mathematica hosted the second, two-day, Big E meeting—revisiting the "could" question and then focusing on the "should" question. The "should" question explores whether the:

1. Intervention would be of interest to the field and policymakers

2. Evaluation would be a good investment for the federal government

3. Evaluation is likely to produce a satisfying test of the effect of the program—that is, one that produces a credible, favorable, and statistically significant impact estimate on policy-relevant outcomes.

Like the first Big E meeting, federal and Mathematica staff, site teams, and 10 consultants attended the meeting to help determine the readiness of the sites for participation in a federally-led summative evaluation. Participants considered critical questions (Exhibit 6).

Exhibit 6. Big E Meeting 2 critical questions

- Is it an innovation that the field will have interest in?
- Do child welfare or homeless youth services use elements of the intervention currently?
- Could other locations implement the intervention with fidelity with relative ease?

Mathematica and ACF structured the meeting with an initial session that reoriented and updated participants on the YARH-3 summative evaluation and the goal of the meeting. The first day of the meeting began with introductions and a discussion about the status of the summative evaluation site selection process. Site-specific sessions followed, during which sites had the opportunity to provide additional data to help meeting participants determine whether the site *should* participate in the summative evaluation. Then, Mathematica elaborated on the site's potential study design, followed by a discussion among meeting participants on the strengths and areas of concern for the site based on the "should" critical questions. As in the first evidence-building meeting, the site teams were only present for their own site-specific session. The meeting concluded with a discussion of next steps with all site teams.

Only the consultants, federal staff, and Mathematica staff attended the second day of the second Big E meeting. Participants discussed options for continued evaluation activities in YARH-3. Four overarching factors, encompassing both "should" and "could" considerations, guided discussions for recommending a YARH-2 intervention for the summative evaluation:

1. **Interest of the field in the intervention.** The policy relevance of the proposed comprehensive service model was a key consideration for the recommendation. The extent to which researchers and practitioners would be interested in the results and the likelihood that other communities would implement similar service models informed whether an intervention would make a useful contribution to the field.

2. **Readiness of the intervention for a summative evaluation.** The clarity of the comprehensive service model and the accompanying program manual were key considerations for whether an intervention was ready for summative evaluation. In addition, the following qualities signaled readiness for the evaluation: professionalism of the site management teams, robustness of the sites' continuous quality improvement and fidelity monitoring processes, availability of administrative data for use in the analysis, and emerging findings from the formative evaluation regarding program fidelity and improving outcomes.

3. **Rigor of evidence that would result from the proposed design for the summative evaluation**. The potential credibility of the evidence from the proposed evaluation design was the third criterion. The team only considered designs that would produce a credible, internally valid test of program effectiveness.

4. **Likelihood of detecting statistically significant favorable impacts.** The statistical power and likelihood of the study to detect any favorable impacts of the program was the fourth consideration. The combination of the sample size available, research design proposed, expected counterfactual condition, and expected magnitude of changes in outcomes based on the dosage and service contrast contributed to the assessment of an intervention's readiness for a summative evaluation.

The Mathematica team used a variety of formats to engage meeting participants in considering these four factors, including a discussion about what's most important to consultants when considering these factors and an activity that prompted participants to list the top three strengths and areas of concern for each site's participation in a summative evaluation.

Summative Evaluation Recommendations

Following the second Big E meeting, Mathematica explored several paths forward and made several recommendations for ACF to consider:

Move forward with a summative evaluation of Colorado's Pathways to Success comprehensive service model. Though consultants agreed that there was interest from the field in multiple sites' interventions and multiple sites had summative evaluation plans with the potential to produce rigorous evidence, consultants assessed that Colorado was best positioned to participate in a summative evaluation in the near term. The greatest area of concern for other sites was low uptake of the intervention.

Augment the summative evaluation with additional analytical approaches, such as Bayesian interpretation and a difference-in-differences analysis using extant administrative data. Consultants had expressed concerns that a summative evaluation might fail to find statistically significant results because of low sample sizes. YARH sites had likely exhausted their abilities to expand their services to increase sample size. Augmenting the impact evaluation with a Bayesian interpretation of the results would make it more likely that the evaluation would provide useful information about the efficacy of the Pathways program, even if the primary impact results were not statistically significant. Adding a difference-in-differences analysis with administrative data offers another avenue for assessing the program that could build on the main impact analysis, potentially with a larger study sample for a more powerful test of program effectiveness.

Continue to support Alameda County in building evidence during YARH-3. Consultants indicated that there was interest in generating rigorous evidence of the efficacy of offering group-based therapeutic interventions, such as dialectical behavior therapy, to transition-age youth and young adults with a history of foster care involvement. Evidence from a successful summative evaluation would be highly useful to practitioners and policymakers. In addition, Alameda County site staff had clearly documented the comprehensive service model in its program manual. Research teams could describe the intervention condition clearly in future reports—and the thorough and clear program manual would help other sites replicate the model if future evidence suggested its effectiveness.

Little e Meeting: Continuing evidence-building work with other YARH sites

In October 2020, Mathematica led an evidence-building meeting that focused on so-called "Little e"—or evidence derived from studies that are not quasi-experimental or experimental. The sites that did not participate in the Big E meetings were the focus of these meetings as they continued to generate and disseminate useful evidence about the implementation and effectiveness of their comprehensive service models.

Site teams, Mathematica staff, and four consultants participated in the Little e meeting. The consultants provided a fresh perspective on the evidence sites had produced and identified ways to strengthen this evidence as sites continue move along the evidence-building continuum. The consultants also highlighted opportunities for sites to use evidence produced during YARH to benefit other practitioners, policymakers, and researchers.

The meeting had a brief introductory session, site-specific sessions, and a final session devoted to discussion of cross-site themes. Before the meeting, Mathematica provided the consultants with background materials on each site's comprehensive service model, including formative evaluation results. During each site-specific section, site staff summarized their comprehensive service model and initial formative evaluation results.

Several key themes emerged in discussions across all sites:

1. **Building evidence is a long-term process.** Evidence-building is long, hard, and expensive, and efficiency can depend on the strength of the evidence base at the beginning of the project. Long-term outcomes are valuable but can take years to materialize.

2. **Evaluation strategies should account for complex program contexts.** Outcomes for at-risk youth and young adults reflect the effects of trauma, racism, gender-based violence, and other systemic problems. It should not be surprising to find modest effect sizes for interventions operating in this context. Richer, more nuanced measures might be more appropriate than the binary measures evaluations often use. Further, though randomized controlled trials are the gold standard, it can be a misperception that evaluation strategies other than randomized controlled trials do not contribute meaningful evidence.

3. **Build the evidence chain by exploring links among intervention components, implementation, and outcomes**. Because of the complexity facing evaluations of child welfare interventions, meeting participants stressed the importance of building the chain of evidence by linking specific program components to specific short-term outcomes and then linking those short-term outcomes to long-term outcomes. For example, it would be useful to understand the relationship between program components such as transition planning and peer coaching and long-term outcomes. Evaluators can use programs' theories of change to understand which components likely relate to short-term outcomes, then build a pathway to long-term outcomes.

4. **Use research to advance equity initiatives.** Meeting participants agreed that addressing equity in program design and evaluation is critically important. Asking a series of "who" questions can be helpful: Who are we building the program for? Who does it work for? Whose

worldview informed the program and intended outcomes? Conducting studies with sample sizes large enough to detect differences in effectiveness across demographic subgroups is important for determining programs' effects on equity.

5. **There are challenges to building evidence in rural areas with small sample sizes.** Consultants agreed that there is no easy way to address some sample size challenges. One solution is to first conduct high quality qualitative research to fine-tune implementation and then evaluate a scaled-up version of the model. In addition, sample size problems in rural locations are not unique to the child welfare field; drawing lessons from other fields could enhance how we think about child welfare in a rural setting.

Lessons Learned

Grantees and funders are eager to participate in summative evaluations to understand what works, for whom, and how. However, it is impossible to know and evaluate every factor that will contribute to the success of a site's participation in a summative evaluation. YARH used four factors to evaluate summative evaluation site readiness, but there are inevitable unknowns that these criteria do not and cannot capture. For example, sites can predict strong sample sizes based on formative evaluations, but in practice, there might be unanticipated challenges with enrollment that hinder the evaluation's strength.

CB, OPRE, and Mathematica took a multistep approach to identifying a site for a summative evaluation, which was possible given the multiphase nature of YARH. This structure enabled grantees to develop and implement programs before committing to a large-scale evaluation. It also allowed time for the funder and summative evaluator to understand the strengths and challenges of each grantee. Using the evidence-building continuum as a framework for producing evidence helped sites consider a summative evaluation, while not rushing into one. Mathematica and ACF kept the summative evaluation in view from the start of YARH-1, but the evaluation TA helped sites focus on the steps to demonstrate readiness, rather than view the summative evaluation as a lofty goal.

Successfully supporting grantees requires careful consideration of evaluation-focused activities while simultaneously providing services. CB and OPRE balanced the need to assess progress and readiness for a summative evaluation with the burden such assessments place on grantees. Grantees focused on the program—its development and implementation—with the goal of improving the lives of youth and young adults in their communities. Collecting and reporting the evidence of readiness could seem like a burden, so CB and OPRE asked Mathematica to help think about what information was necessary and how to systematically collect this information from grantees while not burdening them unnecessarily.

Mathematica, OPRE, and CB needed time to support collaborative planning based on what they knew about grantees and their needs. Grantees were able to serve youth and young adults—and learn about what their intervention looked like in practice—while developing the data and documentation to help inform a future evaluation decision. Relationships between Mathematica liaisons and grantees helped identify useful supports for each grantee and encouraged discussion of grantees' challenges, including brainstorming possible ways to address the challenge or to build other evidence.

Transparency about what information would inform decisions about site readiness to participate in a summative evaluation was vital to the success of the multiphase program and the grantees. During YARH-1, grantees needed to demonstrate that they had support from partners, access to data, and an intervention that looked like it might promote housing stability. YARH-2 required more evidence—including clear articulation of the proposed intervention and how it was innovative, and data showing that the intervention was feasible and would begin to change the path for youth and young adults at risk of homelessness. Being selected for the YARH-3 summative evaluation included bigger hurdles; consultants needed to agree the intervention looked like it could work, the grantee needed to show possible sample sizes to ensure the summative evaluation had sufficient power, and the grantee and partners needed to commit to participating in a rigorous evaluation.

References

Administration for Children and Families. "Implementation Grants to Develop a Model Intervention for Youth/Young Adults with Child Welfare Involvement At-Risk of Homelessness." Washington, DC: U.S. Department of Health and Human Services, 2015.

Davis, L., N. Thoennes, and S. Matyasic. "Colorado Pathways to Success: Formative Evaluation Brief." Denver, CO: Center for Policy Research, December 2020.

Keith, R., N. Islam, R. Sarwar, and M.C. Bradley. "Reducing Homelessness Among Youth with Child Welfare Involvement: Phase II Implementation Experiences in a Multi-Phase Grant." OPRE Report 2020-129. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, September 2020.

Langford, B.H., M. Flynn-Khan, and B.S. Lyght. "Investing in Evidence-Based Approaches for Youth Transitioning Out of Foster Care." n.d. Available at https://www.ytfg.org/wp-content/uploads/2015/02/FCWG_Funder_Guide_Investing.pdf. Accessed May 30, 2019.

Packard Tucker, L., A. Dworsky, and M. Van Duren. "Formative Evaluation Lessons Learned: Evaluating the Implementation of the Youth Transitions Partnership in Alameda County." Chicago, IL: Chapin Hall at the University of Chicago, May 2020.

Stagner, M., L. Klein Vogel, E. Knas, N. Fung, J. Worthington, M.C. Bradley, A. D'Angelo, et al. "Reducing Homelessness Among Youth with Child Welfare Involvement: An Analysis of Phase I Planning Processes in a Multi-Phase Grant." OPRE Report # 2017-51. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, August 2017.