

A Toolkit for Evaluating Initiatives to Improve Child Care Quality



Institute for a Child Care Continuum
Bank Street College of Education
610 West 112th Street
New York, NY 10025

2003
Toni Porter
Sally Mabon
Shannon Kearns
Anne Robertson
J. Lee Kreader

Acknowledgements

The *Toolkit* is the second product from a three-year project funded by the Administration for Children and Families Child Care Bureau. The project was a collaboration in many ways: between the Child Care Bureau and the research team; among the research partners--the Institute for a Child Care Continuum at Bank Street College of Education, the National Center for Children in Poverty at Columbia University's Mailman School of Public Health, and Abt Associates Inc.; between state child care administrators and child care experts who served on the project's Advisory Committee; and with the state policy makers who reviewed and field tested the *Toolkit*.

The list of individuals who contributed to this effort is long. The Advisory Committee members--David Edie, Amparo Garcia, Jeannetta Green, Shari Gruber, Janet McKeon, Tom Olsen, Laura Satterfield, Mary Smithberger, Gwen Stephens, Deb Swenson-Klatt, Marty Zaslow--provided valuable direction about the *Toolkit's* contents as well as practical comments on the instruments in Part II. Ivelisse Martinez-Beck of the Child Care Bureau, another member of the Advisory Committee, offered helpful comments and advice about Part I as well. We also received useful suggestions from the policy makers who reviewed the preliminary draft. They include: Corinne Bennett, North Dakota; Carolyn Drugge, Maine; Pamela Henry, New Mexico; Bonnie Murray, Georgia; Lynette Rasmussen, Utah; Linda Saterfield, Illinois; and Reeva Sullivan-Murphy, Rhode Island. We are also grateful to the three state teams who not only reviewed the *Toolkit* but also participated in the field test. They are: Alice Wommack and Lana Messner, Kansas; Charles McMillon and Jamie Gottesman; Ohio; and Sonja Griffen and Cecelia Callison, Washington.

Thanks are also due to several member of our research team who worked on the project during various stages. At Bank Street, Sulaifa Habeeb helped to create the instruments and Rena Rice copy-edited Part I; Ann Collins contributed to the *Toolkit's* framework; and Fred Glantz from Abt shared his knowledge of theory of change models and summative evaluation.

Defining Child Care Quality

The issue of child care quality is high on the public policy agenda because millions of young children spend their days away from home while their parents are at work. This situation has generated growing concerns about the effects of non-parental care on children. Will children be safe from injury and harm in child care? How will it affect their development? What impact will child care have on children's readiness for school?

Research shows that good child care can be good for children (Phillips, 1995). Children in high quality care demonstrate improved cognitive and language skills. They are also better prepared for school (NICHD ECCRN, 1999) than children who are in child care that is rated poor.

What aspects of child care contribute to these results? Researchers group different characteristics of care into two broad categories: structural features and process features (Vandell & Wolfe, 2000). Structural features include the number of children in care, ratios of adults to children, and caregiver qualifications such as educational levels and special training in early childhood education. Process features include caregiver-child interactions, caregiver sensitivity and responsiveness, and children's interactions with other children as well as with the environment.

Studies have found a positive relationship between the number of children in care and adult-child ratios and children's cognitive and social development, because small group size and low adult-child ratios enable caregivers to provide more individual attention to children (NICHD, 2000; NICHD, 1999; Whitebook, Howes, & Phillips, 1989; Ruopp, Travers, Glantz, & Coelen, 1979). A third structural feature, caregiver education and training in early childhood, is also associated with positive outcomes for children, especially in terms of cognitive and language development.

New research suggests that some of these characteristics may be more influential than others. For example, it indicates that specific early childhood training may have a greater effect than caregivers' educational levels (Tout, Berry, & Zaslow, 2003; Peisner-Feinberg & Burchinal, 1997; Burchinal, Roberts, Nabors, & Bryant, 1996; Helburn, Culkin, Howes, Bryant, Clifford, Cryer, Peisner-Feinberg, & Kagan, 1995; Arnett, 1989). In addition, it found no consistent evidence that previous experience in child care, another structural variable, is linked to quality.

Studies also point to the effect of caregiver-child interactions on children's development. The findings indicate that positive interactions—responding to children with warmth and understanding, talking to and with children, engaging them in a variety of activities that fit their developmental stages, using appropriate disciplinary strategies—result in better socio-emotional, language and cognitive development (NICHD ECCRN, 1999). Children who have caregivers that interact with them in these kinds of ways are also better prepared for school (Pianta, 2003).

Other issues that are related to quality have been the object of research. One is the problem of retention and turnover in the child care workforce. Studies have linked low teacher wages to high turnover, which affects the stability and consistency of the care that children receive (Whitebook, Sakai, Gerber & Howes, 2001; Helburn et al., 1995; Whitebook, Howes, & Phillips, 1989). Wages are also associated with caregivers' educational levels: centers with teachers who had more education and training and were paid higher wages offered better quality care.

States can, and do, use these research findings as the basis for policies that aim to improve child care quality. Regulatory policy can be used to require smaller groups of children in center classes or family child care homes as well as to lower adult-child ratios in these settings. It can also be used to raise educational qualifications for caregivers--

requiring child care credentials, associates (AA) and bachelors (BA) degrees or special training in early childhood education for center teachers, increasing the hours of child development training required for family child care providers, or requiring health and safety training for providers who are exempt from regulation.

States can use research findings to influence subsidy policies as well. To increase providers' wages and reduce turnover, states can raise reimbursement rates. They can also encourage providers to enhance their knowledge and skills by tying reimbursement rates to educational levels or training. Subsidy policy can also be used to improve overall child care quality by linking reimbursement rates to different levels of quality on rating scales or to accreditation by professional organizations such as the National Association for the Education of Young Children or the National Association for Family Child Care.

Funding specific initiatives is another option for integrating research into practice. In our study of states' use of the Child Care Development

Fund (CCDF) quality set-aside, we found that states use a variety of strategies in their attempts to improve quality. Among them are supporting professional development activities to increase providers' education, increasing compensation to reduce turnover, and providing materials as well as technical assistance to improve the quality of the environment. Some states fund an array of initiatives to address individual aspects of quality for different populations of providers or children in different age groups, while others support efforts like provider education or compensation that are intended to improve the child care system as a whole.

There is a great deal at stake in efforts to improve child care quality, because they affect the lives of many children. How do policy makers know whether these initiatives have achieved their objectives? How can they determine which ones work well and which ones don't? How can they measure their results? The *Toolkit* aims to provide some answers to these questions.

Table of Contents

LIST OF TABLES	IV
LIST OF FIGURES	IV
PART I – THE GUIDE	1
Chapter I: Introduction	3
Chapter II: Evaluation Perspectives	5
Chapter III: Theory of Change Models	7
Chapter IV: Using a Theory of Change Model for Evaluation	9
Chapter V: Estimating the Impact of Quality Improvement Initiatives	13
Chapter VI: Designing an Evaluation Plan	15
Chapter VII: Conclusion	19
References	21
APPENDICES	
Appendix A: Selected Evaluation Descriptions	23
Appendix B: Field Test Results	33
Appendix C: Templates	41
Appendix D: Other Resources	45
Appendix E: Developing the <i>Toolkit</i>	49

The Issue

Evaluation research can answer a wide variety of questions. It can provide information about whether an initiative has reached the anticipated population, if it has delivered the planned services, and if it has had the desired results. This evidence can be used to determine if initiatives should be continued, modified or expanded. Evaluation can also be used to determine whether particular policies make a difference for children, if the benefits are equal to the costs, and whether resources should be added to other parts of the child care system to enhance the effects of specific strategies.

The interest in evaluating efforts to improve child care quality has grown enormously in the past several years, largely as a result of three factors. One is the CCDF quality set-aside, which requires states to spend 4% of their CCDF funds on efforts to improve child care quality. Another is increasing emphasis on accountability for public expenditures, especially for social services and education. And the third is growing attention to children's readiness for school.

Across the country, policy makers recognize that evaluation may provide the answers they need. State child care administrators want evidence to determine if their child care quality improvement initiatives have been effective. Members of the National Governors Association (NGA), the National Conference of State Legislatures and the Association of Public Human Services Administrators (APHSA) want to know how and whether these efforts have worked. Congress wants to know how states have used their CCDF quality set-aside funding (GAO, 2002).

This situation raises serious questions for state child care administrators. How can they evaluate their initiatives? What are the most important questions to ask? How can they obtain the answers? This *Toolkit* aims to help them address these issues.

What State Child Care Administrators Say about Evaluation¹

Evaluation can help you:

- Find out whether the initiative did what it was supposed to do and whether it made a difference in child care quality.
- Learn what you need to know to make decisions about funding specific initiatives.
- Determine the impact of the initiative on resources for child care and help you establish whether to maintain the investment if the other resources are not available to support it.

What the GAO Says About Evaluation

In a 2002 report, the General Accounting Office (GAO) examined how states used their CCDF to support quality (*States Have Undertaken a Variety of Quality Improvement Initiatives, but More Evaluations of Effectiveness Are Needed*). Senators Kennedy (Massachusetts) and Dodd (Connecticut) and Congressman Reed (Rhode Island) who commissioned the report, also asked the GAO to identify whether states evaluated the effectiveness of these initiatives. In its review, the GAO used the National Research Council standard that studies of quality improvement initiatives should assess the effects on children's development.

- "Few states have evaluated the effects of their quality improvement initiatives on children's development, [but] some studies provide useful findings about them" (p.4).
- Of the evaluations, "a few had study designs that isolated an initiative's effect and survey response rates that provided reliable estimates" (p.24).
- Only three studies "had methodological approaches sufficient to produce conclusive findings" (p.25), in other words, evaluations that used experimental or quasi-experimental design and that evaluated child outcomes or effects on attributes of child care quality.
- Several states conducted studies that "can provide reliable information that is needed to address program design issues" (p.26), although they did not use experimental or quasi-experimental designs.
- Four of five case study states identified a "trade-off" between conducting evaluation and providing services to more families.

¹ These comments are paraphrased from statements from state child care administrators who serve on the project's Advisory Committee.

What National Organizations Say about Evaluation

- A joint APHSA/Child Trends study found that states are interested in “using research in the planning process” effectively. Evaluation is necessary to decide “the highest priorities...for working on quality issues.”
- The NGA reports that states “want to know where to start” because evaluation has become an issue for governors to consider seriously.
- The National Child Care Information Center receives as many as 25 calls annually from technical assistants to state child care administrators, program managers, governors and policymakers about evaluation to “find out what other states are doing and how they are doing it,” because they want to develop evaluations for their own states.

The Purpose of A Toolkit For Evaluating Initiatives To Improve Child Care Quality

In 2000, Bank Street College of Education, Abt Associates Inc., and the National Center for Children in Poverty of the Mailman School of Public Health at Columbia University began a joint study, *Assessing Child Care Development Fund (CCDF) Investments in Child Care Quality*. A three-year effort funded by the Administration for Children and Families (ACF) Child Care Bureau, it is intended to provide information for policy makers about how to make more efficient use of CCDF quality set-aside funding. Our initial product, *A Study of Selected State Initiatives*, describes 104 initiatives that use a variety of strategies to improve child care for different populations of providers and different populations of children. The companion volume, *Program Profiles*, includes information about the start date, geographic scope, budget, outreach strategies, implementation and evaluation of each initiative.

A Toolkit for Evaluating Initiatives To Improve Child Care Quality presents the results of the second phase of our project. Its purpose is to help state child care administrators design and conduct evaluations of their quality improvement initiatives. The *Toolkit* discusses evaluation in the context of efforts to improve child care quality, and uses examples of specific initiatives to illustrate the evaluation process. It is intended to supplement step-by-step evaluation guides such as the ACF *Program Manager's Guide to Evaluation* (1997) for conducting evaluations of social programs.²

Organization of the Toolkit

The *Toolkit* is divided into two parts. The first part provides a framework for understanding and conducting evaluation of child care quality improvement initiatives. Chapter II offers a general perspective on the two types of evaluation—formative and summative. The following three chapters discuss evaluation in greater detail. Chapter III discusses the theory of change model as a basis for planning and evaluating child care quality improvement initiatives and Chapter IV illustrates how to use a theory of change model to design an evaluation. Chapter V presents some issues related to summative evaluation. The final chapter provides practical advice. It describes specific steps for preparing an evaluation plan based on the ACF *Guide* and the Institute for Women's Policy Research report, *Building a Stronger Child Care Workforce* (2002). The Appendices include other resources for evaluation.

The second part of the *Toolkit* consists of a set of instruments that can be used to evaluate child care quality improvement initiatives. The instruments, which are easy to use, are samples that can be adapted to evaluate these efforts. There is also a table that identifies outcomes for practitioners, programs, and children and families that are linked to specific quality improvement strategies. The Appendix includes an annotated bibliography of selected observation instruments that can be used to assess child care quality.

² Other guides include the W.W. Kellogg Foundation *Evaluation Handbook* (1998) and the United Way's *Measuring Program Outcomes: A Practical Approach* (1996).

The purpose of evaluation is to inform decision-making. There are two broad categories: formative and summative. These categories have different goals. Formative evaluations are intended to guide program improvement, while summative evaluations are intended to provide a judgment about aspects of a program's performance (Rossi, Freeman, & Lipsey, 1999). Evaluations can be used to describe a program or to analyze the results.

Formative Evaluation

Formative evaluation provides information that can be used to improve how a program works and whether it achieves its intended results. This information enables managers to assess the validity of their initial assumptions, to make mid-course corrections, and to adapt the program to changing circumstances.

Formative evaluation typically focuses on intermediate outcomes rather than the long-term outcomes, or effects, that are the ultimate goal of the initiative or program. For example, a formative evaluation of an initiative that aims to increase the number of center teachers with Child Development Associate credentials (CDAs) or AAs by providing scholarships for college courses might examine the following intermediate outcomes:

- The number of providers who apply for scholarships each month;
- The number of scholarships awarded each month;
- The number of scholarship recipients who enroll in AA programs; and
- The number of scholarship recipients who complete their AA degrees each quarter.

Each intermediate outcome is a prerequisite for the initiative's success. Failure to achieve one or more of them points to the need to examine the activities that were expected to produce these results. Program managers can then make the appropriate changes—developing new

outreach strategies to recruit teachers or streamlining enrollment procedures, for example—to make the initiative work better.

Summative Evaluation

Summative evaluation provides decision-makers with information about the impact or effectiveness of initiatives. This information can help policy makers make decisions about continued funding of an initiative or allocation of funding between competing initiatives.

Evaluation

“If formative evaluation is the chef tasting the soup and adjusting the seasoning, then summative evaluation is the customer tasting the soup and passing judgment on its quality.” (Stake, cited in Scriven, 1991).

Summative evaluations typically focus on the long-term goal of the initiative or program. A summative evaluation of the initiative we cited earlier would examine the degree to which it actually increased the number of credentialed center teachers.

A summative evaluation may include additional questions in order to differentiate aspects of the impact of the initiative. For an initiative that provides scholarships to increase the number of providers with CDAs or AAs, some of these questions might be:

- Was there a differential impact among center teachers (by age, marital status, number of years as a child care provider, membership in a professional child care organization)?
- Was there a differential impact across the state (in rural areas compared to urban areas)?

Definition

Evaluators often use theory of change models, because they are useful tools for planning and implementing evaluations. A theory of change model is a map that clearly identifies the features of an initiative that are intended to produce its long-term outcome or goal. Theory of change models can be used for designing formative evaluations that assess the implementation of an initiative as well as for summative evaluations that assess its effects. A theory of change model explicitly specifies how the activities of the initiative lead to the intended outcomes. It also reflects the assumptions on which an initiative or program is based. (Wholey, 1987, in Rogers, Petrosino, Huebner, & Hacsí, 2000).

Three steps are involved in creating a theory of change model. They can be applied to any effort to improve child care. The first is to specify the long-term outcome, or goal, of the initiative. The second is to specify the intermediate outcomes that must be completed to achieve the long-term outcomes. The final step is to specify the activities that will produce the intermediate and long-term outcomes of the initiative.

Specifying intermediate outcomes makes it possible to develop a logical sequence of events between the long-term outcome and the activities needed to reach it. Linking the long-term outcome to intermediate outcomes and activities is the key element of a theory of change model. It provides a clear plan that can be used in an evaluation to document each link in the chain.

Features of a Theory of Change Model

Specifying outcomes

Specifying the desired outcomes of an initiative is a critical early step in building a theory of change model. One approach that can be useful in this process is to ask what will make you happy at the conclusion of your initiative or what will success look like.

Based on your answers to these questions, you will be able to refine the goal of your initiative by specifying the type of provider you want to target, their location, and their initial level of knowledge. The intermediate outcome must be related to the long-term outcome.

Focus on measurable outcomes

The outcomes included in a theory of change model must be concrete

and measurable to the extent that an outside observer can determine whether they have been achieved. Even an apparently simple outcome such as “increased number” must be precise and quantifiable. For example in Connecticut, the School Readiness Initiative aimed to increase the number of accredited centers. This long-term outcome could be measured in a number of ways: 10 more accredited centers than there were before the initiative was implemented; a net increase of 10 accredited centers, because some of the older centers may have lost their accreditation; or 10 new accredited centers in a specific locality. The more specific the outcome, the easier it is to measure the result.

Link specific, detailed activities to desired outcomes

In addition to specific measurable outcomes, a theory of change model must include clearly defined activities. Being specific about the activities helps you determine if you have accomplished them.

Reflect the underlying logic of the initiative’s strategy for achieving its goals

A theory of change model must also be plausible given the context in which the initiative will be implemented. The model should reflect the initiative developers’ knowledge and experience so there is reason to believe that the specified activities will lead to the intended outcomes.

Uses of A Theory of Change Model

A theory of change model can be useful to staff and evaluators throughout the life of an initiative from the planning stage to the final step of assessing long-term outcomes.

Clarify program goals and process

Developing a theory of change model can be useful in helping program staff clearly articulate what they are trying to accomplish and how they will do it, especially in the early stages of program planning. Working out a detailed theory of change model requires a process of repeatedly asking questions about each outcome, so that the elements of an initiative and their relationship to the long-term outcome come to be understood more and more thoroughly. This process can help identify flaws in thinking about how the program should be effective. It also allows staff to make adjustments to the design of the initiative before it is up and running (Weiss, 1995, in Rogers, Petrosino, Heubner, & Hacsí, 2000).

Lead to improvements in the initiative

Using a theory of change model to guide a formative evaluation enables program staff to make improvements or adjustments to the initiative as it is being implemented. In the early stages of an initiative, a theory of change model can help identify where the initiative may be deviating from the initial plan and can point to a mid-course correction.

Assess links between program activities and outcomes

A theory of change model can guide an evaluation by providing a road map that clearly illustrates the evidence needed to assess effectiveness.

It can be used to build a strong case that the activities contributed to the outcomes, but it cannot prove that activities caused the outcomes in the way a true experimental design can.

Suppose that an evaluation of a professional development initiative finds an increase in the number of providers with CDAs. Evidence that each of the intermediate outcomes had also been achieved, would support the conclusion that the initiative was, at least in part, responsible for the result. If the initiative did not result in the anticipated number of providers with CDAs, a theory of change model is also useful. It can be used to track which intermediate outcomes were not achieved, and which activities did not take place.

In this chapter, we show how a theory of change model can be used to develop an evaluation. We create a theory of change model based on Alabama's Statewide Leadership Scholarship initiative, for which we inferred plausible outcomes and activities.³ In the next step, we identify questions related to the theory of change model that could be used to evaluate the Alabama example. Finally, we suggest measures, and instruments from Part II, that could be used to determine if the outcomes in the example have been achieved.

Theory of change models can vary widely in their complexity and level of detail. To develop a model for the Alabama initiative (Figure 1), we identified the long-term outcome; then we identified the intermediate outcomes. In the next step, we identified the evidence needed to document that those outcomes had been achieved. After that, we identified the activities that would lead to the outcomes. In the final step, we developed the evaluation questions to ask whether the intermediate and long-term outcomes had been reached.

The outcomes are indicated by gray boxes; the activities by white boxes. Outcomes are always nouns. They must be measurable and quantifiable. Activities are always verbs. (In Appendix C, we have included a blank theory of change template that you can use to create your own model for your initiative.)

The Alabama example we provide here is intended as an introduction to the process of creating a theory of change model rather than as a detailed guide for developing one.

Step 1: Identify Long-Term Outcome: 100 providers with credentials

As Chapter 3 indicates, one way of determining your initiative's long-term outcome is to ask questions such as, "What will success look like? What will make you happy when the initiative is completed?" that will clarify the aspects of your initiative that are most important to you. In our example, the long-term outcome was to provide scholarships to increase the number of caregivers with credentials by 100. To reach this long-term outcome, the initiative must produce a series of intermediate outcomes.

Step 2: Identify Intermediate Outcomes⁴

Next we identified the intermediate outcomes. The order in which they are listed below is reflected in Figure 1 if it is read left to right.

Outcome: 110 Enrollees

To credential 100 new caregivers in the state, caregivers must enroll in courses. Because some caregivers may not complete all of the required courses, the intermediate outcome is enrollment of 110 caregivers.

Outcome: 125 Scholarships

To ensure that 110 caregivers enroll in courses, 125 scholarships must be awarded because some of the caregivers will not accept the scholarship.

Outcome: 250 Applicants from Pool

To ensure that there is a large enough pool of providers for the scholarship, applications must be received from 250 caregivers.

Outcome: Partnerships with University System

To offer courses to scholarship recipients, partnerships with the University System must be developed and an agreement must be negotiated with it.

Outcome: Selection Criteria

To select caregivers for participation in the scholarship program, criteria must be developed.

Step 3: Identify Activities

The next step is to specify the activities needed to achieve each outcome. To identify the activities, start with the first intermediate outcome and work forward to the long-term outcome. In this exercise you will link activities to specific outcomes.

Activity: Identify Pool

Program staff must identify the population of providers that the initiative will serve.

³ The theory of change model and this chapter are based on the description of Alabama's Statewide Leadership Scholarship initiative in *A Study of Selected State Initiatives: Volume II*. To illustrate the change model, we inferred outcomes and additional activities from that description.

⁴ The numbers we have used to illustrate these outcomes are arbitrary estimates that we have developed.

Activity: Form Committee

Selection criteria must be identified for caregivers who will receive scholarships. In Alabama, a committee of stakeholders was convened to complete this task.

Activity: Reach Out to University System to Create Partnership

Several activities are involved in reaching out to the universities: meetings with system members; presenting the plan for the program; and creating literature that describes the program.

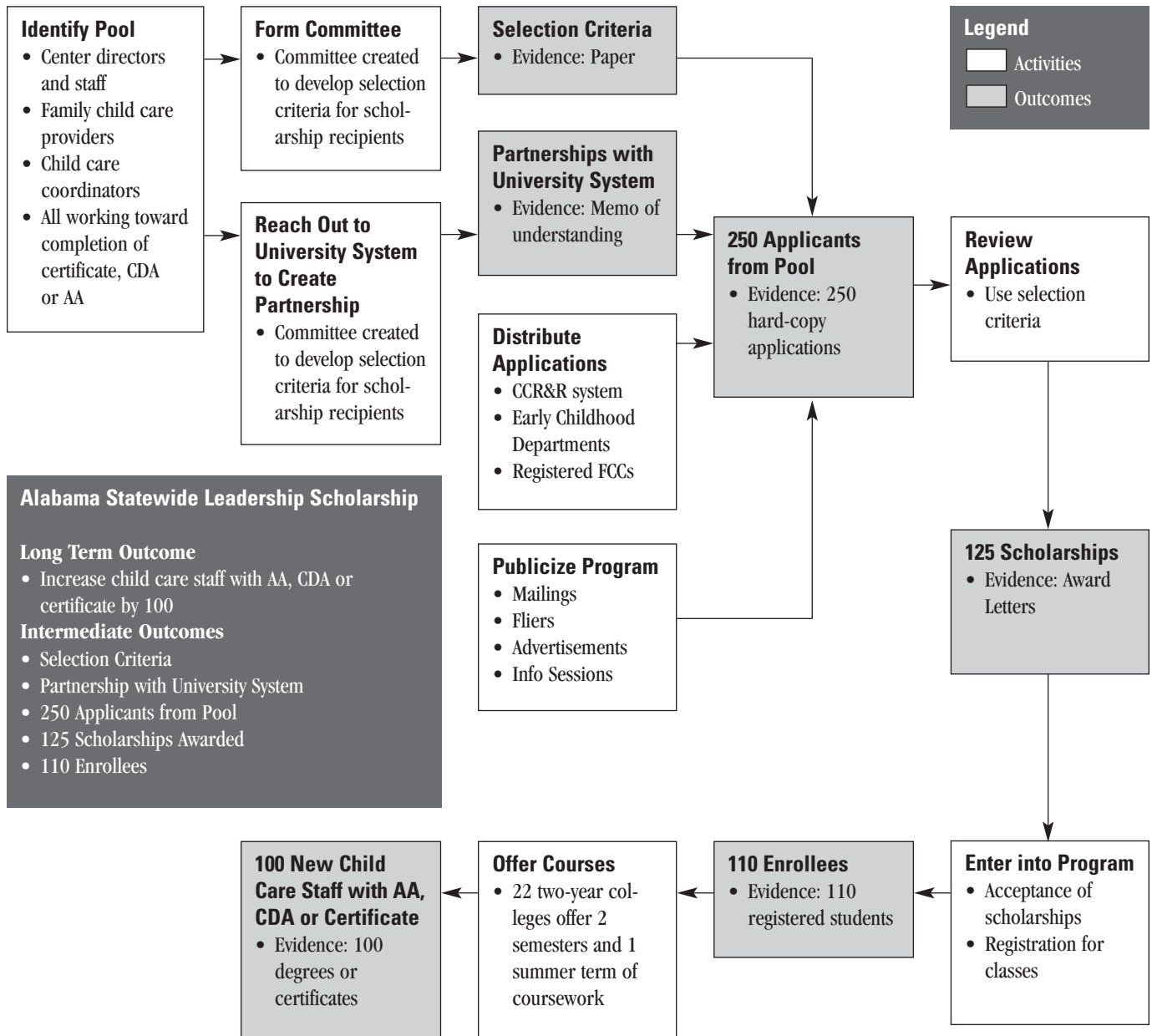
Activity: Publicize Program

The initiative is publicized through mailings, fliers and advertisements in local papers as well as through information sessions hosted by the program at child care centers, CCR&Rs and at institutions of higher education.

Activity: Review Applications

Applications are reviewed and providers are selected using the selection criteria.

Figure 1: Alabama Statewide Leadership Scholarship



Activity: Enter into Program

Caregivers must accept the offer of scholarships and register for classes.

Activity: Offer Courses

Courses are offered at 22 colleges across the state. Scholarship recipients must complete two semesters and one summer term of coursework to receive the credential.

Linking Outcomes to Measurement Instruments

Once you have clearly specified your initiative's long-term and intermediate outcomes, you will want to identify appropriate measurement instruments to collect the data you need to show that the outcomes have been met. In the Alabama initiative, for example, if the type of certification that participants receive is important to you, you will need to collect data on the numbers of participants that received each type of certification (for example, AA, CDA, certificate).

Below we list questions that might be used to assess whether specific long-term outcomes and intermediate outcomes related to the Alabama initiative have been met. To identify your questions, you will want to ask "Do we care?" and "Does it matter?" The possible outcomes vary from increasing the numbers of participants who receive different types of certification to improving the child care practices of the scholarship recipients. We have matched the questions related to different outcomes with appropriate instruments in Part II of the *Toolkit*.

Long-Term Outcome 100 New Child Care Staff with AA, CDA, or certificate

- Does it matter which type of certificate or degree participants complete—an AA, a CDA or a certificate? (Participant Statistics)
- Does it matter if there is a net of 100 caregivers with credentials or simply 100 new caregivers with credentials? (Participant Statistics, Secondary Data)
- Do we care if the AA, CDA or certificates are completed within the required/anticipated timeframe? (Participant Statistics)
- Does it matter who the scholarship recipients are in terms of: the types of child care they provide—center, family child care, or license-exempt care; ethnicity; job title; length of time in the field; location; age; or income? (Participant Characteristics)

- Do we care if scholarship recipients who obtain AAs, CDAs or certificates stay in the same job after 6 months? Stay in the child care field after 6 months? Do we care where they go if they do not remain in the field? Does it matter why? (Retention and Turnover)
- Does it matter if scholarship recipients earn more money as a result of a new degree? (Benefits to the Workforce)
- Do we care about who is offering the training (in terms of which professors/instructors)? Does the format of the training matter? (Participant Characteristics modified as Trainer Characteristics)
- Do we care what scholarship recipients thought of the initiative? (Satisfaction)
- Do we care what specific knowledge and skills the scholarship recipients gained? (Pre/Post, Grades Review)
- Do we care whether their practice changed? And how it changed? (Observation, Self-Report, Parent Report, Director Report)
- Do we care if parents observe changes in the knowledge and skills of caregivers? (Parent Report)
- Do we care if directors/supervisors observe changes in the knowledge and skills of the caregiver? In caregiver's practice? (Director Report)

Intermediate Outcome: 110 Enrollees

- Does it matter who the 110 enrollees are? (Participant Characteristics)
- Do we care which institutions the participants are enrolled in? (Participant Statistics)
- Do we care why they enrolled? (Satisfaction)

Intermediate Outcome: 250 Applicants from Pool

- Does it matter who the applicants are? (Participant Characteristics)
- Do we care what they thought of the process? (Satisfaction)

While theory of change models are most often used to guide formative evaluations directed at improving an ongoing program, they can also be useful in summative evaluations designed to determine a program's effectiveness. For example, a detailed theory of change model can be used to explore possible reasons why a program did or did not have the intended impact. Summative evaluations are most useful if conducted after a program has reached a "steady state" or after its completion (Scriven, 1991).

This chapter provides a broad overview of three key issues that must be addressed in designing a summative evaluation of a program's impact: forming comparison and/or control groups; units of analysis; and sample sizes. In general, these issues are important for evaluations that use quasi-experimental or random assignment designs. Evaluations that are descriptive or that use simple pre/post designs provide useful information that can often meet the needs of policy makers as effectively.

Forming Treatment and Comparison Group

By definition, a summative evaluation of a program's impact compares outcomes between a "treated" group and a "non-treated" group: it compares outcomes of program participants to those of non-participants. The estimated impact of a program is the difference in outcomes between the two groups. This means that you will have evidence that an initiative was successful if you find the desired difference between the treated group and the untreated group. The manner in which these groups are formed is a critical factor that determines the validity of the findings—the degree to which you can be confident that the differences measured between the treated and untreated groups are a direct result of the initiative intervention. Here we distinguish among three general designs for forming these two groups:

- Simple pre/post comparisons;
- Quasi-experimental designs (using external comparison groups); and
- Experimental designs (random assignment).

Simple Pre/Post Comparisons

In this design, program participants serve as their own comparison group. Outcomes are measured before and after participation in the initiative or program. The estimated impact of the program is taken as the difference between post-test and pre-test outcome measures for those that participated in the initiative.

A simple pre/post comparison may be the weakest design for a summative evaluation. The major problem is that it is extremely difficult, if not impossible, to rule out other factors that might have led to, or contributed to, the observed change in outcomes between the pre-test and the post-test. For example, suppose that a training initiative provides workshops for caregivers to enhance their ability to work with children with disabilities. Pre- and post-tests results indicate that caregivers' knowledge of best practices for working with children with disabilities has increased. While this simple pre/post comparison might lead evaluators to conclude that the change in knowledge can be attributed to the caregivers' participation in the program, it is not possible to exclude other explanations for the observed change such as a local television series on children with disabilities. To help rule out other factors that might have led to the measured changes between pre-test and post-test, evaluators must examine the context in which the intervention took place.

Quasi-Experimental Design (external comparison groups)

In this design, a comparison group is formed from individuals who did not participate in the initiative or program (an external comparison group). Again, the estimated impact of the program is the measured difference in outcomes between the treatment and comparison (non-treated) groups. The similarity between the treated and non-treated groups is what allows you to attribute the change in measurable outcomes to your initiative. However, if there are any differences between the two groups that might influence outcomes, the validity of the estimates of the impact of the initiative are questionable.

While there are various statistical methods that evaluators use to adjust for differences between the treatment and comparison group, it is always possible that some unmeasured differences between the two groups may explain all, or part, of the observed difference in measured outcomes. Suppose that the evaluators of the training initiative example we used above formed a comparison group by taking a sample of caregivers from the same community who did not participate in the workshops. Even after adjusting for differences between the two groups on such measurable factors as age, ethnicity, length of time providing child care, type of provider (center-based or home-based), education, previous child care/early childhood education training, and professional credentials, an inherent difference remains between the two groups—the treatment group elected to participate in the training while the comparison group did not. This self-selection bias can never be completely eliminated analytically.

Experimental Designs (Random Assignment)

The “gold standard” for evaluation is an experimental design in which a group of individuals eligible and willing to participate in a program are randomly assigned to a treatment group or to a control group that does not participate in the treatment. Random assignment is the only design that, if properly implemented, allows an evaluator to attribute differences in measured outcomes unambiguously to the treatment group’s participation. Because participants are assigned to the treatment and control groups randomly, there are not likely to be any differences between the groups that could explain the outcomes. Thus, unlike the previous designs, any differences between the groups can be attributed to the intervention. Evaluations with random assignment designs are among the most expensive and most difficult to conduct. As a result, they are not often used in evaluations of local programs or initiatives.

To use random assignment in the example of the training initiative, the evaluator would have to identify and recruit a pool of caregivers who wanted to participate in the program and were willing to accept random assignment to treatment and control groups. While it might be possible to obtain consent from the treatment group to participate in the evaluation, it is substantially more difficult to obtain cooperation of those assigned to the control group. In addition, those assigned to the control group have already demonstrated their interest in the type of training provided by the initiative, and they might obtain similar training elsewhere.

Unit of Analysis

Determining the appropriate unit of analysis and the required sample size are important issues for evaluation, irrespective of the design. There are three likely units of analysis for child care quality improvement initiatives: the child care program, the individual caregiver, or the child. Evaluations that use children as the unit of analysis are often expensive and complex. As a result, we do not focus on them here.

It is important that the unit of analysis for an evaluation correspond to the target of the initiative. Using an inappropriate unit of analysis will not provide a valid estimate of the impact of the initiative. Initiatives that are designed to influence the choices that child care programs make should be assessed at the program level rather than the caregiver level. An example of this kind of initiative is North Carolina’s Five-Star Rated License System, which rates centers on program standards, staff education levels, and compliance with child care regulations. The evaluation should look at changes in the number of programs with higher ratings as evidence of the initiative’s impact.

Caregiver outcomes should be the focus of evaluations of initiatives that are designed to affect them. For example, the Alabama Statewide

Leadership Scholarship initiative focuses on providers, with a goal of increasing the number of providers with CDAs, AAs or BAs. In this case, the evaluation should look at the number of providers with these credentials.

As a practical matter, it is often much more feasible to evaluate an initiative that targets individual caregivers by nesting the sample of caregivers within a sample of centers, that is, including multiple caregivers from each center. This reduces the time and expense necessary to select and recruit the sample of providers. However, such a sampling design reduces the effective size of the caregiver sample because of the similarity of providers within each center. This design effect means that the evaluation will need a larger sample of caregivers than if the caregiver sample was not nested within a sample of centers. The magnitude of the design effect is a function of the number of caregivers selected from each center and the degree of similarity among caregivers within each center.

Required Sample Sizes

A key design issue for any evaluation of a program’s impact is the determination of the sample size for treatment and comparison or control groups. Three factors determine the minimum sample sizes needed for both groups: the desired statistical power; the minimum detectable effect size; and the significance level criterion of the statistical test.

- *Statistical Power.* In general, the minimum acceptable statistical power is .80. If a difference really exists between the treatment and control populations, the experiment should have a minimum of an 80 percent probability of detecting this difference.⁵
- *Effect Size.* Effect size refers to the magnitude of detected differences in outcomes. If you want to detect small differences in your data, you will need a large sample size.
- *Significance Criterion.* The significance level is the rate of rejecting a true null hypothesis (that is, saying that there is a difference between groups when there is not). Substantially larger sample sizes are needed to detect differences between two groups when the desired significance level criterion is smaller. For example, a much larger sample is required to provide a 1% chance of finding a difference between treatment and control groups when, in fact, no difference exists, than the sample size required to find a 5% or 10% difference.

Given the complexity of these issues, you will want to consult someone with evaluation and statistical skills, especially if your evaluation uses a quasi-experimental or experimental design. In most cases—descriptive evaluations or those that use a simple pre/post design—you will not have to address these issues.

⁵ See J. Cohen’s *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed., 1988) for additional information on this topic.

Developing a complete, detailed evaluation plan is an essential first step in conducting a useful evaluation. A comprehensive plan will provide the framework you need to guide you during the evaluation. As we indicated in Chapter III, creating a theory of change model is an important aspect of developing an evaluation plan because it will help you specify outcomes and think about the questions you want to ask. Your evaluation plan should complement your theory of change model.

In this chapter we present an overview of some of the steps involved in designing an evaluation plan. They range from assembling an evaluation team and developing a budget and timeline to designing data collection and data analysis plans and a review of the evaluation plan. These steps apply whether the evaluation is conducted internally (by your own staff) or by an external or third-party evaluator. We recommend reviewing the *ACF Guide* for a more comprehensive discussion of this process.

In Appendix B, we have provided examples of evaluation plans developed by the three states that field tested the *Toolkit*—Ohio, Kansas, and Washington. (We have included a blank evaluation plan template in Appendix C to help you organize these steps for your own evaluation.)

What State Child Care Administrators Say

- Think about evaluation in advance, even as early as designing the initiative.
- Consider involvement of the legislature in some capacity.
- Think about different degrees of complexity and related costs.
- Decide whether you will use your own staff or turn to an outside evaluator.
- Think about the kind of information you want—information about how the initiative works or its effects—and then consider your design.
- Consider how to design and use specific instruments and their constraints.

The Evaluation Framework

An essential precursor to mapping out an evaluation plan is an assessment of your agency's capacity and experience in conducting evaluation as well as the availability of program staff to conduct it. Before you begin, it is important to determine the resources you have. For exam-

ple, is there an information management system that could be used to tabulate and analyze data? Is there an advisory committee with members who have evaluation experience? It is also important to consider the agency's previous evaluation experience. Have other programs been evaluated? Was the experience negative or positive? What problems were encountered? How can they be avoided? Would the design of previous evaluations be applicable to this evaluation?

To better understand the initiative you will be evaluating you will want to know whether the program is incorporated into the agency's other activities or if it is a stand-alone initiative. If it is incorporated into other activities, it is important to consider how you will evaluate it as its own entity.

It is also necessary to consider the agency staff's experience with evaluation. To ensure that the data collection process is properly implemented, staff must be involved in the process of developing the evaluation plan and trained to conduct the evaluation. By addressing any concerns the staff may have about evaluation beforehand, you will be able to avoid such issues before the evaluation is implemented.

Step One: Assemble an Evaluation Team

Assembling a team to plan your evaluation is the first step. Chapter Three of the *ACF Guide* describes different types of evaluation teams, those led by an outside evaluator or consultant and those led by an in-house evaluator. No matter what type of evaluator you use, the team should include initiative and agency staff and—when appropriate—staff from state or regional child care administration offices. The manager or director of the initiative, the person most familiar with the program, is an essential team member. Whether you will be able to rely on in-house staff or will need to hire an outside evaluator or consultant will depend on the resources available in your agency. The evaluation team you assemble will collaborate with the individual in charge of the evaluation to develop the overall plan and your theory of change model.

Ohio, one of the field test sites, proposed to include staff from the Department of Jobs and Family Services and the Department of Education as well as a family child care provider, while Kansas, another field site, identified an evaluation team of trainers from the initiative, the Infant/Toddler Project Director, the grant manager from the Department of Social and Rehabilitation Services, an evaluation professional, and a parent as well as a legislator.

Step Two: Organize an Advisory Committee

In addition to the evaluation team, you may want to also have an advisory committee. Members of an advisory committee may include researchers, experts, program planners, policy makers and members of the population served by the initiative. This committee can assist the evaluation team by providing feedback on the evaluation plan and design, the theory of change model, the data collection plan and interpretation of the findings. An advisory committee can also help you make decisions about the future of the initiative if the evaluation points to negative results.

Typically, an advisory committee consists of the initiative's stakeholders. In Washington, the advisory committee consisted of the STARS stakeholders, an advisory group to the initiative that represents all aspects of the child care community. Kansas proposed to include a legislator, a parent, the initiative's director, a participant and the state grant manager, among others on its committee.

Step Three: Obtain Community Participation

It is important for the members of the community that are affected by your initiative to participate in the evaluation planning process. There are several ways of ensuring community participation. If you are interested in obtaining a response from the community, you can hold a forum or town meeting where you present your evaluation objectives, or you can hold focus groups with community groups to assess their concerns.

Step Four: Create Theory of Change Model

As described in Chapter 3, creating a theory of change model begins with identifying the initiative's long-term and intermediate outcomes. Program staff are the most knowledgeable about what they intend the initiative to accomplish, but the evaluator and the rest of the evaluation team can contribute to identifying outcomes and the activities that will lead to those outcomes. The team and the advisory committee should sit down with the evaluator (whether she/he is an in-house or a third-party evaluator) to discuss plans for the program and to decide what to evaluate.

As Chapter Six of the ACF *Guide* recommends, outcomes should be stated in measurable terms. The theory of change model should explicitly identify what the desired outcomes are and what measures will be used to document that the outcomes were achieved.

Step Five: Develop a Budget and Timeline

It may be helpful to take this step early in the planning process, since your budget and the time you will need to conduct the evaluation will play large roles in determining the kind of evaluation you design. This step helps you set priorities for the kinds of questions you can answer with the budget and the resources you have.

Deciding the amount of money you plan to spend on an evaluation will help you determine the scale of the study (see Chapter Two of the ACF *Guide*). Your timeline should take into account when the data collec-

tion staff and information will be available, how much time will be needed for collecting and analyzing the data, and when the evaluation results are needed. It is important to keep in mind that there may be pressure to collect some outcome data from initial implementation, but it is not appropriate to do so before an initiative has been in existence for more than a year.

Kansas took each of these tasks into account in its four-month timeline, while Ohio's timeline corresponds to the initiative's contract. Washington created a detailed timeline that outlines specific points during the implementation of the initiative when the data will be collected and analyzed.

What State Child Care Administrators Say

When you design your evaluation, you should:

- Think about how much data to collect and whether it will answer your questions.
- Understand the limits of the data: what you can do with it (what it says) and what you can't (what it doesn't say).
- Be careful about using observation instruments or rating scales correctly with the appropriate training to ensure inter-rater reliability.
- Consider whether you can use "natural" control groups in a quasi-experimental design.

Step Six: Develop a Data Collection Plan

Your evaluation plan should clearly state how you will collect the data needed to establish whether the long-term and intermediate outcomes included in your theory of change model have been achieved. In your plan, the data collection instruments you will use, and the procedures and methods you plan to employ to collect the relevant data, should be clearly linked to each outcome. You should use your theory of change model to map how data collection can be integrated into the operation of the initiative. Two aspects of a data collection plan are selecting data sources and monitoring data collection.

Selecting Data Sources

The selection of data sources depends on the kind of information you need to assess the outcomes of your initiative. In choosing data sources you need to identify how the information will be collected as well as who will collect it. Part II of *The Toolkit* provides instruments that will help you collect data. These instruments can be adapted to fit your specific evaluation.

The sources of data and the methods used to collect them vary. Sources range from program memos to participants; methods range from review of program related-materials to surveys, interviews and observations. The following list of sources and methods is taken from *Building a Stronger Child Care Workforce*.

- Document Review can be used to determine how the initiative operates and what works and doesn't work. Possible sources are personnel files, financial statements, meeting minutes, and legal documents. For example, the evaluation of California's Training TANF

Recipients as Child Care Providers initiative included a review of planning memos. (Please see Appendix A for a more complete description of this evaluation.)

- Program applications, utilization forms and other paperwork are often used for evaluations to help understand how the program operates as well as its successes and failures. Examples include intake forms and sign-in logs for activities. The North Carolina T.E.A.C.H. Early Childhood® Project, an initiative that links professional development and compensation, used enrollment forms to track the number of participants. This is probably the quickest and easiest way to collect data. (Please see Appendix A for a more complete description of this evaluation.)
- Questionnaires/Surveys are often used for evaluating satisfaction. They also help determine the effect of the initiative on the target population. California's Child Care Initiative Project, an effort to recruit and train family child care providers, administered surveys to participants and program directors to obtain information about the number of participants who remained in the child care field. (Please see Appendix A for a more complete description of this evaluation.)
- Semi-structured interviews can be used to obtain answers to specific questions as well as to provide opportunities for participants to respond freely. Interviews often help the evaluators assess the operation of the initiative as well as its impact. The evaluators of Michigan's R.E.A.D.Y., an initiative that distributed child development and literacy kits, conducted phone interviews with individuals who had received the kits. The interviews were used to determine the value and usefulness of the kit, to assess the appeal of the material, and to obtain suggestions from recipients about how to improve the kit content, materials and distribution. (Please see Appendix A for a more complete description of this evaluation.)
- Focus group discussions are also used to help determine the strengths and weaknesses of an initiative as well as its effects. The Training TANF Recipients as Child Care Providers evaluation included focus groups at two of the three program sites. The discussions were used to gather information on participants' perceptions of the initiative's strengths and weaknesses, an implementation question; and to assess its effect on participants' self-sufficiency, an anticipated outcome. (Please see Appendix A for a more complete description of this evaluation.)
- Assessments/Observations are often used to assess provider or program quality as well as effects on children. Observations are generally conducted with standardized instruments that have been validated through rigorous research. Observations should be conducted at different intervals—initially, to establish a baseline and then later, to determine if changes have occurred. The Connecticut School Readiness Evaluation used the Early Childhood Environment Rating Scale (ECERS) to assess the effect of the initiative on child care quality. (Please see Appendix A for a more complete description of this evaluation.) Observation is one of the most expensive and time-consuming data collection methods, because it requires trained evaluators and multiple visits. Appendix A in Part II describes several instruments for observing quality in child care settings, with information about their cost and how to obtain them.

Managing and Monitoring Data Collection

To ensure that you will collect all the data you need, you should include a protocol for monitoring the data collection process in your evaluation plan. Chapter Seven of the ACF Guide describes several steps you can take.

- Establish a routine and timeframe for data to be submitted. If you create a data collection manual, you can include the routine and timeframe in it. The routine may include the submission and review of completed data collection instruments to a member of the evaluation team so that any issues regarding the data collection can be addressed quickly.
- Conduct random observations of the data collection process. This will ensure that the data are being collected as intended. Observations could include sitting in on an interview session to ensure the procedures are being properly followed.
- Conduct random checks of respondents. This is another way to ensure that data collection protocols are being followed. An evaluation team member should be assigned to check with a sample of respondents on a regular basis. The research team member should ask respondents if they were asked to complete the confidentiality and consent paperwork, whether any concerns or questions they may have had about the interview were answered, and whether they felt they were treated appropriately during the interview.
- Protect the confidentiality of participants. Data should be kept in a secure location. Using number codes instead of names on completed surveys or instruments is one way to ensure that the confidentiality of the participants is maintained.

Step Seven: Design a Data Analysis Plan

The goal of this step is to determine the types of analyses you will perform on the data you collect. Quantitative methods are useful for summarizing large amounts of information. Qualitative methods are useful for analyzing open-ended questions, which requires extraction of general themes from the information. Usually, evaluations consist of some combination of both types of analysis. Chapter Eight of the ACF *Guide* provides information about different procedures for analyzing evaluation data to help you understand the process more fully. You are likely to want help from someone with evaluation expertise to determine the most appropriate data analysis plan.

The data collection strategies you use will influence the type of data analysis. For example, applications, utilization forms and surveys lend themselves to quantitative analysis. Interviews and focus groups generate data that are appropriate for qualitative analysis. Assessment and observations can produce data for both types of analysis.

Step Eight: Review The Plan

Once an evaluation plan is developed it is a good idea to have it reviewed widely. If an experienced evaluator was not involved in the plan's development, then he or she should be consulted. The advisory committee as well as others who have not been involved in its creation,

such as agency and state administrators, should also review the plan to make sure it is consistent with their resources and evaluation objectives. Feedback should also be elicited from the initiative staff about the demands the evaluation may place on them and participants. You can also ask selected participants and community members to assess the plan's appropriateness and cultural sensitivity.

Practical Issues

Data Collection Procedures

In selecting data sources and data collection methods for an evaluation, it is important to keep several issues in mind. If you will be collecting data from participants who have a low level of literacy, you must ensure that their needs are considered and that you use appropriate language in consent forms, surveys, and other instruments. To address the issue of low literacy levels, for example, you can read questionnaires to participants rather than ask participants to read the survey themselves. In your pilot test, you can make sure that the language is simple and clear. It is also important to think about the most efficient way to collect data in the context of your service delivery strategy. For example, it may make sense to distribute surveys directly to participants in a classroom-based training program, but mailing them or conducting phone interviews may make more sense if the training uses a distance-learning approach.

Training Data Collection Staff

An evaluation plan needs to include a component on training data collection staff as well as pilot testing of data collection instruments and procedures. Before staff begin data collection, it is important to ensure they understand the instruments. You should schedule some sessions with staff to provide training about how to use them. Chapter Seven of the ACF *Guide* identifies the following issues that you should cover in these sessions: item by item review of the instruments; review of the instructions on administering the instruments; discussion of potential problems in administering the instrument; how to protect participant confidentiality; and conducting frequent reviews and checks of the data to ensure consistency.

The training should include an opportunity for staff to use the instruments in a practice session before using them to survey or interview participants. The ACF *Guide* recommends conducting a pilot test of the instruments and the data collection process before you begin the full-fledged evaluation. (Please see Chapter Seven for more information.)

It is important to provide staff with clear written instructions, for using instruments. You can also develop a manual. If you are adapting a survey instrument from the *Toolkit* and using it in a phone interview, for example, it is important to provide a written protocol with instructions for staff who will conduct the interviews. This will ensure that the data you collect are consistent.

Cultural Sensitivity

In preparing staff to collect data, it is important to consider whether the staff should be familiar with the culture or language of the participants whom they will be interviewing or surveying. In addition, you should consider reviewing instruments to identify whether some questions may be inappropriate in certain cultural contexts. It is also important to provide translations of surveys, privacy protocols and any other materials related to data collection from respondents whose primary language is not English.

Confidentiality: Ensuring Participants Privacy

The evaluation plan must also describe the ways in which all participants' rights to privacy will be respected. Whenever participants are asked voluntarily to provide private information or information that may be used in a way that will affect them, they must be assured that the information will be kept confidential. Informing participants that they will not be identified by name in any reports provides assurance that specific information about them will not be made public. It also increases the likelihood that participants will provide the information that you are seeking. They must also know they have the right to refuse to answer any questions or to participate in the evaluation. A common strategy for maintaining confidentiality is to assign a case number, which can be used for all of the participant's records. Data collection instruments should provide written assurance that participants' names will not be printed on any forms and that no one other than the research staff will be able to match names and identification numbers.

The wording below could be used as a heading to explain how participant confidentiality will be protected or how the confidentiality process works. You may want to adapt it for your own needs.

Thank you for taking the time to respond to the following questions. Your answers will be completely private and only the research staff will see your completed survey. Your name will not be printed on any of the forms. We have assigned an identification number to you and no one with the exception of the research staff will be able to match your name with the identification number.

Policy makers at all levels of government are concerned about using their resources effectively. They want to invest in sound programs that work and that produce results. For policy makers who are responsible for child care, decisions about supporting initiatives are especially important because they affect the lives of young children. Effective initiatives can make a difference in child care quality and can benefit children's social, emotional, cognitive, language and physical development as well as their readiness for school.

Evaluation provides the tools that can help policy makers determine how to allocate their child care funding effectively. It can answer questions about whether an initiative worked the way it was designed and whether it achieved the desired results. Evaluations of program implementation can point to weaknesses—recruitment strategies that did not reach the target population, timeframes that were unrealistic, approaches that did not meet providers' needs, or staffing levels that did not match service delivery expectations. They can also identify an initiative's strengths, the features that contributed to its success.

Evaluation can also answer questions about whether an initiative has resulted in the kinds of changes that it was intended to achieve. Although random assignment evaluation designs are considered the best approach for determining if an initiative has produced the desired effects, non-experimental designs that rely on program participants as their own comparison group are legitimate and effective ways to assess outcomes. Non-experimental and quasi-experimental designs can show the initiative has produced changes in the quality of the child care environment, providers' knowledge and practice, turnover, and the ultimate goal of improving quality—children's outcomes.

Evaluation can be used in other ways as well. It can provide information about the costs of initiatives, enabling policy makers to compare actual with estimated expenditures, which can be used to make future funding decisions. Understanding the cost of specific approaches can help policy makers determine whether similar initiatives are feasible—whether results will justify the investment. Comparing the costs of the initiative with its effects can provide information about whether the investment represents an efficient use of resources.

Policy makers can use evaluation in a variety of ways to make decisions about efforts to improve child care quality. They can assess a new initiative to determine if the effort should be modified, expanded, or abandoned. They can look at multiple initiatives that use the same strategy to determine the approach that is most effective for specific child care programs, providers or populations of children. Or they can use evaluation to answer questions about the effectiveness of different strategies for improving child care for specific types of child care programs or providers. Policymakers can also use evaluation to examine their quality improvement efforts in combination to better understand the overall impact on quality.

Whether evaluation is used to assess implementation, effects or costs, it is a worthwhile investment that should be part of every policy maker's portfolio. The returns are valuable, because they affect crucial decisions about funding child care initiatives. Understanding evaluation and how to conduct it will produce long-term rewards for the children these initiatives are intended to serve.

References

- Administration on Children, Youth and Families. (1997). *The Program manager's guide to evaluation* [on-line]. Available: <http://www.acf.hhs.gov/programs/core/pubs_reports/prog_mgr.html>.
- Arnett, J. (1989). Caregivers in day care centers: Does training matter? *Journal of Applied Developmental Psychology, 10*, 541-552.
- Burchinal, M.R., Roberts, J.E., Nabors, L.A., & Bryant, D. M. (1996). Quality of center care and infant cognitive and language development. *Child Development, 67*, (2) 606-620.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, (2nd Ed.). Hillsdale, NJ: Earlbaum.
- Helburn, S.W., Culkin, M., Howes, C., Bryant, D., Clifford, R., Cryer, D., Peisner-Feinberg, E., & Kagan, S.L. (1995). *Cost, quality and child outcomes study: Executive summary*. Denver, CO: University of Colorado.
- NICHD Early Childhood Research Network. (1999, April). *Effect sizes for the NICHD study of early child care*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Albuquerque, NM.
- NICHD Early Childhood Research Network. (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Sciences, 4* (3), 116-135.
- Park-Jadotte, J., Golin, S. C., & Gault, B. (2002). *Building a stronger child care workforce: A review of studies of the effectiveness of public compensation initiatives*. Washington, DC: Institute for Women's Policy Research.
- Peisner-Feinberg, E. S., & Burchinal, M. R. (1997). Relations between preschool children's child-care experiences and concurrent development: The cost, quality, and outcomes study, *Merrill—Palmer Quarterly, 43*, 451-477.
- Phillips, D.A. (Ed.). (1995). *Child care for low-income families: Summary of two workshops*. Washington, DC: National Research Council, Institute of Medicine.
- Pianta, R. C. (2003, February). *Teacher-Child Interactions: The implications of observational research for re-designing professional development*. Presentation given at the Science and Ecology of Early Development (SEED) meeting: Development of a research agenda: Early childhood professional development/training and children's successful transition to elementary school, Washington, DC.
- Porter, T., Habeeb, S., Mabon, S., Robertson, A., Kreader, L., & Collins, A. (2002). *Assessing child care development fund (CCDF) investments in child care quality: A study of selected state initiatives*. New York: Bank Street College of Education.
- Rogers, P.T., Petrosino, A., Huebner, T. A., & Hacs, T.A. (2000). Program theory evaluation: Practice, promise, and problems. *New directions for evaluation, 87*, 5-13.
- Rossi, P.H., Freeman, H.E., & Lipsey, M.W. (1999). *Evaluation: A systematic approach*, (6th Ed.). Thousand Oaks, CA: Sage Publications.
- Ruopp, R., Travers, J., Glantz, F., & Coelen, C. (1979). *Children at the center: Final report of the national day care study*. Cambridge: Abt Associates.
- Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park, CA: Sage Publications.
- Tout, K., Berry, D., & Zaslow, M. (2003, February). *Professional development/training and the quality of early care and education environments*. Paper presented at the Science and Ecology of Early Development (SEED) meeting: Development of a research agenda: Early childhood professional development/training and children's successful transition to elementary school, Washington, DC.
- U.S. General Accounting Office. (2002, September). *States have undertaken a variety of quality improvement initiatives, but more evaluations of effectiveness are needed*. GAO-02-897. Washington, DC.
- Vandell, D. L., & Wolfe, B. (2000). *Child care quality: Does it matter and does it need to be improved?* Madison, WI: University of Wisconsin-Madison.
- Whitebook, M., Howes, C., & Phillips, D. (1989). *Who cares? Child care teachers and the quality of care in America: Final report: National child care staffing study*. Berkeley, CA: Child Care Employee Project.
- Whitebook, M., Sakai, L., Gerber, E., & Howes, C. (2001). *Then and now: Changes in child care staffing 1994-2000*. Washington, DC: Center for the Child Care Workforce.

Appendix A: Selected Evaluation Descriptions

Initiative Purpose

The California Child Care Healthline is a statewide toll-free telephone service to assist providers and parents who have concerns regarding health and safety issues. This initiative strives to provide the child care community with current, accurate and relevant health, safety and development information. Staff respond to telephone calls from the community on health and safety issues; provide linkages to services for children in child care; maintain a health and safety resource library; and provide child care health and safety information to geographic areas that are underserved.

Evaluation Questions

- Are the outreach strategies effective in reaching a broad range of people?
- Are the services effectively promoting child health and safety in the community?

Budget and Timeline

Five years overall. The report details findings for year one.

Advisory Committee/Stakeholders

A representative from a community college, a CCR&R, a child care center, and a child development service agency.

Measures

- Success of outreach
- Satisfaction with services

Data Sources

- Healthline telephone records
- Healthline users

Data Collection Strategy

- Document review
- Surveys
- Interviews

Initiative Purpose	<p>Designed to address the shortage of licensed quality child care in California communities, the program involves a five-stage recruitment and training process to build supply. The objectives include assessing child care supply and demand and targeting shortages for care in specific areas; recruiting individuals who have the potential to become licensed family child care providers; training those individuals to deliver quality care and effectively manage a small business; providing technical assistance to help participants become licensed and begin operations; and providing ongoing support to help family child care providers stay in operation.</p>
Evaluation Questions	<ul style="list-style-type: none"> • Are the CCIP programs addressing areas of the greatest need as determined by the concentration of families receiving TANF? • Is CCIP effective in building the capacity of family child care? • What are the barriers and facilitating factors in achieving CCIP program goals?
Budget and Timeline	<p>One year of a three-year evaluation of multiple programs.</p>
Advisory Committee/Stakeholders	<p>A representative from a community college, a CCR&R, a child care center, and a child development service agency.</p>
Measures	<ul style="list-style-type: none"> • Changes in licensing status • Retention in the child care field
Data Sources	<ul style="list-style-type: none"> • Program narratives, fiscal reports, participant lists, internal evaluation forms, and descriptive reports • CCIP participants • CCIP program directors • Child Care Development Division staff • CCR&R staff
Data Collection Strategy	<ul style="list-style-type: none"> • Documentation review • Surveys • Interviews • Observations

Initiative Purpose	Child Care Plus Maine aims to meet the needs of children who are at risk of expulsion from child care programs because of challenging behaviors. The initiative provides extra support to caregivers of these children through funds for facilities improvement, materials and equipment and professional development.
Evaluation Questions	<ul style="list-style-type: none"> • What are the characteristics of the child care facilities requesting technical assistance (TA)? • What is the nature of TA requests? • To what extent are the TA goals and objectives achieved as reported by provider and staff? • What is the reported outcome of the TA for children? • To what extent are parents able to be relieved of providing child care? • What is the nature of collaborations for children with special needs? • To what extent did the TA result in child retention?
Budget and Timeline	Not available
Advisory Committee/Stakeholders	Not available
Measures	<ul style="list-style-type: none"> • Satisfaction • Change in knowledge and skills, short-term and intermediate
Data Sources	<ul style="list-style-type: none"> • Existing forms • Technical assistance forms • Project staff observation • Parents of special needs children
Data Collection Strategy	<ul style="list-style-type: none"> • Document review • Other paperwork • Interviews • Observation

Initiative Purpose

The School Readiness (SR) program aims to enable child care programs to meet quality components established by the School Readiness legislation (collaboration with community services; parent involvement, education and outreach; referrals for health services including appropriate immunizations and screenings; nutrition services; family literacy; transition planning for kindergarten; professional development; sliding fee scale; admissions policy; annual evaluation; and serving children with disabilities) as well as to encourage accreditation. The combination of School Readiness reimbursement funds, state child care subsidies and parent fees must provide all the funding child care programs need to implement strategies as well as program changes and improvements that will allow them to meet the quality components mandated by the legislation.

Evaluation Questions

- Who is served by SR centers?
- What has been the impact of the SR initiative on the supply of center-based child care?
- Have SR center quality and staff qualifications improved over time as would be expected if the program is effective?
- Has the quality of care and education in SR classrooms improved over time as would be expected if the program is effective?
- Have child outcomes improved over time as would be expected if the program is effective?
- What are the major obstacles to program improvement?
- Is there support for the causal logic underlying the SR initiatives: improved center quality and staff qualifications → improved quality of care and education for the children → improved child outcomes?

Budget and Timeline

The timeline for the project is seven years.

Advisory Committee/Stakeholders

Not available

Measures

- Participant demographics
- Change in child care supply
- Assessment of school readiness
- Quality improvement measured by staff turnover, barriers to staff training, and technical assistance

Data Sources

- Center directors
- Teachers
- Parents
- Four children from each pre-kindergarten center and classroom, selected at random

Data Collection Strategy

- Surveys
- Standardized assessments and observations

Initiative Purpose

The goal of the Michigan R.E.A.D.Y. initiative is to strengthen parent and caregiver involvement in the early childhood years so that children develop the language and literacy skills needed to enter school ready to read and succeed by providing kits of materials to the parent or provider. The kit contains both age-appropriate and general information and materials.

Evaluation Goals

To determine how the R.E.A.D.Y. kit was received by the target audience; to determine the value and usefulness of the kit; to assess the appeal and effectiveness of the kit materials; and to obtain input for improvements to kit material content and delivery.

Budget and Timeline

Not available

Advisory Committee/Stakeholders

Not available

Measures

- Benefits of program to providers
- Satisfaction
- Effectiveness of the program materials
- Assessment of program material content and delivery

Data Sources

- Participants who received one or more R.E.A.D.Y. kits

Data Collection Strategies

- Surveys
- Interviews
- Focus groups

Initiative Purpose	The purpose of the initiative is to increase staff education and compensation and reduce turnover. These are reached by providing scholarships to workers to help pay for classes, books, and transportation. It also helps pay for substitute teacher costs. When the providers complete their education goals, they receive a bonus or a pay increase as long as they commit to stay at their program for an additional year.
Evaluation Goal	To provide information about the education among providers, changes in compensation, and turnover.
Budget and Timeline	The evaluation is ongoing.
Advisory Committee/Stakeholders	Not available
Measures	<ul style="list-style-type: none"> • Change in staff education • Effects on compensation • Turnover rate
Data Sources	<ul style="list-style-type: none"> • Participant applications, grade reports, wage reports, child care sponsor information sheets, state databases • Employers
Data Collection Strategy	<ul style="list-style-type: none"> • Document review • Other paperwork • Interviews

Initiative Purpose

The initiative provides TANF recipients with the community college credits and field experience they need to become child development teachers, simultaneously increasing the number of qualified teachers in California and supporting welfare recipients in becoming economically self-sufficient. Each program is a collaborative arrangement among the CCR&Rs, community colleges, and county welfare departments. To ensure that the program runs smoothly, welfare departments set the eligibility criteria and track and refer the TANF recipients.

Evaluation Questions

- Is the program successful in training TANF recipients to become qualified child development teachers?
- What factors underlie student attrition in the program?
- Does the program offer effective services that support students' participation in the program?
- Does the program support TANF recipients in becoming more self-sufficient?
- What are the facilitating factors and barriers to achieving program goals?

Budget and Timeline

Not available

Advisory Committee/Stakeholders

A representative from a community college, a CCR&R, a child care center, and a child development service agency.

Measures

- Numbers of participants who completed the program
- Satisfaction
- Change in number of qualified teachers
- Change in welfare recipients' employment status
- Change in income

Data Sources

- Students' individual records
- Sites' monthly and annual progress reports
- Students
- Program coordinators
- Campus-level coordinators

Data Collection Strategy

- Document review
- Surveys
- Interviews
- Focus groups

Brief Description of initiative.

The Quality Training Incentive Project (QTIP) was a professional development and incentive project designed by the Kansas Association of Child Care Resource and Referral Agencies (KAC-CRRA) Infant/Toddler Project. The QTIP delivered professional development events across the state of Kansas distributing valuable incentives to encourage participation.

The QTIP was implemented to respond to identified needs from a statewide assessment of quality using the Family Day Care Rating Scale (FDCRS) and Infant Toddler Environment Rating Scales (ITERS) in almost 200 sites. The QTIP topics of the professional development events correlated with the lowest scores of quality in the areas of routine care and cultural sensitivity. Professional development events also included distribution of materials and equipment to improve the quality of infant and toddler child care in these two areas.

The evaluation we propose is to conduct follow-up observations using the rating scales to determine if the quality of care was enhanced in those child care sites who participated in the QTIP.

State briefly the purpose of the initiative.

The purpose of the initiative was to respond to identified needs from a statewide assessment of quality using the FDCRS and ITERS in almost 200 sites. The QTIP topics of the professional development events correlated with the lowest scores of quality in the areas of routine care and cultural sensitivity.

The purpose of the evaluation is to document possible changes in quality of care offered in infant and toddler child care settings who participated in the QTIP. Documentation of results would be used to support current and future funding to continue investing in quality initiatives.

Please list the individuals who would be members of your evaluation team.

Early Learning professionals who participated in the QTIP, QTIP trainers (R&R Infant/Toddler Specialists), Infant/Toddler Project Director, Department of Social and Rehabilitation Services (SRS) Grant Manager, an evaluation professional, a parent, and a legislator.

Please list the individuals who would be members of your advisory committee.

A legislator, the Infant/Toddler Project Director, a researcher, a parent, the SRS grant manager, a QTIP participant, an Infant/Toddler Specialist, a possible funder, a Kansas Children’s Cabinet member.

State briefly your evaluation questions.

1. Did the quality of care improve after participation in the QTIP training?
2. Did the quality of care improve after receipt of the QTIP incentives?
3. Were there differences in family child care settings and child care centers?
4. Was there a difference in quality for those who participated in one QTIP event versus those who participated in both QTIP events?

Please indicate the proposed budget for this evaluation.

	\$10,000
Observation and data collection	\$4,600
Report writing	2,500
Participant incentives	1,250
Administrative costs	1,650

Please describe the proposed timeline for your evaluation.

Completion of data collection, data analysis and report writing four to six months. Review and selection of observation sites–1 to 2 months. Observation and data collection–2 months. Analysis and report writing–1 month.

Please identify the sources of data you would use in your evaluation.

Participants would be identified from records of QTIP professional development events and records of materials received. These lists would be correlated with previous quality data collection efforts conducted by the Infant/Toddler Project evaluation and other recent studies for baseline data. Follow-up evaluation using the FDCRS and ITERS would be conducted for comparison data.

Please explain how you would collect your data and whether you plan to use Instruments from the *Toolkit* for any of the data collection.

Existing data on participation in professional development events will be used to determine selection of child care providers who would receive a second observation using the FDCRS and ITERS. Existing records of materials and equipment received at events would be used to correlate pre- and post- scores related to areas of ITERS and FDCRS. Individuals trained to use the FDCRS and ITERS would collect quality ratings on those providers who participated in the QTIP events. The ITERS & FDCRS would be used in this evaluation plan.

Please describe your data analysis plan, including how you plan to manage your data.

Scores from the FDCRS and ITERS would be used for comparison before participating in the QTIP and after participation in QTIP events. An evaluation consultant would import the data in SPSS to analyze the comparison and write a report on the findings. All participation and observation records will be kept confidential.

Please describe any additional information about your evaluation plan.

A Quality Child Care Study of the Infant/Toddler Project was published in Spring, 2003. This document contains ITERS/FDCRS scores for 196 providers in Kansas who participated in the Infant/Toddler Project. Some of the same providers also participated in QTIP. These providers will be the ones recruited for participation in the QTIP evaluation. For these providers, the data from the Quality Child Care Study will be used as the baseline data for the QTIP evaluation.

Brief Description of initiative.	ODJFS awards a grant to Ohio Educational Television to provide training to family child care providers across the state of Ohio. They also create and run daily educational vignettes targeted to family child care providers.
State briefly the purpose of the initiative.	We believe that Ready To Learn can increase the foundational knowledge of family child care providers which in turn will improve their instructional practices and instructional environments which ultimately will improve the school readiness of children in family child care arrangements.
Please list the individuals who would be members of your evaluation team.	ODJFS, ODE, family child care provider, OETS.
Please list the individuals who would be members of your advisory committee.	ODJFS, ODE, DCAC, family child care provider, OETS, OCCRRRA
State briefly your evaluation questions.	<ol style="list-style-type: none"> 1. Do the Ready to Learn workshops change provider instructional practices? 2. Are the providers planning curriculum? 3. Are the providers incorporating the Early Learning Content Standards into their curriculum? 4. Can providers articulate why they choose one experience over another? 5. Do providers know why vocabulary, phonemic awareness and reading aloud are critical experiences for young children? 6. Do the Ready to Learn workshops change the instructional environments of family child care providers? 7. Is the environment organized and well planned? 8. Are there open-ended materials? 9. Are there numerous opportunities/materials to explore language and literacy?
Please indicate the proposed budget for this evaluation.	The entire grant amount is approximately for \$820,000 per year for two years
Please describe the proposed timeline	July 2003-June 2005
Please identify the sources of data you would use in your evaluation.	<ul style="list-style-type: none"> • Utilize a pre-test and post-test measure for change in provider knowledge and practice. • Utilize a pre-test and post-test measure for change in home environment.
Please explain how you would collect your data and whether you plan to use Instruments from the <i>Toolkit</i> for any of the data collection.	Data will be collected by the coordinators at the eight PBS stations after each workshop and then sent to the statewide coordinator for aggregation. The statewide coordinator would then forward the information to the Department of Jobs and Family Services (DJFS) quarterly. DJFS is planning to use the <i>Toolkit</i> as a template.
Please describe your data analysis plan, including how you plan to manage your data.	We will then evaluate the data against our key indicators of success and look for ways to continuously improve the program to meet our desired outcomes.
Please describe any additional information about your evaluation plan.	

Brief Description of initiative.	Washington State Training and Registry System (STARS) is a career development system for those who work in child care, early education and school-age care. It links providers to approved trainers and training organizations who offer training and provides scholarships to child care providers to take the trainings.
State briefly the purpose of the initiative.	The purpose of Washington STARS is to ensure quality child care for all children in the state through increased opportunities for basic and ongoing training for child care, early education and school-age care providers.
Please list the individuals who would be members of your evaluation team.	Sonja Griffin, Agda Burchard, Judy Serquinia, Metesa Greene, Cecelia Callison and Lijian He
Please list the individuals who would be members of your advisory committee.	STARS Stakeholders (This advisory group meets every other month and represents all aspects of the child care community).
State briefly your evaluation questions.	<ol style="list-style-type: none"> 1. Did the number of trainings increase? 2. Did the number of trainings available to providers in each county increase? 3. Did the number of distance learning trainings increase? 4. Did the number of providers participating in and completing the training increase? 5. Did the number of trainings available in languages other than English increase? 6. Did the number of newly licensed child care centers and family child care home providers who were in compliance with basic health and safety regulations increase after completing the 20-hour basic training? 7. Did participants demonstrate an increased knowledge in early childhood education after completion of the 10-hour continuing training?
Please indicate the proposed budget for this evaluation.	\$250, 000
Please describe the proposed timeline	<p>18-20 months</p> <p>Fall 2003 –</p> <ul style="list-style-type: none"> • Develop and test data base queries and reports for outcomes #1-5 • Develop tool for data collection from licensor check lists for outcome #6 <p>Winter 2004 –</p> <ul style="list-style-type: none"> • Finalize and begin implementation of data base queries and reports for outcomes #1-5 • Implement check list tool for outcome #6 <p>Spring/Summer 2004 –</p> <ul style="list-style-type: none"> • Collect and analyze data for outcomes #1-5 and #6 • Develop and test pre-post test for outcome #7 <p>Fall 2004 –</p> <ul style="list-style-type: none"> • Implement pre-post test for outcome #7 <p>Winter 2004 –</p> <ul style="list-style-type: none"> • Analyze data for outcome #7 • Report evaluation findings outcomes #1-7

Please identify the sources of data you would use in your evaluation.

- Registry reports
- CAMIS reports
- Licensor check lists
- Pre-post test

Please explain how you would collect your data and whether you plan to use Instruments from the *Toolkit* for any of the data collection.

1. For outcomes #1-5, we will develop tools to collect data on an annual basis from the whole population of the sample groups:
 - The STARS Registry data base
 - * Training Information Report
 - * Provider Profile Report
 - CAMIS data base
 - * Provider Report
2. For outcome #6, we will review licensors' check lists for child care centers and family child care homes newly licensed in January 2004 to collect data on compliance for health and safety regulations
3. For outcome #7, we will give two tests on knowledge of child care and development, one before and one after, to participants completing the 10-hour training in the fall of 2004. The test scores will be collected and entered into the database for analysis.

Please describe your data analysis plan, including how you plan to manage your data.

For outcomes #1-6 we will:

1. Design reports to query the data set in a quantitative method
2. Analyze data from the Registry reports to get counts of statewide total numbers of approved trainings over the time period
3. Plot the data to visualize the change over the time period
4. Employ statistical techniques/models to conduct significance test of difference in means and to detect the trend (i.e. t-test and paired t-test)

For outcome #7 we will:

1. Conduct a multiple regression analysis to test the effect of the 10 hour trainings, controlling for participants' past experience and education and delivery method of training
2. Plot the before and after scores to visualize the effectiveness of the 10 hour trainings
 - Queries and reports against the databases will be designed and implemented by Information Technology staff
 - Pre-post test will be sent to trainers for administration; they may be offered an incentive
 - Scores will be uploaded into the Registry; a data base will be created to collect the information
 - Data analysis will be conducted by Research and Data Analysis staff
 - DCCEL Program Manager and STARS staff will manage project

Please describe any additional information about your evaluation plan.

Brief Description of initiative.

State briefly the purpose of the initiative.

Please list the individuals who would be members of your evaluation team.

Please list the individuals who would be members of your advisory committee.

Theory of Change Model

Please create a Theory of Change Model on a separate sheet.

State briefly your evaluation questions.

Please indicate the proposed budget for this evaluation.

Please describe the proposed timeline for your evaluation.

Please identify the sources of data you would use in your evaluation.

Please explain how you would collect your data and whether you plan to use instruments from the *Toolkit* for any of the data collection.

Please describe your data analysis plan, including how you plan to manage your data.

Please describe any additional information about your evaluation plan.

Appendix D: Other Resources

Other Resources

- Administration on Children, Youth and Families. (1997). *The Program Manager's Guide to Evaluation* [on-line]. Available: <http://www.acf.hhs.gov/programs/core/pubs_reports/prog_mgr.html>.
- Curnan, S., LaCava, L., Langenburg, D., Lelle, M., & Reece, M. (1998). *W.K. Kellogg Foundation evaluation handbook*. Battle Creek, MI: W.K. Kellogg Foundation.
- Hatry, H. P., Winnie, R. E., & Fisk, D. M. (1981). *Practical program evaluation for state and local governments*. Second edition. Washington, DC: The Urban Institute Press.
- Herman, J. L., Morris, L. L., & Fitz-Gibbon, C. T. (1987). *Evaluator's handbook*. Los Angeles, CA: Center for the Study of Evaluation, University of California, Los Angeles.
- National Center for Clinical Infant Programs. (1987). *Charting change in infants, families and services: A guide to program evaluation for administrators and practitioners*. Washington, DC: Zero to Three.
- Park-Jadotte, J., Golin, S. C., & Gault, B. (2002). *Building a stronger child care workforce: A review of studies of the effectiveness of public compensation initiatives*. Washington, DC: Institute for Women's Policy Research.
- Shontz, F., Goldman, R., Hellmer, T., & Torxel, M. (1993). *Children's trust fund – Evaluation manual*. Kansas City, MO: The Greater Kansas City Community Foundation.
- United Way. (1996). *Measuring program outcomes: A practical approach*. Alexandria, VA: United Way of America.
- U.S. Department of Health and Human Services, Office of Cancer Communications, National Cancer Institute. (1989). "Chapter 5: Assessing Effectiveness." In *Making health communication programs work: A planner's guide*. Bethesda, MD: National Cancer Institute
- U.S. General Accounting Office (GAO). Program Evaluation and Methodology Division. P.O. Box 6015, Gaithersburg, Maryland 20884-6015. (202) 512-6000. This office publishes a series of evaluation handbooks that address different topics. Single copies are available free from GAO. The series includes the following:
Designing Evaluations GAO/PEMD-10.1.4
Using Structured Interviewing Techniques GAO/PEMD-10.1.5
Developing and Using Questionnaires GAO/PEMD-10.1.7
Case Study Evaluations GAO/PEMD-10.1.9

The Advisory Committee

To help us develop the *Toolkit*, we asked seven state child care administrators to serve on an advisory committee. They were Jeanetta Green, Alabama; Mary Smithberger, California; Amparo Garcia, Connecticut; Janet McKeon, Massachusetts; Deb Swenson-Klatt, Minnesota; Tom Olsen, Oregon; and Dave Edie, Wisconsin. In addition to the child care administrators, three child care experts also served on the Committee: Marty Zaslow of Child Trends and Shari Gruber of AHPA and Ivelisse Martinez-Beck of the Child Care Bureau. Later in the year, Gwen Stephens replaced Mary Smithberger and Laura Satterfield replaced Dave Edie.

Creating the *Toolkit*

Developing The Child Care Quality Logic Models

The Research Team's first task in developing the *Toolkit* was to identify anticipated outcomes from the quality improvement initiatives we had documented in the first year study. (Please see *A Study of Selected State Initiatives*.) We reviewed the 104 initiatives and created a rough matrix of the types of strategies and their expected results.

Then we organized the eight quality improvement strategies into three general categories: transfer of knowledge (professional development, training, technical assistance and mentoring); supply-building (start-up, recruitment, facilities improvement and materials and equipment); and other (compensation and accreditation and rating). We also divided the anticipated results into two categories: implementation (formative) and outcomes (summative).

We shared the revised matrix with the members of the advisory committee to elicit their reactions to it as a model for matching outcomes to quality improvement strategies. They helped us translate it into the table we developed to identify outcomes for specific populations for each strategy. (Please see Part II, Table 1.)

Collecting The Instruments

Of the 104 initiatives included in the first year study, 75 conducted evaluations. We collected 49 instruments from these evaluations. Using the preliminary matrix, we organized the instruments by strategy and outcome to identify gaps for which no instruments were available. Instruments were missing in several categories. For example, there were no instruments for initiatives that used compensation as a strategy

and only one instrument—a satisfaction survey—for initiatives that used start-up funding. In addition, none of the efforts that relied on distributing materials and equipment used instruments to assess the impact on knowledge and skills.

Selecting the Exemplary Instruments

To select instruments for the *Toolkit*, we created a set of criteria based on outcomes, type of instrument, and utility. Instruments were rated on a scale from one to three, with three representing the highest rating. Of the 49 instruments, we scored 22 as “three.” We asked the advisory committee to review the preliminary set. Of the 22, they selected seven to include in the *Toolkit*.

Collection of Additional Exemplary Instruments

Using the logic models we had developed for each strategy, we identified missing instruments. To fill in the gaps, the advisory committee suggested several instruments that had been used in their own states as well as other data collection instruments that we could adapt. These included surveys from the Center for the Child Care Workforce, the Census, the National Association of Child Care Resource and Referral Agencies.

The resulting set consisted of 14 exemplary instruments. We modified several of them, and created several new ones. We also drafted head notes for each instrument with guidelines on how to use it.

We distributed the preliminary draft of the instrument set to the advisory committee for their review. We asked the state child care administrators to consider how they would use the *Toolkit* with two CCDF quality set-aside initiatives, and the other members about how they would use it in their own work. Each member completed a survey with questions about whether the instruments collected appropriate data, were easy to use, and would meet their needs.

Developing Part I

To provide a context for using the instruments, we developed a general guide for evaluating child care quality improvement initiatives. It is intended to supplement other manuals on conducting evaluation of social service programs. The guide includes an overview of the principles of evaluation, a description of a theory of change model as an approach for planning and evaluating efforts to improve child care quality, and a brief discussion of design issues. Child care quality improvement initiatives are used to illustrate each section.

Final Review and Field Test

We asked 11 states to review the *Toolkit*. They are: Georgia, Illinois, Kansas, Maine, New Mexico, North Dakota, Ohio, Rhode Island, Utah, Vermont and Washington. The reviewers offered comments and suggestions that we used to revise the preliminary draft. In addition, three

states—Kansas, Ohio, and Washington—participated in a field test of the *Toolkit* to determine if it could be used independently to develop evaluations of efforts to improve child care quality. We documented the process they used through weekly phone calls that focused on successes and issues. In addition, we convened a meeting of the sites in Washington D.C. to discuss their experience as well as to elicit suggestions for revising the *Toolkit*. We integrated the results into the final draft.