

21st Century Community Learning Centers Grant Monitoring Support

Contract No. ED-04-CO-0027
Task Order No. 0005

For the U.S. Department of Education,
Office of Elementary and Secondary Education

Review of 21st Century Community Learning Centers' Grantee Evaluation Practices *Final Report*

September 24, 2012

Submitted to:

Office of Elementary and Secondary Education
U.S. Department of Education
400 Maryland Avenue, SW
Washington, DC 20202



Table of Contents

I. INTRODUCTION.....	1
Purpose and Overview of the Evaluation Review Task.....	1
Overview of the Evaluation Review Methodology.....	2
Overview of this Report.....	2
II. EVALUATION REVIEW METHODOLOGY	4
Evaluation Review Questions	4
Summary of Data Collection Sources	4
Summary of Analysis Strategies	5
III. 21ST CCLC PROGRAM EVALUATION APPROACHES	8
Frequency of Evaluation	8
Types of Evaluators	9
Evaluation Cost	13
Key Evaluation Questions	15
Evaluation Measures Used by Grantees	19
Evaluation Research Designs.....	27
Changes in Evaluation Approaches over Time.....	38
Promising Practices in Evaluation Approaches	42
Summary.....	43
IV. KEY FINDINGS FROM 21ST CCLC EVALUATIONS	45
Key Findings from SEA Evaluations	45
Key Findings from Sub-Grantee Evaluations.....	46
Summary.....	47
V. USE AND DISSEMINATION OF EVALUATION RESULTS.....	49
SEA's Use of Evaluation Results	49
Sub-Grantees' Use of Evaluation Results.....	50

SEA Dissemination of Evaluation Results	51
Sub-Grantees' Dissemination of Evaluation Results	52
Promising Practices of Use and Dissemination	53
Summary	54
VI. EVALUATION GUIDANCE AND TECHNICAL ASSISTANCE PROVIDED BY SEAS TO SUB-GRANTEES.....	55
Materials.....	55
Instructions for Data Entry/Reporting.....	55
QIP Training	56
Conferences/Meetings/Trainings	56
Strong Guidance/ TA.....	57
Changes in Requirements and Guidance	60
Summary	61
VII. GRANTEE EVALUATION NEEDS AND CHALLENGES	62
Challenges in Conducting State Comprehensive Evaluations.....	62
SEA-Identified Technical Assistance and Support Needs	64
Recommendations for Methods of Delivering Technical Assistance to SEAs.....	66
Summary	67
VIII. IMPLICATIONS FOR DEPARTMENT OF EDUCATION: RECOMMENDATIONS FOR EVALUATION TA AND GUIDANCE TO SEAS	68
Selecting a Qualified Evaluator	68
Articulating Program Goals and Measurable Objectives	68
Using Designs Appropriate for Measuring Program Quality and Effectiveness	69
Analysis and Reporting	69
Use of Evaluation Results	70
IX. SUMMARY AND CONCLUSIONS.....	71
Extent to Which Grantees and Sub-Grantees Conduct Evaluation	71

Review of 21st Century Community Learning Centers' Grantee Evaluation Practices
Final Report

Evaluation Approaches 72

Findings of State and Local Program Evaluations 73

SEA Guidance and Technical Assistance to Sub-Grantees 73

Use and Dissemination of Evaluation Results 74

Challenges and Needs 74

BPA Recommendations 75

REFERENCES..... 76

APPENDICES..... 77

Appendix A: Detailed Evaluation Review Methods and Procedures

Appendix B: Analysis Plan

Appendix C: Evaluation Review Documents Requested

Appendix D: Interview Guide

Appendix E: Codebook for Reviewing Documents

Appendix F: Document Review Instructions and Guidance to Reviewers

Appendix G: Detailed Information on 21st CCLC Evaluation Measures Used by SEAs and Sub-Grantees

Appendix H: Evaluation Framework for 21st CCLC Programs

Review of 21st Century Community Learning Centers' Grantee Evaluation Practices

Final Report

September 24, 2012

I. Introduction

The 21st Century Community Learning Centers (CCLC) program, funded by the U.S. Department of Education, is designed to assist communities in creating learning centers that provide academic enrichment opportunities during non-school hours¹ for children, especially those who attend high-poverty and low-performing schools. The assistance is provided through grants to state education agencies (SEAs) that in turn hold their own competitions to fund local projects through state-administered grants. The 21st CCLC program supports community efforts to:

- Help students meet state and local standards in core academic subjects, such as reading and math;
- Offer students a broad array of enrichment activities that can complement their regular academic programs; and
- Offer literacy and other educational services to the families of participating children.

SEAs are responsible for monitoring their sub-grantees and evaluating their programs and activities. As part of this responsibility, SEAs are required to:

1. "Conduct a comprehensive evaluation (directly, or through a grant or contract) of the effectiveness of programs and activities provided with 21st CCLC funds."²
2. "Monitor the periodic evaluations of local programs" and "disseminate the results of these evaluations to the public."³ Each sub-grantee must "assess its progress toward achieving its goal of providing high-quality opportunities for academic enrichment" and "use the results to refine, improve, and strengthen the program and to refine the performance measures. Local grantees, working with their SEAs, must evaluate the academic progress of children participating in the 21st CCLC program."⁴

Purpose and Overview of the Evaluation Review Task

The Department of Education has contracted with Berkeley Policy Associates (BPA) and its partner organizations, Westat and Learning Point Associates (LPA) (now American Institutes for

¹ In 2012, the Department allowed state education agencies to apply for a waiver that would allow grantees to use 21st CCLC funds for academic enrichment during school hours as well.

² §4203(a)(13) A-B; Non-regulatory guidance H-5

³ Ibid.

⁴ §4203(a)(13) A-B; Non-regulatory guidance H-6

Research/AIR), to support monitoring and program evaluation review efforts across all states and jurisdictions receiving 21st CCLC funds. The purpose of this review is to ensure compliance with §4203(a)(13)A-B and §4205(b)(2)A-B of the Elementary and Secondary Education Act of 1965, as amended. The Department is also interested in identifying areas for Department support for and technical assistance with conducting effective program evaluations.

Specifically, the objectives of this task include:

- Determine the extent to which states and sub-grantees are conducting evaluations;
- Describe the programs' approaches to evaluation and the comprehensiveness of the evaluations being conducted;
- Provide information about how evaluation results are used to inform program improvements; and
- Identify promising practices, technical assistance and types of support that can help improve state and sub-grantee evaluation efforts.

Overview of the Evaluation Review Methodology

This review required examining the evaluation practices of approximately 18 states and jurisdictions annually in order to complete a review of all 54 states/jurisdictions over the three-year contract period.

To address this task, the BPA team reviewed the evaluation practices of 54 SEAs and a representative sample of sub-grantees from 2010 to 2012. In the third year of contract, the team also reviewed updated state-level documents for states that were part of the 2010 sample. Both qualitative and quantitative data were collected and analyzed as part of this task. In addition to reviewing evaluation documents, the team conducted interviews about evaluation practices with program staff and evaluators from SEAs and sub-grantees.

Overview of this Report

This final report summarizes the work completed for Task 2.3, Review of 21st CCLC Grantee Evaluation Practices, during BPA's three-year contract with the Department of Education. This report builds upon the past two interim reports (2010 and 2011) and expands data analysis to include all three years of state and sub-grantee data.

This report consists of nine chapters and eight appendices. The first chapter has introduced the overall purpose of the project and the task. In the second chapter, the BPA team presents the methodology used for conducting the review. Chapter 3 describes the evaluation approaches of all SEAs and the sub-grantee sample. Chapter 4 presents key findings of the grantee and sub-grantee program evaluations, and Chapter 5 describes the ways that programs used and disseminated evaluation results. Chapter 6 describes the evaluation guidance and technical assistance provided by grantees, including examples of particularly strong guidance. Chapter 7 describes some challenges SEAs have with program evaluation and particular needs for guidance. In Chapter 8, the BPA team provides recommendations for the Department in providing guidance and support to SEAs related to program evaluation in the future. Chapter 9 presents a summary of the report. The appendices include: detailed evaluation review methods and procedures (Appendix A), the BPA team's detailed analysis plan (Appendix B), memos requesting documents sent to SEAs (Appendix C), interview guides for on-site discussions with

SEA and sub-grantee staff and evaluators (Appendix D), codebook used for reviewing documents (Appendix E), guidance and instructional materials provided to the BPA analysis team for reviewing and coding documents (Appendix F), detailed information about evaluation measures used by SEAs and sub-grantees (Appendix G) and the *Evaluation Framework for 21st CCLC Programs* developed by BPA (Appendix H).

II. Evaluation Review Methodology

This chapter is divided into three sections. The first section presents the key questions guiding the Review of 21st CCLC Grantee Evaluation Practices. The second section describes the major data sources (i.e., state evaluation-related documents and on-site interviews) used to answer the research questions about grantee evaluation practices. The third section summarizes the team's analysis strategies. For a detailed description of methods and procedures, see Appendix A.

Evaluation Review Questions

During 2011-2012, BPA finalized the key questions related to this task as well as the analysis plan to guide the review of evaluation practices. This year, BPA combined two analysis plans (one for the state level and one for the sub-grantee level) into one comprehensive plan to address the key questions (see Appendix B). The evaluation review questions are based on the goals of the project and discussions between BPA and the Department about the specific objectives of this task listed in Chapter 1 above.

The key questions guiding this evaluation review are:

1. To what extent are SEAs and sub-grantees conducting evaluations of their 21st CCLC programs?
2. What is the range of approaches that are being used to evaluate 21st CCLC programs?
3. What research designs do programs use to evaluate student outcomes?
 - a. How does research design vary by selected SEA characteristics?
 - b. How have evaluation practices been changing over time?⁵
 - c. What are promising practices in evaluation among 21st CCLC grantees?
4. What are the key findings from 21st CCLC program evaluations?
5. What guidance and technical assistance are SEAs providing to sub-grantees regarding program evaluation, and what topics do they cover?
6. How do programs use evaluation and disseminate evaluation results?
7. What information or support do grantees need to conduct more rigorous evaluations?

Summary of Data Collection Sources

Two main sources of data informed the review of evaluation practices: 1) evaluation-related documents obtained from the SEAs; and 2) interviews conducted during monitoring site visits about evaluation practices. Other sources include brief interviews conducted via desk monitoring follow-up phone calls, and selected state profiles.

- **State and Sub-Grantee Documents.** Evaluation documents produced by states, sub-grantees and/or third-party evaluators provided key information about the range of

⁵ This question came about in 2012 as the BPA team gathered updated information from SEAs that were reviewed in the first year of the study to learn about how program evaluation had changed in the last two years. Addressing this question also includes a summary of lessons learned and plans for improving future evaluation efforts that SEAs shared in the second and third year of the study with the BPA team.

approaches used by states and sub-grantees to evaluate their programs. Over the three years, documents were collected from 54 SEAs and a sample of their sub-grantees. The full list of documents requested of SEAs and sub-grantees is included as Appendix C. In general, the types of documents reviewed included:

- Evaluation reports;
 - SEA requests for proposals (RFPs) or grant solicitation packages;
 - Scopes of work for evaluators;
 - Guidance materials and documentation of technical assistance provided by SEAs to sub-grantees; and
 - Evidence of use and dissemination of evaluation results.
- **Interviews Conducted during On-Site Monitoring Visits.** Interviews were conducted with SEA representatives and evaluators for all 54 SEAs that received on-site monitoring visits from the Department and BPA's monitoring team over the three years of the contract. SEA Coordinators were asked to provide their most recent evaluation reports, grant applications and evaluation reports from the sub-grantees selected to be included in the on-site visit. During these visits, the designated BPA/AIR Evaluation Lead interviewed program staff and evaluators about SEA and sub-grantee evaluation practices (see Appendix D for the Interview Guides). The information gathered on site provided valuable context and background for understanding the documents being reviewed.
 - **Additional Information from 2010 Sample.** For states that were part of the Evaluation Review sample in 2010 (Base Year), additional sources of data were used in the final analysis, including updated SEA documents, follow up phone interviews and state profiles. In 2011 (Option Year 1), a set of updated evaluation documents was collected from nine of the 2010 states as part of the monitoring process. In 2012 (Option Year 2), a set of updated evaluation documents was collected from the other nine 2010 states. In addition, BPA conducted follow-up phone calls with some of the 2010 states in July 2012 to gather updated information on evaluation practices and challenges, and changes to evaluators, design and guidance to sub-grantees. Finally, state profiles (developed as part of each monitoring visit) were also analyzed for 2010 states, since structured interview summaries were not submitted in the first year of the review. In 2011 and 2012, structured and comprehensive interviews were conducted with each SEA and notes were written up and submitted to the Task Leader, so review of the state profiles was not necessary for these more recently visited states.

Summary of Analysis Strategies

Selection of States and Sub-Grantees for Document Review

During each year of the study, evaluation practices for approximately one third of the SEAs were reviewed. For each SEA, a representative sample of sub-grantees for each grantee was selected using a random sampling method stratified by funding amount and type of sub-grantee (e.g., school, district or other type of agency). A sample of at least 10 sub-grantees was randomly selected in each state, with 10 percent of the sub-grantees randomly selected in states with more than 100 total sub-grantees. In states with fewer than 10 sub-grantees, all sub-grantees were

sampled⁶. As shown in Table 1, this sampling plan resulted in a total sample of 595 sub-grantees (about 20 percent of the sub-grantees in the selected states), representing 1,620 centers, or about 20 percent of the active centers in the selected states. While evaluation documents were requested for all sub-grantees in the sample, 126 sub-grantees (21 percent) did not submit any evaluation-related documents.

Table 1. Sub-Grantee Sample

	Total Sub-grantees	Total Sub-grantees in Sample	Total Active Centers	% of Active Centers	Sub-grantees Submitting documents	Sub-Grantees Not Submitting Documents
Base Year	1385	213	793	20%	177	36
OY1	560	215	287	23%	155	60
OY2	975	167	540	18%	126	41
TOTAL	2920	595	1620	20%	469	126

Document Review Process

As documents were received from SEAs and sub-grantees, BPA staff screened the materials to classify the types of documents received, and to determine whether the documents received were relevant for the purpose of this review task and whether any documents were missing. Similar documents were grouped into records. For example, if a sub-grantee evaluation report had two parts, it might include two documents, but would eventually be grouped into one record with one identification number assigned to it. A total of over 3,447 documents were received and screened over the three years of the review. About 29% (1,175) of the documents received were deemed inappropriate for review because they were not relevant to state or sub-grantee evaluations (e.g., outreach documents used to recruit program participants, or dissemination materials about the program that did not include evaluation results). A total of 2,272 documents were reviewed over the three years of the study, with the final database ultimately including 1,138 records. Detailed information about the types and numbers of documents coded is provided in Appendix A.

During the first year of this project, BPA developed a set of criteria for assessing state and sub-grantee evaluations based on evaluation research literature (GAO, 2009; McDavid & Hawthorn, 2006). In the second year, standards published by the Joint Committee on Standards for Educational Evaluation (2011) were used to refine the review criteria. These criteria include, among other elements, standards for research design, measures, analysis and use of evaluation findings. These criteria, as well as preliminary review of documents collected in 2010, inter-rater reliability data and feedback from analysts, were used to refine the codebook (Appendix E) for use in the review of the evaluation documents. Reviewers were instructed to code various aspects of the documents including: type of evaluator, measures used, research design used, description of sample, evidence of use, etc. The data entry system, developed in Microsoft Access, used both closed-ended variables (e.g., “Does the SEA require that specific performance measures are used in the sub-grantee evaluations?”); check boxes for specific items (e.g., “Evaluation measures include: math achievement using state standardized tests”); and fields for open-ended comments

⁶ See Appendix E for detailed sample information by year.

(e.g., “According to the document, how were evaluation results used?”). The instructions in Appendix F describe the procedures for coding and entering data and completing reliability analysis, and also provide some guidance to document reviewers.

Training the Document Review Team

BPA conducted several training meetings with the analysis team, beginning in January 2011. During these meetings, team leaders demonstrated coding and entering data into the Access database. Coders were instructed to work in pairs and consult their co-reviewers with coding questions, and also to consult the task leader and management team. The Access database was simplified and some slight revisions made based on lessons learned in previous years. Due to the team’s experience coding the documents in 2011, minimal changes were made to the coding process throughout 2012.

Methods for Synthesizing and Analyzing Data

To prepare this final report, the team collected, cleaned and analyzed both quantitative and qualitative types of data. For the quantitative analyses, the BPA team ran frequencies and cross-tabulations of key variables (e.g., type of evaluation design, process measures, outcome measures) in the document review database to describe the range of approaches to evaluation for states and the sample of sub-grantees. To examine how sub-grantee methods varied by selected SEA characteristics, cross-tabulations were created and Chi-Square tests of association were run.

In addition, notes from the interviews were systematically summarized using an Evaluation Interview Guide (Appendix D). These summaries were entered into a qualitative data analysis program (NVivo) for integration with qualitative document review data. Information entered into open-ended fields in the Access database was coded in NVivo so that all data could be synthesized to address the research questions as completely as possible. The analysis of qualitative text also allowed the team to extract examples of specific practices for this report.

The results of these analyses are presented in the next five chapters. Results include descriptions of the evaluation approaches (Chapter 3) taken by SEAs and sub-grantees (including research designs, types of evaluators used, and process and outcome measures); key findings from program evaluations (Chapter 4); use and dissemination of evaluation results (Chapter 5); guidance and technical assistance provided to sub-grantees (Chapter 6); and SEA evaluation needs and challenges (Chapter 7).

III. 21st CCLC Program Evaluation Approaches

This chapter describes the range of evaluation approaches undertaken by the SEAs and sub-grantees reviewed between 2010 and 2012. Various practices are described, including frequency of evaluation, types of evaluators, cost of evaluation, key evaluation questions, evaluation measures, research designs used, and how some SEA evaluation approaches have changed over time. Where appropriate, each section presents information at the SEA level and then at the sub-grantee level.

The chapter begins by presenting information about how many SEAs and sub-grantees conducted program evaluation. Table 2 presents the number of SEAs and sub-grantees that submitted completed evaluation reports to the BPA team over the three years of the study.

Table 2. Number and Proportion of SEAs and Sub-Grantees Submitting Evaluation Reports

Type of Entity	Number Submitting Reports	Percentage
SEAs (N=54)	48	89%
Sub-Grantees (N=595)	458	78%

Although 48 (out of 54) SEAs submitted evaluation reports, the BPA team gathered information through interviews suggesting that *all* SEAs were engaged in some sort of statewide evaluation activity. Some were in the process of developing evaluation designs, changing evaluation plans or evaluators, or having their reports written or revised, and therefore did not have reports available for submission at the time of data collection.

On the other hand, the team found that many sub-grantees did not conduct local program evaluations. While 78% of the sample submitted evaluation reports, these documents did not all represent formal evaluations (reports that assess program outcomes, implementation or both, supported by a systematic collection of objective data). Instead, some of the documents submitted included monitoring reports, data tables, self-assessments, peer observations and perceptions of program effectiveness based on anecdotal evidence.

Frequency of Evaluation

During interviews with SEA and sub-grantee directors and evaluators, the BPA team members asked how frequently program evaluations were conducted. In most cases, evaluations are conducted annually, although frequency often varied by capacity, resources and data availability.

Frequency of SEA Evaluation

The majority of SEAs conducted a statewide evaluation annually, with several states conducting evaluations every two or three years. At the time of data collection, two SEAs had not conducted complete evaluations since 2006, and two others had not completed an evaluation report since 2007. At least one SEA had only submitted one report for their 21st CCLC program that included PPICS data for the years 2006-2007, 2007-2008 and 2008-2009. A few others did not have

reports available at the time of data collection. The Virgin Islands had not yet conducted an evaluation of the 21st CCLC program at the time of BPA's document review. In California, a comprehensive evaluation report was being written at the time of document collection⁷; California published an evaluation of its high school afterschool program) After School Safety and Enrichment for Teens Program (ASSETS), which includes 21st CCLC programs, in 2007.

Frequency of Sub-Grantee Evaluations

Many sub-grantees conducted evaluations annually as part of an SEA requirement. In at least one SEA, sub-grantees were required to submit an annual evaluation *as well as* quarterly reports. A few states, such as Illinois, ensured that sub-grantees submitted their annual evaluation report as part of their continuation application. A few sub-grantees from the sample, such as those from New Mexico, reported submitting their evaluations bi-annually. All of the New Jersey sub-grantees that were interviewed submitted a mid-year report and a final report every year.

In some cases, the frequency with which sub-grantees completed evaluations varied within states. For example, two of Tennessee's sub-grantees submitted their reports annually, while another had not completed an evaluation report at all. One Indiana sub-grantee submitted both a mid- and end-year report, whereas other sub-grantees in Indiana completed annual reports only. A Puerto Rico sub-grantee had already conducted three evaluations in the last nine years, in contrast to others from the same SEA that had done one or none at all.

Some sub-grantees did not conduct formal evaluations, but instead used monitoring reports and informal evaluations to assess program quality and effectiveness. These SEAs collected data within their own sites/centers at least once a year using a variety of tools and assessments. One of the Virgin Islands sub-grantees examined the territory's standardized assessment (VITAL) results each year to assess individual student progress. Another Virgin Islands sub-grantee did not submit formal evaluation reports, but instead did pre/post testing for reading, literacy and math using a software package. A Virginia sub-grantee reported anecdotally on student progress on their continuation application, but did not use systematic objective data for measuring results. One of Tennessee's sub-grantees did not submit a formal evaluation report; however, students were assessed for academic performance three times a year, and the program also evaluated the performance of its tutors and coordinators.

Types of Evaluators

One of the questions guiding this review asks about the type of evaluators programs used for their 21st CCLC evaluations. A key characteristic of a comprehensive and high-quality evaluation is utilizing the services and expertise of a qualified and independent evaluator. The BPA team explored the types of evaluators SEAs and sub-grantees used over the last three years.

Evaluators of SEA Programs

Evaluators of SEA 21st CCLC programs included state or SEA personnel, academic institutions, independent consultants and research agencies. As shown in Table 3, SEAs were most likely to contract with research agencies to conduct evaluations. These independent agencies were external to the SEA. Examples include American Institutes for Research (AIR), formerly Learning Point Associates (LPA), and RMC Research Corporation. The next most common

⁷ More recent California statewide evaluation reports (2012) for both the elementary and middle school programs (Afterschool Education and Safety/ASES) and the high school programs (ASSETS) were published and are posted on the California Department of Education website after BPA collected the documents for review.

category of evaluator was academic institutions (12). These evaluators included professors or research centers based at universities. For example, California contracted with the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), based at University of California, Los Angeles.

Table 3. Evaluators of 21st CCLC SEA Grants

Type of Evaluator Used	N	%
Program staff	0	0%
State personnel	7	13%
Academic Institution	12	22%
Independent Consultant	6	11%
Research Agency	19	35%
Missing	10	19%
TOTAL	54	100%

In seven states, state research personnel conduct the state-level grant evaluations. These evaluators were internal to the state, but since they were external to the 21st CCLC program, they could serve as independent evaluators and conduct evaluation with objectivity. In West Virginia, staff from the State’s Office of School Research provided independent evaluation, research and analysis for the WV Department of Education. These researchers were able to remain independent as they are housed outside the office that operates the 21st CCLC program and are prohibited from making policy decisions that impact the program. Six SEAs hired independent consultants as evaluators. For 10 SEAs, evaluator information was missing or unknown. These include SEAs that were not currently conducting evaluations and SEAs that were in the midst of selecting an evaluator. None of the SEAs had their own program staff conduct the statewide program evaluations.

SEA Evaluator Selection Methods and Challenges

Most SEAs hired an external evaluator through a competitive bid process. For example, California released an RFP for an evaluator after state Proposition 49 passed and greatly expanded funding for afterschool programs, enabling the SEA to seek an external evaluator. Texas also used an RFP process, but unlike the other states, required the highest scored applicants to give an oral presentation before being awarded a grant. A few states, such as Minnesota and Oregon, were working towards releasing an RFP for an external evaluator at the time of their monitoring visits.

Other states had non-competitive selection processes, where evaluators were selected based on a recommendation or preference, or by using an independent research office within the SEA. Arkansas chose their contractor, MGT, through sole-source procurement. Delaware sought RMC Research Corporation as the statewide evaluator; since the contract was for less than \$50,000, the SEA did not have to go through a formal bidding process. A few other SEAs, like Georgia, obtained recommendations for external evaluators from other governmental agencies.

Many states reported having a good relationship with their evaluators; however, a few states reported challenges with their evaluators as well. For example, Arkansas found that having

separate contracts with one agency for data collection services and another for their quality assessment and continuous improvement process was not only expensive, but also meant that their evaluation did not combine the student outcome and process data, and was not giving them useful results. The SEA noted that they intend to combine the two efforts into one contract and work more closely with David P. Weikart Center, which administers the Youth Program Quality Assessment (YPQA), to incorporate student academic outcome measures during site observations. In another case, Oklahoma staff reported that their evaluator did not provide enough information about program quality across the state and that they planned to work towards selecting a new evaluator. West Virginia reported working with an external evaluator that had increasingly higher costs while not providing very useful information, which prompted the SEA to ask the state's Office of Research to conduct the evaluation.

SEA Evaluator Roles/Responsibilities

SEAs consistently established clear and specific responsibilities for their evaluators in scopes of work or other documents. Some SEAs, like Florida and BIE, had a very broad scope of responsibilities for the evaluator, including monitoring and technical assistance in addition to evaluation services. In the case of West Virginia, however, the independent evaluator was the WVDE's Office of Research, and the SEA 21st CCLC staff provided technical assistance and conducted monitoring. Other states approached external evaluation services differently, by contracting specific portions of the evaluation to certain agencies. For example, Oklahoma contracted with David P. Weikart Center for the YPQA and West Virginia contracted with 21st Century Grant Services for tasks specific to their Continuous Improvement Process for After School (CIPAS). Arkansas contracted with an organization purely for data management and analysis services. Like Arkansas, quite a number of states also sought primarily data analysis support, but were deliberate in seeking support from American Institutes of Research (AIR) due to their direct knowledge of the national PPICS data system.

Evaluators of Sub-Grantee Programs

Like SEAs, sub-grantees most commonly hired research agencies (24%) to conduct evaluations.. Sub-grantees also commonly hired program staff (22%) as evaluators, unlike SEAs, none of which used program staff. Program staff members serving as sub-grantee evaluators were usually program directors or a teacher or other staff member who worked directly on the grant program. The next most common types of sub-grantee evaluators were independent consultants (18%) and academic institutions (11%). A few sub-grantees used state or district personnel (5%); these include district evaluators or offices within the state or district that conducted research and evaluation. A very small proportion of programs used an evaluator that does not fall into the above categories (1% "other"). In some of these cases, grantees reported that a parent or other volunteer served as an informal evaluator. For approximately 19% of sub-grantees, the evaluation documents did not identify a specific evaluator, or there was not enough information provided to determine evaluator type.

Table 4. Evaluators of 21st CCLC Sub-Grantee Programs

Type of Evaluator Used	N	%
Program staff	99	22%
State/ district personnel	24	5%
Academic Institution	52	11%
Independent Consultant	81	18%
Research Agency	110	24%
Other	6	1%
Missing	86	19%
TOTAL	458	100%

SEA Guidance to Sub-Grantees on Selection of Evaluators

Some SEAs provided requirements regarding the types of evaluators sub-grantees should hire and the amount of funds that should be spent on local program evaluations. For example, California, Washington, D.C., and Georgia, required that sub-grantees hire an evaluator external to the program. Pennsylvania encouraged but did not require their sub-grantees to hire local, external evaluators. And Arkansas assigned evaluators to their sub-grantees (one company manages PPICS data and the Weikart Center operates the YPQA).

Several SEAs established a limit to the amount of funds sub-grantees could spend on evaluation, such as Texas, which required sub-grantees to spend \$3,500 to \$5,000 per site on evaluation. Other SEAs prohibited sub-grantees from expending more than a certain percentage of their grants on program evaluation, ranging from 5 to 6% of the total grant on external evaluation services.

At least 10 SEAs offered technical assistance to sub-grantees for selecting an evaluator. For example, Maryland and New Jersey offered some information about selecting an evaluator in their RFPs, including recommended responsibilities of the local evaluator. Michigan and Texas created evaluation guides that discussed external evaluator cost guidelines and evaluator responsibilities, and included a template for an agreement between the sub-grantee and the chosen evaluator. Florida and Pennsylvania discussed evaluator selection during trainings. Florida SEA staff discussed how to work with an external evaluator and what to look for in evaluators during the bidders' conference, and Pennsylvania held a training through its statewide evaluator (Allegheny Intermediate Unit) on what to look for and how to select an evaluator.

Sub-Grantee Evaluator Selection Methods and Challenges

Sub-grantees used a range of processes to select different types of evaluators, but repeatedly reported challenges with selecting a quality evaluator. One of the more commonly reported methods sub-grantees used for obtaining an evaluator was through a solicitation process. Other sub-grantees chose an evaluator based on internal recommendations; in a few cases, the superintendent of a school district recommended an evaluator, or friends or acquaintances of program staff served as evaluators. SEAs such as Delaware, New Mexico and North Dakota simply assigned an external evaluator to the sub-grantees; in these cases, the assigned evaluator often conducted continuous or quality improvement assessments, rather than more formal program evaluation services. Multiple sub-grantees from one state indicated that the SEA

referred the evaluator to them, while other SEAs (New Jersey and Florida) specifically reported that they are prohibited from recommending evaluators to sub-grantees.

Sub-grantees selected evaluators on many criteria, such as a background with out-of-school programs or evaluation, experience working with the sub-grantees through other programs, familiarity with the local region and/or state, evaluation or research reputation, and affordability. In Alaska, sub-grantees felt it was important to hire evaluators that were local or had extensive work experience in the state; they reported that conducting evaluation through the lens of regional needs or familiarity with local rules or regulations can benefit the program. Community-based organizations in other states chose their evaluators based on the organizations' reputations as premiere researchers and evaluators.

Sub-Grantee Evaluator Roles/Responsibilities

Sub-grantee evaluators' roles varied, ranging from data collection, comprehensive evaluation, preparing reports and evaluation monitoring, to technical assistance or guidance related to evaluation. At the sub-grantee level, the role of evaluators seemed to encompass a broader scope compared to the SEA level evaluators. In the case of a Georgia sub-grantee, the evaluator's scope of work included developing the evaluation plan, designing the data collection instruments (e.g., interview protocol), collecting data (e.g., school records, observations, site visits), analyzing data, completing reports and other State-required documents, and providing recommendations on how to improve the program. While these are common and expected responsibilities of evaluators, other sub-grantees assigned additional responsibilities to evaluators to meet SEA evaluation requirements. Sub-grantees in Washington, for example, included in the responsibilities that the evaluator assist the sub-grantee in meeting the requirements of the grant, including having the evaluator upload PPICS data by the specified deadlines. In the case of states with state-assigned evaluators conducting quality or continuous improvement processes, the evaluators' roles primarily or exclusively entailed conducting monitoring-related activities such as visits, technical assistance or reporting. Several sub-grantees had evaluators provide trainings or technical assistance to program staff in preparation for the evaluation, or based on sub-grantee needs as identified from the evaluation results.

Evaluation Cost

A comprehensive and rigorous evaluation can be a high cost endeavor, and 21st CCLC programs are limited in the amount of grant funds available for evaluation. Because grant award amounts varied across SEAs and sub-grantees, the funds spent on evaluation also varied widely. This section presents a summary of information gathered about evaluation costs for SEAs and sub-grantees.

Costs of SEA Evaluations

State level data reveal that most, but not all of SEAs supported evaluations of their state 21st CCLC programs using funds from the allocated 3% allotment (for monitoring, technical assistance and evaluation) of the grant. Actual amounts varied widely, as did cost structures. Some states held annual contracts with evaluation firms and others had multi-year contracts. In some cases, SEAs used in-house personnel as evaluators and the 21st CCLC grant funded a portion (or all) of their time. For a couple of states (West Virginia and Montana), no grant funds were spent on evaluation. In the case of WV, the SEA's Office of Research conducted the evaluation and in Montana, a staff member generated a PPICS reports.

Actual costs of evaluation ranged from over \$1 million in large states like California and Florida, to less than \$10,000 in smaller, less populated states. California spent \$1 million for a four-year contract with its external evaluator, a university-based research center, and Florida spent \$1.6 million per year (however, this included technical assistance and monitoring in addition to evaluation). New York spent \$162,792 per year, Texas spent \$699,000 over two years, and Puerto Rico spent over \$600,000. In contrast, Colorado spent \$9,300 on its statewide evaluation, Hawaii spent \$11,000, and Missouri spent \$11,500. About 10 SEAs spent between \$50,000 and \$100,000 on evaluation per year; these included small and medium-sized states (Table 5).

In several cases, it was difficult to determine precisely how much SEAs spent on evaluation, because some SEAs had contracts with agencies that included evaluation, monitoring, and technical assistance (TA). The University of Florida received a roughly \$1.6 million contract to conduct evaluation, monitoring and TA services for the Florida 21st CCLC program. This contract was the full amount of the 3% state activities fund. BIE also held a contract for \$240,000 with an agency that provided evaluation, monitoring and TA services.

Because the cost of evaluation is difficult to report accurately, as illustrated in the examples above, several figures in Table 5 represent estimates of evaluation costs. As the table shows, only a handful of SEAs had sufficient resources for evaluation available to support a large-scale, rigorous statewide comprehensive evaluation. Many SEAs spent a significant amount of program funds on evaluation, and with a well-structured design, should be able to produce evaluation information that can help support program improvement. Those that spent less than \$100,000 a year may need to consider spending as much as one-half of their 3% funds on program evaluation and/or conducting evaluations every two to three years in order to be able to afford a meaningful evaluation effort.

Table 5. Costs of SEA Evaluations

Cost of Evaluation	# of SEAs
\$ 250,000 or more	5
\$100,000 to \$250,00	11
\$50,000 - \$100,000	10
Less than \$50,000	16
<i>Missing</i>	12
TOTAL	54

While some of the cost information collected from SEAs was very general (e.g., total contract amounts with little detail), several states provided evidence of specific cost breakdowns. This information provides a clearer picture of how evaluation funds were spent. For example, in Maine the total \$23,500 cost estimate is broken down in the following way: Client Meetings (\$3,200), Data Analysis (\$7,200), Data Collection (\$3,500), Report Writing (\$5,000), Presentations (\$3,600), and Travel, Materials and Phone Calls (\$1,000). Arkansas spent \$24,900 for MGT of America to provide data management and collection services, and \$175,000 per year for YPQA services paid to Arkansas State University-Jonesboro, which holds the state contract with the Weikart Center to provide YPQA services. In New Mexico, the cost of the most recent evaluation contract was \$50,000. This included only \$12,500 to develop an evaluation plan,

analyze data, and prepare both an annual and a four-year evaluation report. The remaining funds were for researching database vendors for federal APR reporting and improving the quality of PPICS data through TA and monitoring of the data.

Cost of Sub-Grantee Evaluations

The data reflecting how much sub-grantees spent on evaluation is much less clear than the SEA-level cost data. Some states provided cost guidelines or requirements for sub-grantees about how much to spend on evaluation services. For example, Alaska permitted 5% of grants to be used for evaluation, and in Pennsylvania, the SEA encouraged sub-grantees to use 5 to 8% of grants for evaluation. In West Virginia, the SEA suggested that sub-grantees set aside \$1,500 for the CIPAS (continuous improvement process) activities. In California, one large sub-grantee used 6% of their grant on a contract with ERC, an external evaluator. At the higher end, a Maryland sub-grantee used 15% of their budget on evaluation. This high proportion of grant funds used on evaluation was rare. Most sub-grantee evaluations cost between \$5,000 and \$15,000, with a small number of sub-grantee evaluations costing closer to \$20,000.

Overall, sub-grantees did not spend a great deal of funds on evaluation, and they usually had little to no direction from their SEA on what to spend. Most sub-grantees reported that their grant funds were limited and they were already trying to stretch those dollars to provide services to students and families.

Many sub-grantees in the study sample revealed that they did not spend any specific funds on evaluation (note that Table 4 shows 22% of sub-grantee evaluations were conducted by program staff), but that program directors and staff conducted the evaluation (usually informally) as a part of their roles on the 21st CCLC program.

Challenges Related to Cost

Some states reported cost challenges. In most of these cases, SEAs did not feel that the evaluation services were worth the large amount of funds being spent. For example, the Kansas SEA reportedly thought their evaluator (Kansas University) charged too much for their services. In California, SEA personnel reported that a huge issue was the cost related to making sure the data submitted by sub-grantees for state evaluations was of high quality before analyzing it. The California Department of Education (CDE) expressed the desire to develop an in-house online system of data collection, believing it would be a more cost-effective strategy for evaluation. Another reason to develop CDE capacity was that CDE data submission is required by school districts, whereas CRESST (the external evaluator) data collection is voluntary among participants (sub-grantees).

Key Evaluation Questions

An important component of a strong comprehensive evaluation is the articulation of key evaluation or research questions. The BPA team reviewed documents to determine the extent to which programs articulated specific evaluation questions. This section describes the types of evaluation questions or objectives presented in the evaluation reports, and includes questions related to both outcomes and program implementation.

Key Questions in SEA Evaluations

The specificity and focus of research and evaluation questions varied across the SEA-level evaluations. Evaluation questions inquired about the impact of 21st CCLC programs on student achievement and other student outcomes, program implementation, and the relationship between

aspects of implementation and outcomes. These questions ranged in level of specificity from rather broad (e.g., Did the program do what it said it would do? Did the project accomplish what it intended?) to very detailed (e.g., To what degree is recreational and wellness programming explicitly focused on improving health?). Some evaluations did not articulate specific questions, but explained the goals or objectives of the evaluation. As is the case with research questions, the evaluation goals ranged in their level of specificity from broad (e.g., to provide descriptive information regarding programs) to detailed (to look at data gathered and determine whether benchmarks set for the next three years for Vermont are reasonable, achievable, and push Vermont to continue to grow at the same pace that the last three years have seen).

Outcome Questions

Questions about academic outcomes focused on student achievement generally or performance in specific subject areas. Examples of this first type of research question include:

- “Are centers impacting student achievement?”
- “Does participation in after school programs appear to contribute to improved academic achievement?”

Examples of research questions focused on academic performance in specific subject areas include:

- “Do participants of [Louisiana’s] afterschool programs demonstrate improved academic achievement in reading and math, when compared to non-participants?”
- “To what extent is there evidence that students participating in services and activities funded by 21st CCLCs demonstrated better performance on state assessments in reading and mathematics, compared to similar students not participating in the program?”

Questions also focused on non-academic outcomes, including student behavior, school attendance, and social and/or emotional development. Examples include:

- “Does participation in afterschool programs affect other behaviors such as: school day attendance, homework completion, positive behavior, skill development, and healthy youth development?”
- “Did Illinois meet Goal 3 - Schools will see an increase in the social-emotional skills of their students?”

Implementation Questions

Evaluation questions inquiring about implementation centered around several aspects of programming, including program characteristics, participant characteristics, program participation, services or programming offered, program goals and objectives, program sustainability, compliance with their RFP, partnerships, parent involvement, stakeholder satisfaction, staffing, staff qualifications and professional development. Examples of research questions about specific aspects of program implementation include:

- “What are the key characteristics of [the] afterschool program and participants?”
- “What is nature of the [Virginia] 21st CCLC grant program and level of participation by students?”
- “Are 21st CCLCs using evidence-based programs that target social and emotional

development?”

- “Did Illinois meet Goal 7 - Programs will collaborate with schools and community based organizations to provide sustainable programs?”
- “Are key stakeholders satisfied with program performance?”
- “How were the centers staffed?”

High Quality Questions

Some state evaluations also included research questions that sought to identify if there were aspects of program implementation that influenced student outcomes. These questions were not limited to investigating whether the program is working or how the program was implemented, but inquired about reasons the program may be producing particular results. These were designated in our review as “high quality” questions. Examples include:

- “Were there specific features or characteristics of 21st CCLC programs that are associated with sustainability and exemplary outcomes over the long term? Did student participation beyond required minimum days (dosage) affect behavior changes, academic achievement, and/or school involvement?” (Illinois)
- “To what extent is there evidence of a relationship between select program and student characteristics and the likelihood that students demonstrated: (a) higher academic achievement in reading/language arts and mathematics; and (b) an improvement in behaviors likely to be supportive of better academic achievement?” (New Jersey)
- “To what extent was there evidence of a relationship between higher levels of attendance in 21st CCLC programming and the achievement of desired academic and behavior outcomes? To what extent was there evidence that particular center and student characteristics and attributes were associated with student academic and behavioral improvement?” (Washington)

Evaluation Goals

Not all evaluations articulated research questions, but instead explained the goals or purpose of the evaluation. The types of goals varied, but several evaluations specified that one of the objectives of the evaluation was to inform program improvement. Examples of goals related to program improvement include:

- “Provide process data that will assist project staff in continually improving the quality of services to the children and their families.” (Nebraska)
- “Present information, based on scientific principles using defensible data and methods that will improve the outcomes of the program in South Carolina and elsewhere.”
- “To provide formative evaluation information and direction to program managers, teachers and other practitioners administering Florida's 21st CCLC programs and to determine needs for technical assistance and trainings.”

Key Questions in Sub-Grantee Evaluations

The data indicate that most sub-grantee evaluations did not articulate research questions. Among those that did, questions included both outcome and implementation questions. Similar to the

research questions in statewide evaluation reports, outcome questions inquired about academic performance on math and reading assessments and non-student outcomes such as student behavior, discipline and attendance. Implementation questions inquired about a wide range of topics as detailed below. Although sub-grantees in only a small group of states specified research questions in their evaluations, several sub-grantee evaluations clearly stated the purpose or objective of their evaluations.

Outcome Questions

Examples of academic outcome-focused questions include:

- “Did Camp Zion students improve their math and reading scores as measured by locally developed tests?”
- “To what extent did regular program participants make progress on reading and math short-term assessment measures (e.g., DIBELS, Star Reading and Math, Acuity)?”
- “What was the effect of attending BRIDGES on students’ academic progress?”
- “What do second year math and reading grades and test scores reveal about 2nd year participation?”
- “To what extent are participants scoring proficiently in each of the three science domains?”

A few questions focused on sub-group analysis:

- “Do students participating in the Heads Up program attain higher attendance rates and higher scores on standardized assessments than comparable students who did not participate in the program, and how do those results compare to the previous school year?”
- “What proportion of frequent participants during the previous (2009-10) school year improved their [state standardized tests] performance category scores in reading and math from fall 2009 to fall 2010?”, and “How did those proportions compare to district non-frequent participants in the same grade levels?”

Examples of non-academic outcome questions include:

- “Did Camp Zion students increase their knowledge of enrichment activities such as art, recreational activities, dancing, and sports as measured by locally developed tests and/or student surveys?”
- “Do parents perceive the afterschool program to have a positive impact on student behavior?”
- To what extent are regular attending students being suspended from school for disruptive or violent behaviors?
- “Did the percentage of students expressing positive attitudes towards school and schoolwork increase by 10%?”

Implementation Questions

Research questions about implementation inquired about a variety of topics including, but not limited to, satisfaction with the program, program activities, program participation and program quality. Examples include:

- “What are school administrators' perceptions of [afterschool] programs, and how have they changed over the last five years? Do students perceive the afterschool program to have been a positive experience?”
- “Were students provided a balanced curriculum that included required academic activities and sufficient youth development enrichment activities?”
- “What are the characteristics of students participating in the Heads Up program, and how do they differ from other students attending the same schools?”
- “Were students provided a safe, supervised, and structured program to attend? What are the students', parents', and staff members' perceptions related to key dimensions of program quality and areas of improvement?”
- “Which community based partnerships as planned in the grant application have been established and maintained and which ones have not? Did the center serve children and community members with the greatest need for expanded learning opportunities in 2009-10?”

Evaluation Objectives

Several sub-grantee evaluations did not articulate specific research questions and instead stated the overall purpose or the objectives of the evaluation. Examples of these objectives include: to examine the extent to which the goals of the grant have been met and the extent to which the activities outlined in the grant have been met; to identify current strengths of the program, current program operations, areas for improvement, areas for corrective action and needs for technical assistance; and, to determine the effectiveness of program implementation and determine the impact of program activities on student academic performance and graduation completion.

Evaluation Measures Used by Grantees

The BPA team gathered data about measures used to evaluate 21st CCLC programs through interviews and document review of evaluation reports. Evaluation measures fall into four major categories:

- Student academic outcome measures;
- Student non-academic outcome measures;
- Parent/family outcome measures; and
- Process measures.

Information was gathered about the content of the measure and the source of data used. For example, math achievement was measured using a variety of sources including state assessments, school grades and teacher surveys. Below, measures are described first for state-level evaluations and then for sub-grantee level evaluations.

SEA Evaluation Measures

Table 6 displays a summary of SEA evaluation measures and the most common data sources used. Complete information about measures and data sources is provided in Appendix G.

Table 6. Types of Measures Used in State Evaluation Reports

Student Academic Outcome Measures	Number of states using this measure	Most Common Data Source
Math Achievement	38	State Standardized Test
Reading/ELA Achievement	38	State Standardized Test
Other Measure - Academic	15	Teacher Survey
Other Content Area Achievement	8	School/Classroom Grades, Other or Unspecified Source
Student Non-Academic Outcome Measures		
Behavior	38	Teacher Survey
School/Classroom Attendance	28	Teacher Survey
Homework Completion	24	Teacher Survey
Satisfaction/Attitude Toward School	15	Teacher Survey
Other Measure – Non-Academic	15	Teacher Survey
Disciplinary Incidents	8	Other or Unspecified Source
Parent/Family Outcomes		
Other Measure - Family	7	Parent Survey
Parent Involvement in School/Classroom	4	Other or Unspecified Source
Satisfaction with Child's School	3	Parent Survey
Parent Employment	1	Other or Unspecified Source
Process Measures		
Program Attendance	40	Program Records/PPICS
Recreational and Enrichment Activities	36	Program Records/PPICS
Core Academic Activities	35	Program Records/PPICS
Community Partnerships	29	Program Records/PPICS, Other or Unspecified Source
Service Hours Provided	27	Program Records/PPICS
Services to Adults	23	Program Records/PPICS
Parent Involvement in 21 st CCLC Program	21	Parent Survey, Other or Unspecified Source
Other Measure - Process	18	Other or Unspecified Source
Links to School Day	15	Other or Unspecified Source
Parent Satisfaction with 21 st CCLC Program	13	Parent Survey
Counseling and Mentoring to Students	13	Program Records/PPICS
Program Implementation Issues	12	YPQA ⁸ , Other or Unspecified
Communications with Parents	12	Other or Unspecified Source
Student Satisfaction with 21 st CCLC Program	11	Student Survey

⁸ Highscope Youth Program Quality Assessment

Student Academic Outcome Measures

As Table 6 shows, the most common types of student academic outcome measures used for SEA evaluations were math and reading achievement, assessed using state standardized test scores. Twenty-one SEAs used standardized state assessments to measure math and reading. Examples of these assessments include Northwest Evaluation Association Measures of Academic Progress (used by SEAs such as BIE and Massachusetts), the Virginia Standards of Learning, the Colorado State Assessment Program (CSAP) and the Arizona Instrument to Measure Standards (AIMS).

Other common measures of academic performance include grades and teacher reports of academic performance or academic improvement. The majority of SEAs that used a teacher survey used the PPICS Teacher Survey developed by LPA/AIR which has an item asking teachers to rate students' academic performance. However, not all states used this survey. For example, Massachusetts used the Survey of Afterschool Youth Outcomes (SAYO) for teachers, which had items asking teachers to rate students' reading, verbal communication, written communication, math communication, math reasoning, math problem-solving, science and social studies outcomes. "Other" sources of data used to measure student academic outcomes include local reading and mathematics assessments (e.g., Dynamic Indicators of Basic Early Literacy Skills (DIBELS), Acuity); SATs and PSAT; Palmetto Assessment of State Standards Scores for math, English, science, social studies and writing; parent and student surveys assessing student academic performance; and individualized education plans (IEPs).

Non-Academic Outcome Measures

In addition to academic achievement, SEAs used non-academic measures. The most common were behavior (38 SEAs), school or classroom attendance (28), and homework completion (24). The most common data source was a teacher assessment. Many SEAs used the teacher survey developed by LPA/AIR, or a modified version of it. The survey includes measures of homework completion and quality, class participation, volunteering, class attendance, attentiveness in class, class behavior, motivation and getting along with other students. Other sources of data included school records, student and parent surveys, and PPICS.

SEAs also collected data to assess students' social and emotional indicators, student safety and behavior. These measures included disciplinary incidents such as suspensions, expulsions, criminal and non-criminal referrals, and instances of student violence. Other indicators used included drop-out rates, graduation rates, parent and student ratings of student behavior, students' involvement in school activities, students' involvement in extracurricular activities, students' communication skills and relationships with adults.

Parent/Family Outcomes

Very few SEAs included parent or family outcome measures as part of their statewide evaluations. Such measures included parents learning new skills as a result of participation at the CCLC, family and civic engagement, parent employment and GED completion, attainment of computer skills, and attainment/increase of parenting skills. Most of these data were gathered through parent surveys.

Process Measures

According to the evaluation review data, SEAs used a variety of measures to assess program implementation, quality and process. The most commonly used measures include student

program attendance, types of activities (academic and recreational), community partnerships, service hours provided, services to adults and parent involvement (Table 6). All of these were most often measured using PPICS or program tracking data, except for parent involvement, which was most commonly assessed using a parent survey (or other type of parent assessment, such as focus groups).

Participant attendance and student characteristics were measured by collecting data from student attendance sheets, records reflecting the proportion of students receiving free or reduced lunch, grade levels served and other student demographic data. Stakeholder and participants' satisfaction with programs were measured by surveys with multiple stakeholders, including parents, staff, school liaisons, community stakeholders and students. Other constructs measured in examining program implementation included staff engagement and interaction with students, students' feelings of safety (parent and student surveys), program climate and instructional quality.

In several SEAs, program implementation and challenges were measured using Program Director Surveys that addressed accomplishments, barriers or challenges, strengths, issues/concerns, lessons learned and training needs; focus groups to collect information on program implementation and challenges; grantee reports on implementation challenges and issues; and various stakeholder surveys.

Formal Quality Assessment and Continuous Improvement Processes

Some SEAs have put formal continuous improvement processes into place using quality assessment tools such as the YPQA, which focuses on safe and engaging learning environments, or the Assessment of Afterschool Program Practices Tool (APT-O), a measure used to assess program implementation relative to key quality constructs. These tools are designed to support a continuous quality improvement process (QIP) by identifying areas for improvement and providing a structure for improvement planning and follow-up. The SEA makes an investment in the tools themselves, training, and in some cases, for staff or consultants to conduct the assessments or produce reports. However, the assessment process and its use in continuous improvement primarily takes place at the sub-grantee level.

The roles and responsibilities of SEAs and sub-grantees in the formal QIP vary across states. SEA staff may simply require sub-grantees to conduct a self-assessment, contract with a TA provider to support sub-grantees with their self-assessments, review self-assessment results, or may conduct part or all of the QIP. One example of the first scenario is Minnesota, which requires sub-grantees to choose from a list of approved self-assessment tools (e.g., YPQA, NYSAN-QSA⁹). Sub-grantees must indicate which tool they will use in their applications and report on their program quality improvement efforts, but do not submit actual QIP results to the SEA. Similarly, in Massachusetts, the SEA requires sub-grantees to conduct an annual self-assessment using the APT-O and to report on how the results informed program improvement efforts in their continuation applications. In New York, the SEA requires sub-grantees to complete a self-assessment, yet encourages local evaluators to facilitate the assessment to offset any tendencies for programs to rate themselves too generously. In addition, some states contract with a TA provider to coordinate sub-grantee self-assessments. For example, both Arkansas and Michigan contract with the Weikart Center to provide assistance to sub-grantees in administering

⁹ New York State Afterschool Network Quality Assessment Tool

the YPQA. In both cases, sub-grantees complete a self-assessment, and both sub-grantee and SEA staff use the results to inform program improvement.

In other states, SEA staff review self-assessment data or uses a QIP in conjunction with their monitoring processes, but do not conduct the QIP themselves. For example, in West Virginia sub-grantees complete the self-assessment tool and upload it virtually. The SEA contractor then reviews the documentary evidence and self-assessment and produces an official assessment of the program. In Maryland sub-grantees complete a self-assessment tool and provide evidence and documentation to support their ratings. During monitoring visits, the SEA reviews the ratings with sub-grantee staff, makes observations, reviews supportive documents, and determines areas of strength and areas of improvement. Similarly, in Wisconsin, sub-grantees choose from one of four approved tools and complete a self-assessment before their monitoring visits from the SEA. The SEA then reviews the information from the self-assessment summary and on-site monitoring interviews, and uses the information to determine if there are any findings.

Although many states with formal QIPs rely on self-assessment data, several states also rely on SEA staff or SEA contractors' assessments of program quality. For example, in Kentucky the SEA requires sub-grantees to self-administer the YPQA, and the SEA evaluator (Center for Evaluation and Education Policy) administers the YPQA when sub-grantees are in their third year of programming as a check on fidelity and continuous improvement. In Florida, the SEA contracts with the University of Florida to do TA, monitoring and evaluation. The contractor uses a QIP tool during monitoring visits with sub-grantees. These visits result in corrective action, follow-up and/or individual TA if needed. In Nebraska, the state evaluator designed the Continuous Improvement Process model and works with local evaluators who support sub-grantees with data collection, data review and setting goals for program improvement. The state evaluator facilitates the CIP, but a local management team, comprised of the project director, site coordinator, school principal, local evaluator and partners, conducts and reviews the progress of the CIP.

Among states with a formal QIP, the relationship between the QIP and program evaluation can differ from state to state. SEAs may use QIP data to evaluate implementation while using outcome data or PPICS data to evaluate performance, have a separate implementation evaluation that does not incorporate QIP data, or may consider the formal QIP as evaluation at the sub-grantee or SEA level. In Wisconsin, the SEA requires sub-grantees to complete the YPQA, but the QIP results are not incorporated into the statewide evaluation or sub-grantee evaluations. The state-level evaluation primarily relies on PPICS data, and the evidence indicates that the SEA does not gather information beyond what is included in the QIP or PPICS data to evaluate implementation. In Maryland, the SEA included both process measures and outcome measures in its statewide evaluation, but did not incorporate QIP results in its evaluation. To assess implementation, the state evaluator administered program satisfaction surveys with multiple stakeholders, gathered focus group and interview data on program implementation challenges, and examined PPICS data on participant demographics, program attendance and program activities. In Colorado, the QIP tool (Monitoring and Quality Improvement Tool) includes both process and outcome measures. The SEA requires that all sub-grantees complete the QIP and views the process as meeting the requirement for sub-grantee program assessment. Similarly, in Nebraska, the SEA coordinates evaluation at the sub-grantee level through the QIP.

Sub-Grantee Evaluation Measures

Table 7 displays a summary of the most commonly used sub-grantee evaluation measures, including common data sources. Detailed information is provided in Appendix G. Examples of each type of measure are described below.

Table 7. Types of Measures Used in Sub-grantee Evaluation Reports

Student Academic Outcome Measures	Percent Sub-grantees using this measure	Most Common Data Source
Math Achievement	62	State Standardized Test
Reading/ELA Achievement	62	State Standardized Test
Other Measure - Academic	19	Other or Unspecified Source
Other Content Area Achievement	9	State Standardized Test, School/Classroom Grades
Student Non-Academic Outcome Measures		
Behavior	52	Teacher Survey
School/Classroom Attendance	47	Teacher Survey
Homework Completion	33	Teacher Survey
Satisfaction/Attitude Toward School	30	Teacher Survey
Disciplinary Incidents	21	School Records
Other Measure – Non-Academic	19	Teacher Survey
Parent/Family Outcomes		
Other Measure - Family	10	Other or Unspecified Source
Parent Involvement in School/Classroom	9	Other or Unspecified Source
Satisfaction with Child's School	6	Parent Survey
Process Measures		
Program Attendance	54	Program Records/PPICS
Recreational and Enrichment Activities	47	Program Records/PPICS
Core Academic Activities	45	Program Records/PPICS
Community Partnerships	38	Program Records/PPICS, Other or Unspecified Source
Parent Involvement in 21 st CCLC Program	33	Parent Survey, Program Records/PPICS
Parent Satisfaction with 21 st CCLC Program	31	Parent Survey
Student Satisfaction with 21 st CCLC Program	30	Student Survey
Program Implementation Issues	30	YPQA ¹⁰ , Other or Unspecified
Services to Adults	29	Program Records/PPICS
Service Hours Provided	28	Program Records/PPICS
Communications with Parents	26	Parent Survey
Links to School Day	19	Other or Unspecified Source
Other Measure - Process	17	Other or Unspecified Source
Adult-to-Student Ratio	16	Program Records/PPICS
Teacher/Administrator Satisfaction with 21 st CCLC Program	13	Teacher Survey
Counseling and Mentoring to Students	11	Program Records/PPICS

¹⁰ Highscope Youth Program Quality Assessment

Student Academic Outcome Measures

Of the sub-grantees that submitted evaluation reports, 62% assessed math and reading achievement as part of their evaluation. State standardized tests were the most common source for this measure (43% for math; 45% for reading), followed by school or classroom grades (34% for math; 35% for reading). Examples of standardized achievement assessments include TerraNova (used in states such as Arizona and Indiana); Northwest Evaluation Association Measure of Academic Progress; Dynamic Indicators of Basic Early Literacy Skill (DIBELS); Measures of Academic Progress, a state-aligned computerized adaptive assessment instrument used in Kentucky; Mississippi Curriculum Test; and the Michigan Education Assessment Program.

Sub-grantee evaluations also relied on teacher, parent, student and program staff ratings of academic performance. For example, some sub-grantee evaluations presented aggregated data from teacher surveys that included one or more items asking teachers to rate students' overall academic performance or academic improvement.

Non-Academic Outcome Measures

The most common measures of non-academic outcomes used by sub-grantees were behavior, school attendance and homework completion. To assess these areas, many sub-grantees used a teacher survey of academically related behaviors and attitudes developed by LPA/AIR, or a modified version of it. The survey has measures of homework completion and quality, class participation, volunteering, attentiveness in class, motivation and getting along with other students. Some sub-grantees used parent or student surveys to measure similar behaviors and attitudes. Examples of these measures include parent survey items concerning students' feelings and attitudes towards school and school attendance, and a student self-assessment of homework completion. Examples of other measures of student behavior include daily tutor logs on behavior, a student connectedness survey and stakeholder perceptions of increased engagement in school among participants.

Parent/Family Outcome Measures

Very few sub-grantees (10%) examined parent and family outcome measures. Examples of family and parent outcome measures included teacher survey ratings of parent involvement in school, parent enrollment in higher education courses, percent of parents able to hold a full-time job as a result of afterschool services, computer and English language literacy skills, and stakeholder perceptions of effective parenting skills.

Process Measures

The most common process measures used in sub-grantee evaluations included program attendance, types of activities offered, community partnerships and parent involvement.

Student participation was most often gathered through PPICS data or school records. Examples of the types of data include total number of participants engaged in programming, number of days attended, frequency of attendance (attending at least one day, 30 days, or 60 days or more), demographic characteristics of participants and numbers of families serviced.

The types of activities described in sub-grantee evaluation reports included academic and enrichment activities. Examples of data sources include hours of activities targeting reading, math, science and other subjects; average hours of academic support (reading, math, other) per regular attendee; average hours of ELL instruction for LEP students; and the percentage of centers emphasizing at least one core academic area and percentage of centers offering

enrichment and support activities. Examples of other types of activities include prevention activities (including nutrition, drugs/violence and sex education); average hours of entertainment activities per regular attendee; number and types of arts and poetry activities; and the offering of at least two educational opportunities for families during each year of the program. Most of these data were gathered through PPICS and program tracking.

Community partnerships were also most often measured by tabulating the numbers and types of partnerships in PPICS. Other sources include partner surveys and self-assessments of program sustainability, including additional funding sources, collection of anecdotal "good news" stories to disseminate in the community, and partnerships with businesses and foundations to expand program resources.

Examples of parent involvement measures include parent surveys assessing program attendance and volunteering for program activities, teacher surveys assessing parent involvement, and other stakeholder perceptions of parental input on 21st CCLC program decisions.

Evaluation Research Designs

A key characteristic of an evaluation approach includes the research design used to analyze outcomes. For evaluations of 21st CCLC programs geared toward improved academic achievement and other student outcomes (such as school attendance and behavior), it is important to understand whether and how program evaluators are measuring the programs' effects on student outcomes. This section describes the designs used to analyze and present student outcome data. While program evaluations normally consist of other information as well (process or implementation measures, descriptions of programs and indicators of quality), this section on design focuses specifically on methods used to analyze student achievement outcomes.

When sufficient information was available, SEA and sub-grantee evaluation reports were each grouped into one of the following design categories:

- **Experimental design using random assignment.** This is the strongest and most rigorous research design used in social sciences and educational research, although it can be costly and difficult to implement. This design entails that students are randomly assigned to treatment and control groups (in this context, students would be randomly assigned to 21st CCLC programs, which is impractical to expect). While the BPA team included this category in their coding system, none of the evaluations reviewed used this design.
- **Comparison group is rigorously formed.** This is considered the most rigorous design used. In these studies, evaluators formed a comparison group using statistical controlling or matching techniques to ensure that the comparison group best mirrored the participant group, but did not themselves participate in the 21st CCLC programs. Typically, evaluators that used this method followed the participant and comparison groups over at least two points in time to compare progress. These types of designs are often described as "quasi-experimental".
- **Comparison to districts, states or a national average.** These designs entail that the 21st CCLC participants' average scores are compared to some combination of the following: averages for non-participants in the same schools/districts, district averages overall, state

averages or national averages. For these analyses, the comparison group is not rigorously formed, but usually these types of comparison data are conveniently available.

- **Single group, multiple points in time.** For these designs, a single group of student participants are tracked over time. Typically student-level data is gathered and compared between two points (usually pre- and post- program), or over more than two points in time (e.g., at the end of each school year for three consecutive years).
- **Program comparisons made over time.** In many cases, evaluators examine student outcomes for a group of students that belong to a participant group (e.g., a Boys and Girls Club afterschool program for middle school students in 2009) and compare outcomes to the next group (or cohort) of students who participate in that program (e.g., the same Boys and Girls Club afterschool program for middle school students in 2010). In these cases, different groups of students are compared, but the intent is to examine the progress of the program, not of the individual students.
- **Single group of students, single point in time.** This type of presentation of student outcomes offers the least amount of analysis and information to interpret. For these evaluations, a “snapshot” of data is presented; in other words, data is presented for program participants only, at only one point in time (e.g., math and reading grades at the end of the first year of a program).

Designs used in SEA Evaluations

Table 8 shows the types of research designs used by the SEAs to examine student outcomes and the number of SEAs employing each type of design, based on the evaluation reports reviewed. Our review of these documents and our analysis of the interview data found that SEAs frequently used descriptive designs, in which student outcome data was presented for a single group/multiple points in time (N=17) or as program comparisons made over time (N=11). For the first design type, the SEA looked at the student outcomes for a single group of participants in the 21st CCLC program over time. Oklahoma and Puerto Rico analyzed 1st and 4th quarter grades for regular attendees; Oklahoma also analyzed pre- and post-means on the PPICS Teachers' Survey. Other states also made program comparisons over time, comparing the achievement scores of different cohorts of students across the years of their program's implementation. For example, North Carolina reported conducting a cross-year analysis of three school years and a comparison of the progress of three cohorts across the years. A number of SEAs did, however, develop strong research designs and analytic reports. Examples of these designs are described below.

Table 8. Research Designs used in SEA Evaluations

Research Design	N
Experimental: random assignment	0
Comparison group is rigorously formed	6
Comparison to state or national average	5
Single group, multiple points in time	17
Program comparisons made over time	11
Single group, single point in time	4
Not enough information	5
Missing	6
Total	54

Although no state utilized the strongest experimental: random assignment design, six states used a quasi-experimental design in which the evaluators formed comparison groups using rigorous matching techniques, a common technique being propensity score matching¹¹. The state of Washington, for example, had a strong research design¹² that clearly articulated evaluation questions that focused on the association between 21st CCLC participation and desired program outcomes, and guided a rigorous analysis that specifically addressed these questions. To assess 1) the extent to which there was evidence of a relationship between higher levels of attendance in 21st CCLC programming and the achievement of desired academic and behavior outcomes, and 2) the extent to which there was evidence that particular center and student characteristics and attributes were associated with student academic and behavioral improvement, Washington’s evaluator conducted two types of outcome analysis. A correlation analysis included variables in programs related to math, reading and behavior, and a matched comparison group analysis examined 21st CCLC-eligible students participating in the program for over 30 days and a statistically matched group of non-participants. Within the group of participants, the evaluator also analyzed students’ propensity to attend.

In Virginia, achievement in reading and math was assessed using regression models with two years of test data for students in grades 3-8. Regular attendees (i.e., 30 days or more) were compared to students who were matched on several demographic variables and who were eligible to attend but had zero days of attendance. The use of student achievement data from multiple tests (with standardized scale scores converted to z-scores) allowed the state not only to include analysis of sub-groups that took tests other than the main state standardized test (i.e., ELLs or students with disabilities), but also to capture incremental changes within proficiency levels. Virginia also analyzed center-level variables (e.g., total hours open) to examine impact on student proficiency.

¹¹ This is a method of constructing a comparison group of individuals who are matched to individuals in the intervention group on key characteristics.

¹² Only experimental designs can establish causal relationships between the program and outcomes. However, strong quasi-experimental designs can also be used to estimate program effects.

Another state with a strong research design was New Jersey, which made within-participant comparisons in a two-level model (students at Level 1 and centers at Level 2) that accounted for the nested structure of the data and allowed for the exploration of relationships among center-level characteristics, student-level characteristics, and student-level outcomes that included individual assessment scores in reading and mathematics and the PPICS teacher assessment surveys. To create the comparison group, propensity score matching was used. In addition to the states described above, California will also use multi-level modeling to analyze student outcome data in comparison groups matched using propensity scores¹³.

Less Rigorous Designs

A commonly used design was comparison of 21st CCLC student outcome data to state or national averages (N=5). Colorado, for example, compared 21st CCLC standardized achievement test results on the Colorado State Assessment Program (CSAP) to national data. Illinois compared their GPRA results to a seven-state cohort deemed to be similar (California, Florida, Michigan, New York, Ohio, Pennsylvania and Texas), while Alaska's statewide evaluation likewise compared their student outcome results with those of a group of five neighboring states (Washington, Montana, Oregon, Idaho and Wyoming).

According to document reviews, four SEA statewide evaluations utilized a single group, single point in time design in evaluating student outcomes. Included in this number were reports – often extremely detailed and comprehensive – that summarized data that were reported in PPICS or in data collection and reporting systems developed by the states. Pennsylvania reported PPICS and state data on school attendance, behavior, and reading and math assessment scores, disaggregated by participant sub-groups (those who participated 30/60/90 days); however, no statistical analysis was performed on these data. Montana's statewide program evaluation also reported on data generated from PPICS, with no analysis of the data.

As noted in past years' reports, the BPA team found that some of the documents submitted to the team by the SEAs as "evaluation reports" were actually monitoring reports or implementation progress reports. There was little evidence in reports of this type that the SEA had developed an evaluation research design. In addition, some reports that *did* purport to include a design for evaluation of student outcome data were weak in their reporting or documentation of relevant information, making it difficult to identify what research design was used or how effective it was in guiding the evaluation research and analysis. For five SEAs, not enough information was available to determine the evaluation design. Six SEAs did not submit evaluation reports (or other relevant information), so the team could not determine the design used.

Designs used in Sub-Grantee Evaluations

The evaluation team also reviewed sub-grantee evaluation reports to identify evaluation research designs that included analysis of student outcome data. Table 9 shows the types of research designs used by the 21st CCLC sub-grantees to examine student outcomes and the percentage of sub-grantees employing each type of design, based on the evaluation reports reviewed. Similar to the statewide evaluations, a commonly-used research design by the sub-grantees included single group, multiple points in time (23% of all sub-grantees in sample), program comparisons made

¹³ Although California did not submit a state evaluation report, interviews and examinations of other documents (e.g., evaluator scope of work) revealed that the current statewide evaluation uses a rigorous matching technique to develop comparison groups for the purposes of student outcome analysis.

over time (12%), and comparisons to district, state or national averages (11%), resulting in primarily descriptive evaluation reports.

Table 9. Research Designs used in Sub-Grantee Evaluations

Research Design	N	%
Experimental: random assignment	0	0%
Comparison group is rigorously formed	18	4%
Comparison to district, state or national average	51	11%
Single group, multiple points in time	105	23%
Program comparisons made over time	55	12%
Single group, single point in time	42	9%
Not enough information	50	11%
Missing	137	30%
Total	458	100%

In very few cases (4%), sub-grantee evaluation reports displayed a high level of rigor in the analysis. In Maryland, for example, a sub-grantee developed a quasi-experimental design, forming comparison groups using matching techniques for a comparison group/time-series evaluation design. The sub-grantee planned to collect outcome data for comparison groups drawn from the student population that was not participating in the program at each of four schools and analyze program participants' test scores according to program site, gender, grade, race, special education status and free or reduced lunch status.

Using a comparison of 21st CCLC student outcome data to district, state or national averages, a sub-grantee in Virginia evaluated its 21st CCLC program's progress in meeting objectives that included increasing the percentage of students in the economically disadvantaged subgroup scoring "proficient" on the state math assessment. Participants' gains were compared with average percentage gains of economically disadvantaged students at the district and state level. In Arizona, a sub-grantee presented charts comparing their grant to state and national statistics.

The most commonly used design was single group, multiple points in time. For example, a sub-grantee in South Carolina compared student grades in the fall and spring, i.e., before and after program participation, to gauge improvement in math, science, ELA and social studies, while several sub-grantees compared student standardized assessment scores from fall to spring. An Arizona sub-grantee also compared participant grades from fall to spring; these participants appeared to show improvement according to post-program data. A sub-grantee in Georgia evaluated its 21st CCLC program with a design that involved making program comparisons over time, by presenting academic outcome data for different cohorts of participants over multiple points in time (the 2010-2011, 2009-2010 and 2008-2009 school years). Similarly, an Alaska sub-grantee presented the standardized test scores of regular attendees at each of its centers over

a three-year period. In neither case was there additional analysis beyond the descriptive comparison of the different groups over time.

As occurred with the SEAs, some of the documents submitted to the BPA team by sub-grantees as evaluation reports were actually monitoring reports, implementation progress reports, or data summaries containing little or no analytic detail. This occurred frequently with the sub-grantee evaluation reports and accounts for the 11% of the sub-grantees for which not enough information was available to determine the research design. For almost one third of the sub-grantees in the sample (30%), evaluation reports were not submitted, making it impossible to determine if any evaluation was conducted and/or what type of design was used to evaluate student outcomes.

Variations in Research Design across SEAs

The BPA team conducted several analyses to determine if SEA and sub-grantee research designs varied by selected state characteristics. Below, analyses are presented that examine SEA size; type of evaluator; sub-grantee variation with SEAs; and sub-grantee design based on SEA design, required SEA performance measures and required targets.

Research Design by State Size

Analysts examined the relationship between the size of the SEA and the research design used to analyze student outcomes to determine if states with more funding would have more rigorous evaluation designs. Size of state was defined as follows:

1. Small: SEAs awarded grants of under \$10 million in 2011.
2. Medium: SEAs awarded grants between \$10 million to under \$30 million in 2011.
3. Large: SEAs awarded grants of \$30 million or more in 2011.

Table 10 displays the variation in research designs across the three size categories. There was a greater percentage of large states that used a rigorous design compared to small or medium-sized states, though the sample size across the three groups varied. None of the large states used the most basic design of presenting data for a single group at a single point in time. The most common design for large states (N=10) was single group, multiple points in time. This includes designs in which student outcomes were analyzed using pre-/post-program measures, and designs which include data collected for more than two points in time. For medium-sized states, the most common design used was also single group, multiple points in time.

Table 10. SEA Research Design by Size of State

Research Design	Small	Medium	Large
Comparison group is rigorously formed	2	2	2
Comparison to state or national average	4	1	0
Single group, multiple points in time	3	8	6
Program comparisons made over time	5	4	2
Single group, single point in time	3	1	0
Missing	7	4	0
Total	24	20	10

Small SEAs were more varied as a group. The most commonly used design was program comparisons made over time, in which different cohorts or groups of students were compared. Four small SEAs compared participant student outcomes to state or national averages. For four small SEAs, design information was missing or unable to be determined.

Sub-Grantee Design by SEA Size

The team also examined how sub-grantee research designs varied by the size of their SEAs. Table 11 presents the results of the cross-tabulation of SEA size with sub-grantee design. The Chi Square test reveals an association between the size of the SEAs and their sub-grantees' research designs (p=0.000).

Table 11. Sub-Grantee Research Design by Size of State**

Research Design	Small	Medium	Large
Comparison group is rigorously formed	1%	13%	3%
Comparison to district, state or national average	24%	9%	28%
Single group, multiple points in time	38%	40%	37%
Program comparisons made over time	21%	29%	7%
Single group, single point in time	15%	9%	25%
Total N	84	112	75
<i>** Pearson Chi Square (8) = 42.1245 Pr = 0.000</i>			

The BPA team expected that large states would have sub-grantees more likely to use rigorous designs. However, most of the sub-grantees that used rigorous designs (comparison group is rigorously formed) came from the medium-sized states (13%). Sub-grantees that used the most basic design came from large states (25%).

State Design by Type of Evaluator

In addition, BPA analysts explored the relationship between the research designs used in evaluating student outcomes and the types of evaluators SEAs used. Table 12 shows a cross-tabulation between these two variables. The SEAs that used rigorous designs had evaluators that came from academic institutions or research agencies. In addition, most evaluators from academic institutions (7) examined student outcomes by following a single group of students over multiple points in time.

Table 12. SEA Design by Type of Evaluator*

	Comp. group is rigorously formed	Comp. to state or national average	Single group, multiple points in time	Program comp. made over time	Single group, single point in time	Not enough info/ Missing
State/ district personnel	0	0	2	3	1	1
Academic institution	2	0	7	1	0	2
Independent consultant	0	0	2	3	0	1
Research agency	3	3	5	3	3	2
Missing	1*	2	1	1	0	5
Total N	6	5	17	11	4	11

*Missing information refers to California's evaluator; data is missing because a report was not submitted. However, the BPA team gathered information from other sources indicating that the evaluator is a research center based at a university.

Sub-Grantee Design by Type of Evaluator Used

The team also explored sub-grantee research designs in relation to the types of evaluators they used (Table 13). A Chi Square test revealed that there is a significant relationship between the two variables. For 75% of the sub-grantees that used a rigorous design, the evaluator was either an independent consultant or a research agency.

Table 13. Sub-Grantee Designs by Sub-Grantee Evaluators**

	Comp. group is rigorously formed	Comp. to district, state or national average	Single group, multiple points in time	Program comp. made over time	Single group, single point in time
Program staff	17%	12%	24%	37%	31%
State/ district personnel	0%	2%	4%	2%	0%
Academic institution	8%	38%	17%	14%	6%
Independent consultant	33%	22%	16%	33%	28%
Research agency	42%	24%	38%	14%	28%
Other	0%	2%	1%	0%	6%
Total N	12	50	89	43	32

**Pearson Chi Square (20) = 39.6530 Pr = 0.006

For sub-grantees that used comparisons to district, state or national averages, 38% of evaluators were academic institutions, 24% were research agencies and 22% were independent consultants. For the most basic design (single group, single point in time), the evaluator was most likely program staff (31%). These results should be interpreted with caution since a large number of sub-grantee-level evaluation designs are unknown (11%) or missing (30%), and the numbers reflected in the table represent only about 60% of all sampled sub-grantees.

Variations of Sub-Grantee Design within States

Another characteristic the BPA team examined was the variation of sub-grantee research designs used within SEAs. Did sub-grantees within the same SEA use the same research design?

According to Table 14, there was a great deal of variation within SEAs. In only five SEAs did all sub-grantees use the same research design to examine student outcomes. In four SEAs, most sub-grantees (one outlier) used the same design, and in eight SEAs, sub-grantees used one of two designs. For 24 SEAs, sub-grantees used one of three (or more) research designs. For 13 of the SEAs, the team could not determine sub-grantee variation for several reasons. It is possible the SEA only had research design data for one or two sub-grantees, or SEA and/or sub-grantee data was missing.

Table 14. Variation of Sub-Grantee Designs within SEAs

	States (N)
Sub-Grantees all use same design	5
All sub-grantees except one use same design	4
Sub-grantees use one of two designs	8
Sub-grantees use one of three (or more) designs	24
No data	13
TOTAL	54

Table 14 suggests that overall a great deal of variation exists not only across states, but *within* SEAs, about how sub-grantees approached evaluation design.

Sub-grantee design by SEA research design

One theory the BPA team examined was that states that used rigorous research designs for their student outcome analysis would have sub-grantees that also used rigorous designs in their evaluations. Table 15 displays how sub-grantee research designs vary by their state research designs. The Chi Square test results were significant, indicating that SEA design and sub-grantee research design are related. For SEAs that used a rigorous design (comparison group formed using matching technique), 21% of sub-grantees from these SEAs also used this design; 27% used comparisons with district, state or national averages, and 12% presented data for a single group/multiple points in time. The most commonly used design in this group was program comparisons made over time (30%).

Table 15. Sub-Grantee Evaluation Designs by State Level Design Categories**

Research Design	1	2	3	4	5
Comparison group is rigorously formed (1)	21%	5%	2%	6%	0%
Comparison to district, state, or national average (2)	27%	10%	21%	6%	75%
Single group, multiple points in time (3)	12%	55%	49%	34%	8%
Program Comparisons made over time (4)	30%	30%	11%	34%	8%
Single group, single point in time (5)	9%	0%	17%	19%	8%
Total N	33	20	118	67	12
<i>**Pearson Chi square(24) = 100.6074 Pr = 0.000</i>					

For SEAs that use the most basic design (single group of students, single point in time), 75% of their sub-grantees used comparisons to district, state or national averages in their evaluations. For SEAs that used the other three designs, most of their sub-grantees used single group, multiple points in time; program comparisons made over time; or single group, single point in time as their research design.

Sub-Grantee Design and Required Performance Measures

The BPA team sought to explore whether SEAs that required specific performance measures had sub-grantees that use rigorous research designs. In providing guidance and requirements to sub-grantees, many SEAs required that sub-grantees use specific performance measures in the sub-grantee evaluations. Out of all 54 SEAs reviewed, 44 required some specific outcome measures be used in evaluations. Outcome measures required by the states include national, state or local assessments of academic achievement, grades, and teacher-reported academic performance and student behavior. According to documents reviewed, the most common measures of academic achievement required by the SEAs were standardized state assessments in reading and math. Idaho, for example, requires applicants to describe their evaluation plans and advises them that these plans must utilize the Idaho State Achievement Test and the Idaho Reading Indicator. Colorado requires the Colorado State Assessment Program (CSAP) and Arizona requires the Arizona Instrument to Measure Standards (AIMS). In addition to state assessments in reading and math, some states require sub-grantees to use other state or local academic achievement measures, or national standardized tests such as the SAT (Illinois), the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) assessment (Oregon, Arkansas), or Northwest Evaluation Association Measures of Academic Progress (NWEA MAP) (Missouri, Indiana).

Many states also required that sub-grantees utilize classroom grades as measures of academic achievement, with the usual requirement being quarterly report cards. Georgia sub-grantees are required to collect both benchmark grades and reports of grades after one year of participation in the program. Illinois sub-grantees collect pre/post grades specifically in the subjects of reading and math. Other outcome measures that states required were the PPICS teacher survey and school or classroom attendance. The PPICS survey includes teacher perceptions of student academic improvement as well as teacher reports of non-academic outcomes, such as homework completion and student behavior (e.g., participation in class, getting along with others, etc). Table 16 displays the results of this analysis.

Table 16. Sub-Grantee Design within SEAs that Require Performance Measures**

Research Design	Require (SEAs = 44)	Do not require (SEAs = 10)
Comparison group is rigorously formed	5%	7%
Comparison to district, state or national average	30%	15%
Single group, multiple points in time	28%	42%
Program comparisons made over time	33%	16%
Single group, single point in time	5%	19%
Total Sub-Grantees	64	207
<i>**Pearson Chi Square (4) = 20.9670 Pr = 0.000</i>		

According to the Chi Square analysis, it appears that requiring performance measures is associated with research designs among sub-grantees; however, results lack a strong, clear pattern. A single group, multiple points in time approach was the most common approach used by sub-grantees within states that did not require performance measures, and it was used less often in states that did prescribe performance measures. Use of the single group, single point in time approach, which is the least rigorous, was more common in states that did not require specific performance measures. On the other hand, among sub-grantees that used the most rigorous design, there seems to be very little difference; 7% came from SEAs that did not require performance measures and 5% were from SEAs that did require such measures. These results should be interpreted with caution since an overwhelming majority of SEAs required some sort of performance measures.

Sub-Grantee Design and Required Targets

Another relationship the team explored was between SEAs that required specific performance targets and sub-grantee research designs. Out of 54 SEAs reviewed, the team found that 12 set specific targets for student achievement. In New Mexico, the SEA required that each 21st CCLC program increase student proficiency in reading and math as measured by the NMSBA (the New Mexico Standards Based Assessment), requiring that they meet AMOs (Annual Measurable Objectives), “meet proficiency by confidence interval,” or reduce non-proficiency by 10% or more over the prior year. Alabama required that 70% of students in the extended day/year program who attended over 30 days and took the SAT 10 improve at least two percentile ranks on the SAT 10 test in either reading or math between each year. In addition, “the program must show improvement,” in the most recent Alabama Reading & Mathematics Test (ARMT), i.e., students attending over 30 days must perform at or above proficiency in ARMT Reading and Math.

In Kansas, the state set overall performance goals and indicators that followed the federal 21st CCLC criteria (the GPRA requirements) and required sub-grantees to add their own goals/targets. According to the document review, several sub-grantees in Kansas chose the following targets:

- 80% will maintain or improve high academic achievement in reading and math;
- 60% of participants will have increased homework completion;

- 90% of students will not be suspended;
- All students will graduate; and
- 80% will not engage in violent activity at school.

Wisconsin set targets based on teacher-reported academic achievement and student behavior, including:

- 95% of programs to have at least 40% of teachers report improved academic performance for program participants; and
- 95% of programs to have at least 40% of teachers report improved homework completion to the teachers' satisfaction.

Table 17 displays the proportion of sub-grantees using each type of research design within each group of states (SEA requires specific targets=Y; SEA does not require targets or it is not known=N).

Table 17. Sub-Grantee Design within SEAs that Require Targets

Research Design	Require (SEAs = 12)	Do Not Require (SEAs = 42)
Comparison group is rigorously formed	0%	7%
Comparison to district, state or national average	27%	18%
Single group, multiple points in time	36%	39%
Program comparisons made over time	0%	21%
Single group, single point in time	36%	15%
Total Sub-Grantees	11	260

Due to missing data, it is difficult to draw conclusions about whether specifying targets is associated with rigorous design. Furthermore, the Chi Square test did not yield significant results about the association between the two variables. The results show that the few sub-grantees (7%) that did use the most rigorous design did not come from states that have set targets.

Changes in Evaluation Approaches over Time

Over the three years of the Evaluation Review, the BPA team has engaged in discussions with SEA representatives and evaluators about various aspects of their 21st CCLC program evaluations. Respondents from many SEAs shared background and context for present practice, such as how the SEA had done evaluation in the past and how they have made changes to present practice, and lessons learned and plans for changing their evaluation practices in the future. In addition to document reviews and on-site interviews, the evaluation team conducted follow-up phone interviews with the following states (from the 2010 sample) in 2012 to gain updated information about evaluation practice: Idaho, Illinois, New Hampshire, New York and Oregon.

Interviews and discussions with SEAs revealed that many of the states are making substantial changes in their statewide evaluations. In this section, changes are discussed related to the

following evaluation practices: evaluation plans, measures, frequency of evaluation, evaluator roles, data collection, research design, and alignment with state programs and sub-grantee evaluations.

Evaluation Plans

Some states discussed changes in evaluation planning over the two-year period. A few SEAs that did not have a statewide evaluation in place at the time of the Base Year review (2010) have made much needed progress in implementing a statewide evaluation. For example, at the time of the monitoring visit in 2010, Oregon had not yet conducted a statewide evaluation (though they were in the process of developing an evaluation plan) and received findings pertaining to this issue. Immediately following the monitoring visit, the state contracted with Gibson Consulting for site-visits and with AIR, which is currently in the process of completing the first evaluation report for the 2010-2011 academic year. The evaluation plan has changed from simple aggregate data analysis to looking at outcomes by sub-groups: school, district and grade level. Virgin Islands is similarly working with an external evaluator to conduct the first statewide evaluation for the 21st CCLC program. Other states, like Georgia, have not conducted a statewide evaluation for several years; Georgia is working towards having a statewide evaluation conducted by the end of 2012, with a statewide evaluation conducted annually thereafter.

Other states that had some type of evaluation in the first year of evaluation review have made a complete overhaul of the evaluation design between 2010 and 2012. For example, Alaska will be using new data collection tools that measure positive youth development; the Alaska Assessment Tool and Alaska Site Observation Tool were originally developed and tested in Colorado and Massachusetts and adapted to the needs of Alaska. New Hampshire has also made improvements to the statewide evaluation over the last few years. In 2008, through its contractor Susan Frankel, New Hampshire conducted an overall assessment of availability of afterschool programming across the state; while this report included information on 21st CCLC out-of-school programs, it focused more on access to services rather than student outcomes. The state has since contracted with Policy Studies Associates (PSA) and is working with them to develop systems to conduct 21st CCLC-specific and outcome-oriented evaluation. Because of the marked change in the evaluation design, the state anticipates having major changes in the way the evaluation results will be used for program improvement. Arkansas experienced a related issue, having received findings related to its statewide evaluation. The state, during the monitoring visit, expressed their awareness that the statewide evaluation report did not provide much useful information on the 21st CCLC program. The SEA contracted with MGT for data collection services and Arkansas State University-Jonesboro for services in YPQA. The SEA Coordinator planned on working more closely with Arkansas State University-Jonesboro to incorporate academic outcome measures with the quality improvement work the center is already conducting.

Evaluation Measures

Several states made changes to the outcomes measures they use in their evaluations. Some examples include the following:

- West Virginia, in the most recent evaluation, only assessed student academic outcome through teacher surveys inquiring about changes in students' academic performance; the SEA plans to use WESTEST 2, the state's standardized assessment in reading and math, to look at students' academic outcomes in the future.

- Kentucky incorporated a QIP process into the sub-grantee evaluation. SEA representatives reported that sub-grantees better understood the process of conducting a QIP rather than the general instruction to conduct evaluation, and saw the QIP as highly valuable.
- Florida's sub-grantees used a wide spectrum of academic assessment tools for their local evaluations, hindering the state evaluator's efforts to aggregate that information at the state level. The state evaluator would like to either develop an index for each tool so that the evaluator can make results comparable across tools and aggregate at the state level, or move towards having sub-grantees use the same tools for academic assessment.

Frequency of Evaluation

A few states were interested in changing the frequency of evaluation. In the case of Alaska, at the time of the monitoring visit, the state's most recent evaluation report was a five-year analysis from FY 2004-2008. The state noted that, in the future, the statewide 21st CCLC evaluations will be conducted annually.

Evaluator Qualifications and Roles

The evaluators, evaluator roles and responsibilities, and contract amounts have also changed from the first to last year of the evaluation review. One prominent example is in Idaho, which contracted with Creating Change, Inc., for monitoring, data management, evaluation and technical assistance services. One major challenge was that Creating Change's data systems were not designed for program evaluation, and the state received several evaluation-related findings during their last monitoring visit. Ultimately, with a new SEA coordinator in 2011 and unsuccessfully trying to adapt Creating Change's systems to meet program needs, Idaho decided to contract to a new organization, Education Northwest, to "tell the story more effectively."

The state of West Virginia is unique because, in the face of increasing costs with their evaluation contractor, the SEA decided to have the Office of Research within the state's Department of Education serve as the external evaluator. To ensure objective evaluation, the Office of Research cannot participate in, for example, technical assistance activities to the state; however, this change in evaluators has addressed the issue of cost in regards to evaluation services.

New Hampshire is another unique case in that the SEA has been working with PSA in what it describes as a three-year process, which not only includes data collection and analysis but the initial time needed to build systems and protocols. The SEA appears to have sought external services to develop the initial infrastructure but seeks to conduct statewide evaluations internally in the future. The California SEA would also like to internalize some of the work related to evaluation. Interview respondents explained that the SEA would be able to better monitor the data collection process if it were done internally. Furthermore, the SEA has more accurate data than what sub-grantees are submitting or to which the evaluator, CRESST, has access. The SEA believed internalizing data collection but hiring an external evaluator to conduct the data analysis and reporting would streamline the evaluation process.

Illinois requires a new evaluator but is seeking to keep the roles and responsibilities and scope of evaluation nearly identical to the previous evaluator, who retired. For the new contract, the state is requesting that bidders implement the existing evaluation design, use tools that have already been developed and analyze already collected data. While the previous contract was for \$167,000 per year, the new contract will be for \$200,000 per year, based on the same scope of work.

Data Collection

Several SEAs are improving their data collection and management systems. The state of New York is working with a contractor to develop what the SEA describes as a “data hub” for all of the program’s data related to technical assistance, monitoring and evaluation, housed on a secure website. Colorado’s sub-grantees have experienced issues with collecting data for submission into PPICS. Therefore, the SEA, at the time of the monitoring visit, was pilot-testing two alternative data collection systems with select sub-grantees to determine if there are easier and more efficient data management systems to use in the state. New Jersey similarly made improvements to their data collection system, the Evaluation Template and Reporting System (ETRS), developed by AIR. The system now allows sub-grantees to assess center-level data through the system.

Regarding data collection tools, Minnesota plans to make changes to the QIP tool used in the state. Minnesota required cohort 4 to choose a QIP process from the Forum for Youth Investment’s Measuring Youth Program Quality Compendium, which includes tools like the YPQA or NYSAN. Minnesota would like to examine the types of QIP tools sub-grantees are choosing and how they are reporting data; through this review, the state would like to narrow the list of QIP tools that will be used in the state to assess this program.

Rigorous Design

A couple of states expressed a desire to move towards developing comparison groups within their evaluations. West Virginia is planning to use propensity score matching to create a non-participant comparison group to analyze student academic outcomes, primarily using state assessment scores, between the participant and non-participant groups. The University of Florida (UF), the external statewide evaluator for Florida, would also like to move towards including a comparison group in the analysis. UF is in the process of working with three CBOs to potentially get a group of students to serve as a comparison to program participants. The primary barrier to conducting a comparison between program and non-program participants is getting student data for the latter group, as districts are reluctant to release information on non-program participants.

Alignment with Other State Programs or Sub-Grantee Evaluations

Some SEAs expressed the desire to integrate the efforts related to 21st CCLC with other related state programs. For example, Idaho would like to fold the monitoring efforts for 21st CCLC in with Title I programs. Previously, the 21st CCLC program operated in isolation. However, due to a paradigm shift in recent years, the state’s Department of Education seeks to look at “coherence across systems.” SEA representatives reported that having an evaluation of all programs, including 21st CCLC, will be a positive shift towards both integrating related programs as well as creating, as stated by the SEA coordinator, a “positive lack of comfort” – indicating that state program staff cannot become complacent if the 21st CCLC program is no longer assessed in isolation. Oregon expressed similar goals of integrating assessment efforts, like evaluation or monitoring, across SEA programs.

Another example is New Hampshire, which now integrates 21st CCLC data with the data system for the state’s Department of Education. Therefore, 21st CCLC data, such as teacher, student, principal or program director surveys, are linked to other student data using the designated student identification. Matching student program data to student data in the state’s data system has, as expressed by the SEA, “open[ed] the door to analysis.”

Other changes that states have expressed wanting to make is tying sub-grantee evaluations with that of the states. Illinois has asked sub-grantees to create goals and objectives to align with the goals and objectives for the state, like increasing involvement of parents or improving student achievement based on the annual yearly progress (AYP). Oregon has also expressed a broad goal of wanting to align local evaluations with the state evaluation.

Promising Practices in Evaluation Approaches

The BPA team has identified several promising practices related to evaluation design, data collection tools and procedures, evaluator expertise, and reporting. Examples of these are discussed below, organized by topic.

Evaluation Design

The BPA team classified some types of evaluation research designs as promising. For example, the evaluator in Virginia analyzed student-level data to measure longitudinal changes, enabling program providers and evaluators to accurately measure student progress over time. Also, a few states reported plans to conduct more rigorous analysis in the future, for example, West Virginia's plan to use propensity score matching to create a controlled comparison group (made up of non-participants). The BPA team also identified sub-grantee designs that rigorously compared outcomes between participants and non-participants. For example, a sub-grantee in Illinois compared academic outcomes, behavioral outcomes and school climate outcomes (as measured by a student survey) between participants and non-participants. A sub-grantee in Illinois compared test scores and grades between program attendees and non-attendees. In addition, a sub-grantee in South Carolina analyzed differences in the number of [disciplinary] referrals among "local regular students" and among 21st CCLC students. Sub-grantee staff reported that they were proud of being able to show this difference in order to demonstrate that their program had a positive impact.

Data Collection Tools

The team identified several instances in which the state provided effective data collection instruments for sub-grantees as well as instances of sub-grantees developing their own effective data collection instruments. For example, in California, the SEA is working with a University to develop and validate evaluation instruments for local afterschool programs that will eventually be available for all sub-grantees. Also, Delaware provided a user-friendly quality assessment tool that has been adapted from the North Carolina Center for Afterschool Programs. The tool has 11 program goals, worksheets for diagnosing deficiencies, action plans for improvement and measures for assessing whether the improvement has occurred. A sub-grantee evaluator in Colorado worked closely with the program staff on designing evaluation tools. The staff received training on the tools and appreciated that they could modify them as program improvements occurred or program needs changed. In addition, a sub-grantee in South Carolina reported that their development of a survey that asked alumni about their personal experiences with the program was a promising practice.

Data Collection Procedures

SEA and sub-grantee staff reported several practices that helped improve their data collection. In Georgia, the SEA developed a process to use a state testing ID to extract students' state standardized test scores in order to facilitate data collection among non-LEA programs. SEA staff reported that non-LEAs had to secure agreements with schools to get test data and felt that it was more difficult for these programs to obtain data than LEA programs. The SEA staff in

Puerto Rico reported that they used to have trouble getting necessary data when programs provided the data and decided to get all pertinent data directly from a state system. Sub-grantees reported that strong partnerships with school and parents helped them obtain data. A sub-grantee in DC reported that they increased their parent survey response rates by increasing communication with parents. More specifically, they reminded them about the surveys when parents picked up their kids and sent email reminders. A sub-grantee in Maryland felt that their strong relationship with the school helped them increase response rates. The sub-grantee reported that they were able to administer surveys during parent events. Similarly, a sub-grantee in Michigan reported that the active involvement of the LEA streamlined their access to data.

Evaluator Competency

Several SEAs and sub-grantees reported that they had good relationships with their evaluator and deemed this as a promising practice. For example, a sub-grantee in Indiana reported that one of their strengths was that the evaluator was able to translate state requirements into information that is easily understood by program staff, engage with program staff directly about the results and explain what the results mean in terms of the programming. The state staff in North Carolina reported that the evaluator helped the state learn about needed program improvements at the state and sub-grantee level. Also, the SEA staff in Virginia reported that their evaluator's research and evaluation expertise created confidence among the program staff that they were receiving a high quality evaluation.

Evaluation Reports

The research team classified several instances of high quality evaluation reports as a promising practice. For example the Illinois statewide evaluation report included four clear research questions, concrete recommendations for program improvement as well as data collection, and an explanation of the limitations of the data. Another example is the Wyoming statewide evaluation report, which clearly explained the response rates and warned the reader about inferring too much from data based on low response rates. The research team also identified several sub-grantee evaluations that were of higher quality than the norm. For example, a Georgia sub-grantee evaluation report included an explanation of program goals, research questions, a discussion of data collection methods, recommendations related to program improvement and improving the evaluation process, as well as a coherent narrative. An Indiana sub-grantee report was also identified as promising because it included program objectives, research questions, a crosswalk between research questions and objectives, a detailed methodology section and recommendations.

Summary

This chapter has covered a great deal of information pertaining to the evaluation approaches undertaken by SEAs and sub-grantees, including frequency of evaluation, types of evaluators used and how they are selected, the amount of funds spent on evaluation, key evaluation questions, measures, research designs, changes in evaluation approaches over time and some examples of promising practices. The data reveals that evaluation approaches varied widely across states and even across sub-grantees within a state. With a few exceptions, most SEAs and sub-grantees used descriptive designs to analyze student outcomes. The team has noted SEAs that use sophisticated and rigorous designs (e.g., New Jersey, Washington, Virginia). The results of some analyses were surprising; for example, most sub-grantees that used rigorous research designs came from medium-sized states (rather than large or small). Other findings were not surprising, for example, that the most common outcome measures used were state standardized

assessments for math and reading/English language arts. Overall, programs used a variety of approaches to evaluation, some incorporating a quality improvement process and others strictly using PPICS data to assess outcomes, though this assessment provides little information about specific program effects.

The next chapter summarizes key findings from 21st CCLC grantee and sub-grantee program evaluations.

IV. Key Findings from 21st CCLC Evaluations

Through document review, data was gathered on key findings from SEA and sub-grantee evaluations. Analysts recorded findings that primarily related to student academic outcomes and were reported in evaluation reports. This chapter summarizes the different types of key findings the BPA team found through this review. In the first section, findings from SEA evaluations are discussed; in the second section, findings from sub-grantees are described.

Key Findings from SEA Evaluations

SEAs examined a wide range of measures in their evaluations and revealed various levels of detail and rigor in the findings they presented in evaluation reports. This section will focus on key findings presented in SEA evaluation reports that focus on student academic outcomes.

Findings Based on Descriptive Data

Some states presented findings based on descriptive data. These findings were typically the results of evaluations that presented data for a single group of students (or participants) at a single point in time (usually the end of the academic year), or compared participant outcomes to a state or national average. Examples include:

- 91% of students report feeling more prepared when going to school; 96% of parents report that their children are in better academic standing (Alabama).
- The percentage of elementary 21st CCLC regular program participants who improved from “Not Proficient” to “Proficient” on a reading state assessment (48.9%) was notably higher than the national average (22.8%). 48.7% of BIE’s regularly participating middle and high school students improved from “Not Proficient” to “Proficient” on math state assessments compared to 15.9% nationally.

Findings Based on Rigorous Analysis

Other SEA reports showed a higher quality of research and richer findings based on their respective evaluations. Below are some examples of more specific results, based on rigorous analysis, found across SEA evaluations:

- The evaluator for Maine, using regression analysis, identified key characteristics across centers that accounted for much of the variation in students’ changes in performance outcomes. The key explanatory variables were number of paid teachers, number of volunteers, average daily attendance of students and average daily attendance in homework activities.
- Nebraska’s evaluator found that students who attended 90 days or more had the best rating in math and reading among program participants. Using ANOVA analysis, the evaluator determined that these sub-group differences were statistically significant.
- In New Jersey, for programs operating during the 2008-2009 and 2009-2010 academic years, there was a small but positive significant impact of the program on the mathematics state assessment results for students that participated in the program regularly, especially for 70 days or more, compared to a comparison group made up of students from the same schools that did not participate in 21st CCLC programming. In addition, multiple years of participation in 21st CCLC were found to be positively associated with student performance on state assessment outcomes in both reading and

mathematics.

- The most recent Virginia evaluation report found that programs with "too many" activities led to lower academic outcomes. The total hours of activities was associated with lower standardized Standard of Learning (SOL) math scores, while the total number of activities was associated with a lower number of students scoring proficient in reading.

Not all SEA reports reflected the same results; Washington found that increased program attendance did not correlate with higher standardized scores, while states like Maryland concluded the opposite. Conflicting outcomes were also reported within states. Key findings from sub-grantee evaluations are presented below.

Key Findings from Sub-Grantee Evaluations

Most findings presented in sub-grantee evaluation reports were inconclusive about the effectiveness of 21st CCLC programs for the following reasons: the findings presented in the reports were not given sufficient context for the proper interpretation of data; findings were often not supported by sufficient evidence; and sub-grantees reported varying results of their 21st CCLC programs' effectiveness. Despite these limitations, a few sub-grantees presented solid findings based on clear evidence of data analysis. Examples are provided below.

Findings Based on Descriptive Data

One of the ways in which sub-grantees reported on student outcomes was to report descriptive statistics, primarily sums or percentages, of the number of students that met a threshold. Several sub-grantees reported on the proportion or number of participants who met a target – either one specified by the SEA or generated by the sub-grantees themselves. For example, a sub-grantee in South Carolina set the following targets: at least 50% of students attending the program at least 85% of the time will not engage in drug use or violent acts and will not be suspended in school or out-of-school during their involvement in the program. The sub-grantee did not meet this objective, though it reported that disciplinary reports show that 36.2% of the participants had no behavioral incidents. By assessing the program based on pre-determined (and possibly arbitrary) targets, it is not clear if the targets were set with pre-program performance in mind or whether the targets were set too high or too low in measuring adequate progress.

Another sub-grantee reported that, based on teacher reports, 70 students in the program graduated to the next grade. However, not knowing the total population size limits interpretation of the data. Furthermore, as these are descriptive statistics, the data does not reflect whether the students made an improvement after joining the program or if the program led to improvements the students would otherwise be making. The lack of any comparison greatly limits the usefulness of the results. Another limitation of findings, particularly related to survey data, has been that some of the results are based upon low response rates.

Lack of Evidence

Another challenge to reporting findings has been to support them with data or evidence. Roughly half of the sub-grantee documents specified what data was used, how it was analyzed and how the sub-grantee determined the findings. However, the other half of the sub-grantee documents reviewed did not clearly support the findings with data or evidence. Some omissions included 1) type of data collected and data collection source or method; 2) survey response rates; or 3) reporting of actual data and presentation of data tabulation (e.g., numbers, percentages, statistical tests). For example, a sub-grantee in D.C. reported that the program participants had significantly

higher average grades from the first to fourth quarter of the year, but did not provide evidence to support that claim.

Mixed Results

Despite some limitations of the reported findings, sub-grantees reported some promising findings. A South Carolina sub-grantee found that state standardized ELA scores for program participants increased by 11.3%, science scores increased by 12.8%, social studies scores increased by 2.4%, math increased by 17.5%, and writing increased by 17.7%, providing evidence that math and writing results were statistically significant. Other sub-grantees reported varying levels of program effectiveness through their evaluation reports. Two centers in D.C. reported grades in language arts actually decreased over the course of the year. Others, such as sub-grantees from Maine, reported positive outcomes, with teachers finding that 89% of students made gains in academic performance while the rest had some combination of positive and negative outcomes. However, for the vast majority, sub-grantees reported improvements in student outcomes to some degree, whether minimal or substantial.

Findings Based on Rigorous Analysis

Some of the strongest findings included sub-grantees that compared program outcomes across sub-groups based on duration of program attendance (e.g., attended less than 30 days versus attended more than 60 days), between participants and non-participants, and between sub-grantee participants and the outcomes of 21st CCLC participants across the state or nation. Below are some examples of the performance outcomes found in the more rigorous sub-grantee evaluation reports:

- A sub-grantee in DC found:
 - In reading, the benchmark assessment scores for program participants in terms of growth level is 1.03. For non-program students, the growth level is .98. The more a student attends [the program], the greater the level of observed growth. Students that attended the program 120 days or more observed growth levels of 1.11 while students that attended the program 150 days or more observed growth levels of 1.27.
 - On the D.C. Comprehensive Assessment System (CAS), the percentage of program students that tested at the Below Basic level increased by 13%, which was similar for non-program students at the end of the academic year.
- In Indiana, a sub-grantee found that 6th to 8th grade students attending 60 or more days in the program had higher grade point averages than students in other attendance levels.
- Another Indiana sub-grantee found: 80% of regular attendees who had unsatisfactory grades during the first grading period improved their reading grades by the end of the year, and 83% improved math grades. Approximately 27% percent of regular attendees who had unsatisfactory grades during the second grading period improved their reading grades by the end of the year and 67% of students improved their math grades.

Summary

Key findings presented in both SEA and sub-grantee evaluations were mostly positive, but in many cases (especially for sub-grantees), lacked sufficient evidence to justify their claims. The insight from the statewide evaluations tended to be richer than those found at the sub-grantee level, which provided descriptive data but less context and analysis. State evaluations also

provided a greater focus on student, particularly academic, outcomes than the sub-grantee evaluations. Nearly all SEA evaluations reviewed supported findings with data or evidence, whereas sub-grantees were less consistent in corroborating findings with data in documentation. There were a few examples of program evaluation reports that provided findings based on rigorous analysis. However, due to the limitations discussed above, audiences should be cautious in interpreting findings from 21st CCLC evaluation reports.

V. Use and Dissemination of Evaluation Results

In this chapter, the BPA team presents information about how SEAs and sub-grantees used and disseminated evaluation results. Federal regulations require SEAs to use evaluation results for program improvement and to make evaluation reports public. In addition, SEAs must ensure that sub-grantee evaluation or assessment results are:

- (1) used to refine, improve and strengthen the program and to refine the performance measures; and
- (2) made available to the public upon request.¹⁴

This chapter begins with descriptions of how SEAs and sub-grantees used evaluation results. It then presents information about how SEAs and sub-grantees disseminated evaluation results to stakeholders and the public.

SEA's Use of Evaluation Results

SEA's use of evaluation results falls into four broad categories: program improvement, informing technical assistance, improving evaluation measures or procedures, and program advocacy and procurement of additional funding. Examples of these types of use are described below.

Program Improvement

The majority of SEAs primarily used statewide evaluation results for program improvement and program implementation. For example, Illinois used its evaluation to provide procedure and policy recommendations for the 21st CCLC program and for setting priorities for the coming year. The SEA also used the evaluation to assess economic viability of each sub-grantee, the resources needed to sustain the program and appropriate budget allocation for the next year. Kentucky used their evaluation results to identify successful programs, spot areas of program improvement, determine what the goals should be in the state action plan, and to inform sub-grantees during the director's meetings about the program improvement process. Self-assessment results were also used by Kentucky to develop plans for program improvement. Louisiana used evaluation results by presenting PPICS data as a tool that sub-grantees could use for continuous program improvement, and to support program improvement and implementation. Some states used evaluation results to identify promising practices, challenges and barriers within their programs. The SEA coordinator for Puerto Rico used evaluation results to identify needs and establish a monitoring agenda when meeting with program providers. Wisconsin used evaluation results to improve the coaching model that was used for the self-assessment tool called Wisconsin After School Continuous Improvement Program (WASCIP). In addition, the SEA staff reviewed the success stories that sub-grantees submitted to determine best practices to share with other sub-grantees at conferences.

Technical Assistance

Many SEAs used evaluation results for technical assistance purposes. Pennsylvania used evaluation results to follow up with sub-grantees that had challenges to determine if they needed support in addition to the scheduled monitoring/TA visits. In addition, Wisconsin reviewed the sub-grantee-level evaluations during the on-site monitoring visits to determine sub-grantees'

¹⁴ §4203(a)(13) A-B; Non-regulatory guidance H-6

specific areas of need or follow-up. More specifically, Arizona used evaluation results to design technical assistance activities and conduct informal surveys to determine what the sub-grantees' challenges were in obtaining positive outcomes and how the SEA could support them in addressing those challenges. Oregon used evaluation results to inform and revise the state technical assistance plan, monitoring process and future program evaluation. Utah's technical assistance provider used evaluation results to share key issues that arose during the quality assessment process with the SEA 21st CCLC Coordinator as input into the topics that were addressed in the monthly director's meeting. Several other SEAs also used evaluation results during conferences and presentations for dissemination and technical assistance purposes. Alabama shared the evaluation report findings with sub-grantees and community members at the Annual Alabama Community Education Association (ACEA) Conference in the opening session. Based on evaluation results, Illinois included some presentations on best practices in areas identified by the sub-grantees as needing more support in its 2009 fall kick-off meetings. Maryland presented evaluation findings to sub-grantees at a networking meeting, which was followed by a break-out session and discussion on how to use the evaluation findings. Idaho presented results of the evaluation to the Association of Idaho Cities at public safety conferences and other education related venues across the state.

Evaluation Improvement

A few SEAs used evaluation results for evaluation improvement. For example, BIE used evaluation results to refine their performance measures. Iowa used evaluation results to discuss future plans to do a more in-depth analysis. Plans include creating new charts and graphs so two years of data can be compared between state and sub-grantee data. Louisiana used evaluation results to pilot the use of pre-test/post-test assessment measures for afterschool programs to determine the impact on student academic outcomes and program effectiveness, and to design, develop and pilot a site observation tool that will identify a program's use of identified best practices. Oregon used evaluation results to support a future impact analysis based on comparing 21st CCLC program participants with non-participants.

Funding and Advocacy

Finally, a few SEAs used evaluation results to seek more funding and for program advocacy. Connecticut used evaluation results to create advocacy for additional funding. The state also added smaller networking opportunities for high and middle school programs based on evaluation materials and those programs looking for more support. Vermont used evaluation results to secure additional funding and partnerships through the use of data. New Hampshire used evaluation results when focused on advocacy with policy makers to increase resources for programming, and to attract additional funding and community partners.

Sub-Grantees' Use of Evaluation Results

This section describes the ways that sub-grantees used evaluation results, which include program improvement, discussions during conferences and meetings, evaluation improvement, and funding and program advocacy.

Program Improvement

The majority of sub-grantees primarily used statewide evaluation results for program improvement and program implementation. Examples included hiring more personnel to work one-on-one to improve students test scores, offering parent workshops on how to interpret test results and modifying programming. One sub-grantee added enrichment classes to the math and

reading program to make academics more interesting for students, incorporating music and cooking recipes. Another used evaluation results to change program structure such as the rotation of activities, including shortening some activities and making them more student-centered. A Florida sub-grantee used evaluation data to discuss program implementation issues at staff meetings. Another Florida sub-grantee used evaluation results to refine, improve and strengthen the program performance and delivery. For instance, the sub-grantee made minor adjustments to the center's educational offerings based on standardized test scores, and added summer services for students that were left with no resources during summer months based on student survey results. A New Mexico sub-grantee used evaluation results to identify its three weakest areas and develop action plans for continuous improvement efforts.

Conferences and Meetings

A few sub-grantees used evaluation results during conferences to make presentations and for technical assistance purposes. For example, a Kansas sub-grantee used recommendations from the evaluator in looking at how it could improve staff development and staff training. A New York sub-grantee used evaluation results to figure out training needs during the program advisory group meetings.

Evaluation Improvement

Some sub-grantees used evaluation results for evaluation improvement. A California CBO sub-grantee used evaluation results to refine evaluation measures. For instance, it added on-track high school graduation as a measure and then modified the homework measure based on responses. An Illinois sub-grantee used evaluation results for the development of program objectives and targets, and in designing the outcome evaluation. A Georgia sub-grantee used evaluation results to discuss the strengths and challenges of the previous year. Discussion topics included how to better obtain data on students and whether to refine one of the objectives. A Washington sub-grantee reviewed evaluation results to identify gaps in their assessments and to ensure they collected better data in the future.

Funding and Advocacy

Finally, a few sub-grantees use evaluation results for funding and program advocacy. A Kansas sub-grantee used evaluation results to secure additional funding through the city or through United Way. A New Hampshire sub-grantee used evaluation results to convince a local school board of the value of the program, leading to financial support and potential additional funding sources. Also, a New Jersey sub-grantee used evaluation results to apply for other grants and seek funding from foundations, private companies, social entrepreneurs and angel investors.

SEA Dissemination of Evaluation Results

This section discusses the ways in which SEAs disseminated evaluation results. Forums for dissemination include SEA websites, state conferences, communication with state and government agencies, and various training sessions.

Website

The majority of SEAs primarily disseminated their statewide evaluation results through the SEA websites and/or the states' 21st CCLC websites. For example, Florida posted statewide evaluations on its state program website and disseminated the report to a program listserv. Within their website, Hawaii offered sub-grantees a series of links to help them better understand the 21st CCLC programs. Links included statewide and sub-grantee evaluations.

State Conferences

Some SEAs disseminated their statewide evaluation results at conferences. Alaska disseminated evaluation information at the Alaska 21st CCLC Grant Director Meetings (which occur twice a year) and at the Annual State Conferences hosted for sub-grantees and others interested in afterschool programs. Illinois presented the 2009 statewide evaluation report findings at the three fall regional workshops. North Dakota disseminated its statewide evaluation at the summer conference to 21st CCLC sub-grantees. Executive summaries of the South Carolina statewide evaluations were presented at a program director meeting and the annual 21st CCLC Conference.

State and Government Agencies

SEAs also disseminated their statewide evaluation results to state and government agencies, including groups that work with the 21st CCLC program. For example, Alabama disseminated statewide evaluations to their advisory council, superintendents, federal program coordinators and other appropriate entities. Kentucky presented evaluation results to state education agency officials, at the statewide program directors' meeting and to the statewide afterschool network. Annual descriptive reports were given yearly to the California Department of Education to inform the governor and the legislature in making data-based decisions on policies and regulations affecting afterschool programs. Idaho made dissemination efforts and did evaluation presentations to other governmental groups, including mayors, police departments and LEAs.

Trainings

A few SEAs disseminated their statewide evaluation results to sub-grantees, often during 21st CCLC trainings and webinars. Kentucky shared evaluation results in PowerPoint trainings by providing examples of different methods of sharing data depending on the audience. Indiana shared statewide site visit results with sub-grantees via webinars. Topics covered included academic achievement, attendance data, student outcomes, program areas of strength and program areas in need of improvement. Georgia statewide evaluation results were presented to the board and discussed during sub-grantee trainings. Georgia also disseminated a PowerPoint presentation about the statewide evaluation for continuing sub-grantees. The first part of the presentation summarized statewide evaluation results from the previous year (2010-2011). Finally, Maryland disseminated the statewide evaluation results at a networking meeting for sub-grantee administrators and staff by asking the external evaluator to present evaluation findings, which was followed by a break-out session and discussions on how to use evaluation findings.

Most states made their statewide evaluation report available to the public upon request, with one exception: while BIE makes program evaluations available to the public upon request, notification of the existence and availability of program evaluations is typically limited to the tribal authorities, BIE Line Officials (Superintendents) and other BIE officials.

Sub-Grantees' Dissemination of Evaluation Results

This section describes methods by which sub-grantees disseminated evaluation results. Sub-grantees disseminated evaluation results to interested stakeholders commonly via websites, sub-grantee meetings and school newsletters.

Website

Several sub-grantees disseminated their evaluation results on their own websites and/or their SEA's 21st CCLC website. For example, a large sub-grantee in California posted the evaluation on the school district's website. A sub-grantee in Florida shared its evaluation report on the

SEA's website. The sub-grantee representatives reported that they appreciate this dissemination because it holds them accountable as a sub-grantee.

Meetings

Sub-grantees also disseminated evaluation results during a variety of meetings. An Alabama sub-grantee disseminated results through advisory meetings, teacher and parent meetings, and newsletters, fliers, and bulletins at schools and in the community. An Arizona sub-grantee's formative and summative project evaluation data and analysis were shared as an information item by the Superintendent during three School Board Meetings in 2008 and 2009. An Illinois sub-grantee disseminated its evaluation report in meetings with school principals and in meetings of site coordinators. Indiana presented results at one of the school board meetings each year and also provided them to the program's advisory group, which was made up of a classroom teacher, the Title I teacher and the family literacy coordinator for the school in question. A Pennsylvania sub-grantee provided evidence that the evaluation results were presented and discussed with a parent advisory board at one of their meetings. Pennsylvania evaluators also made a formal presentation of their findings at a school board meeting.

School Newsletters

A few sub-grantees disseminated evaluation results through school newsletters. Georgia sub-grantees disseminated evaluation results in the "Back-to-School" newsletter from the project director. On the first page, it described program success over the past few years and summarized some of the findings from the project's evaluation report. An Indiana sub-grantee summarized evaluation results in the school newsletter that went to parents and other key stakeholders.

Promising Practices of Use and Dissemination

Examples of a few use and dissemination practices classified as promising by the BPA team include:

- In Nebraska, a sub-grantee reported that the continuous improvement process (CIP) is the biggest strength of the 21st CCLC program because a local management team (made up of a principal, representatives from the initiative leadership council, the local evaluator and local funders) reviewed results on an ongoing basis. Sub-grantee staff reported that this allowed evaluation to be "infused into the program." The staff also noted that they shared a snapshot of these results with community partners.
- In Wisconsin, the SEA reported that the self-assessment process, an evaluation requirement for sub-grantees, has changed the way sub-grantees approach local evaluation. They believed sub-grantees viewed evaluations as a way to improve the program rather than a requirement to get continued funding.
- In West Virginia, sub-grantees used results from their CIPAS (which many considered to be their evaluations) to form peer learning groups. As a result of CIPAS reports, sub-grantees developed action plans focused on specific goals. The SEA then grouped sub-grantees into learning groups based on programs that had similar action plans and goals. The groups met regularly to share promising practices and resources, discuss challenges and lessons learned, and generally support each other in progressing toward their goals.

Summary

SEAs and sub-grantees were generally fulfilling the requirements to use evaluation for program improvement and to make evaluation results available to the public. Most programs used evaluation results for program improvement by identifying weaknesses or areas that needed more support and modifying program services or increasing support for staff training. Several SEA and sub-grantee programs also used evaluation results to modify and improve their evaluation approaches. SEAs disseminated evaluation reports in a variety of ways, including posting them on websites and sharing reports with sub-grantees. Sub-grantees used similar methods of dissemination and some programs also shared evaluation results through school newsletters. In this chapter, the BPA team also identified examples of promising practices with regard to evaluation use and dissemination. These practices involve the ongoing use and examination of evaluation results throughout program service delivery, and reflection on evaluation results through sub-grantee self-assessments.

VI. Evaluation Guidance and Technical Assistance Provided by SEAs to Sub-Grantees

This chapter describes evaluation guidance and technical assistance provided by SEAs to their sub-grantees. The first section describes different types of guidance and then summarizes information the BPA team has identified as particularly strong guidance. The next part of the chapter presents analyses of sub-grantee evaluation design related to whether SEAs provided strong guidance, and the size of states related to whether SEAs provided strong guidance. Finally, the chapter includes some discussion of changes in SEA evaluation guidance, support and requirements for sub-grantees over the years.

According to state and sub-grantee level data, much of the guidance and training regarding evaluation provided to sub-grantees by SEAs (or their contractors) included materials, data entry and compliance reporting, training on some aspect of evaluation (usually quality improvement processes or program self-assessments), and training delivered at conferences and meetings. About one third of SEAs provided strong guidance on comprehensive evaluation, including support for programs with regard to choosing an evaluator, developing measurable objectives, collecting and analyzing data, and including both process and outcome measures.

Materials

Several SEAs provided useful guidance materials to their sub-grantees, including various templates and sample measures. In Pennsylvania, the state-contracted evaluator provided a performance indicator “bank” to give sub-grantees examples of indicators they could use to evaluate their progress, and provided training on how to select evaluator during a yearly conference. Another state provided a sample evaluation plan, which consists of a table with columns indicating the objectives and sub-objectives, indicators of the objectives and targets for each school included. Sub-grantees in Colorado reported that materials distributed by the SEA were particularly useful in evaluating their programs. In addition to providing sub-grantees with orientation workshops and meetings that covered collection, reporting and utilization of program data in depth, the SEA gave each sub-grantee a large “evidence box” at the beginning of the grant for organizing the data they were required to collect; each box contained pre-marked folders and tabs corresponding to the data elements and supporting documents required for completing the MQI (monitoring and quality improvement tool). The Colorado SEA also provided a “Monitoring and Quality Improvement Tool Evidence Bank” of materials on its website.

Although SEAs provided materials, they did not always follow up and support use of the materials. While Pennsylvania provided a bank of indicators and other support, including tips on what qualities to look for in an evaluator and on cost, the sub-grantee evaluations were not reviewed for quality by the SEA or state evaluator. In another case, the SEA provided sample evaluation plans and documents describing how to conduct peer reviews; however, there was no evidence reflecting how these plans and documents were actually used by sub-grantees.

Instructions for Data Entry/Reporting

Common types of guidance provided to sub-grantees included instructions for data collection and mandatory reporting, such as PPICS data entry and data entry for statewide evaluation purposes. This guidance was more about standardized data collection for the state and federal government than data analysis for evaluation and program improvement for sub-grantees. Although many of

the same data collection and reporting tasks are needed for both evaluation and monitoring, oftentimes SEAs did not make the distinction between the two clear. Materials that SEAs considered evaluation guidance included a document explaining how sub-grantees should enter information into the Monitoring and Quality Improvement Tool (MQIT), an SEA monitoring tool, instructions on how to fill out the Evaluation Template for the SEA's web-based Evaluation Template and Reporting System (ETRS), and other state required data input procedures.

Several states also provided support to sub-grantees for the data collection and reporting that was the basis for the statewide evaluation. In New Hampshire, a PowerPoint presentation from the SEA and the contractor presented information on evaluation goals, parameters surrounding evaluation (e.g., research-based), a logic model displaying program inputs and outcomes, and data sources, but the presentation primarily focused on the responsibilities sub-grantees had for the statewide evaluation. Another similar PowerPoint presented information on how to complete a standardized template for sub-grantee evaluations, which was used to feed information into the statewide evaluation.

QIP Training

A common type of guidance SEAs provided to sub-grantees was training on a quality improvement process (QIP) which focused primarily on program implementation. In West Virginia and New Hampshire, the Continuous Improvement Process for After School (CIPAS) was used, and in West Virginia, thorough trainings were provided to all sub-grantees via webinar. Guidance on how to conduct CIPAS peer review visits were provided in New Hampshire, but no additional guidance on how sub-grantees should conduct program evaluation that utilizes outcome measures was included. Other states provided manuals, webinars and in-person training sessions on the QIP and using data for continuous improvement. Several states used the Youth Program Quality Assessment (YPQA) and included training on this tool, including how to conduct program observations. Oklahoma included sessions during their fall conference on tips, timelines, due dates, training opportunities and other supports to help smoothly implement the YPQA process. The Ohio Department of Education and Ohio State University presented a day-long training for all sub-grantees on using the Ohio Quality Assessment Rubric (O-QAR) for continuous program improvement. In Rhode Island, the SEA conducted an extensive technical assistance process called the Rhode Island Program Quality Assessment (RIPQA) that included site visits by contracted technical assistance providers who examined program operations in great detail in order to assess areas needing further training. In Wisconsin, sub-grantees could choose their own QIP/self assessment, but the state offered three annual trainings on the Wisconsin After-School Continuous Improvement Process (WASCIP) assessment.

Conferences/Meetings/Trainings

Other types of trainings and conferences were also a common method for SEAs to deliver support for evaluation to sub-grantees. In Kansas, the SEA hosts a conference in the fall during which the state evaluator trains sub-grantee evaluators. Sub-grantee representatives reported that the presenters do "a good job speaking to a variety of skill levels of evaluators". In Washington and Minnesota, networking meetings allowed sub-grantees to share promising practices and challenges related to evaluation. In Washington, evaluators from all sites were invited to attend the director's meeting; respondents say this networking "is the most important way of getting TA during meetings". In Delaware, a two-day workshop on evaluation was put on by the University

of Delaware. At state conferences in Indiana, evaluators attended sessions on how to conduct effective local evaluation and how to use these data to support program improvement efforts. Respondents reported that a significant amount of information is provided by the state regarding what an effective evaluator should be doing from a data collection and analysis standpoint. This is especially the case in trainings conducted by the state for first year grantees, where multiple examples are provided of what a good evaluation should look like.

Strong Guidance/ TA

The evaluation team determined that about one third (17) of all SEAs provided “strong guidance” on evaluation to their sub-grantees. Criteria for defining strong guidance include some combination of the following:

- Guidance goes beyond instructions for data entry (e.g., PPICS, reporting of attendance and activities, or other state/federal compliance);
- Guidance includes particularly useful materials such as an evaluation manual, sample logic models, sample RFPs for selecting evaluators and examples of measurable objectives;
- Guidance includes a process/mechanism for assessing the quality of sub-grantee evaluations and providing feedback to sub-grantees on their evaluation reports;
- Guidance includes individual TA to sub-grantees on evaluation; and/or
- Guidance includes a requirement that sub-grantees use an independent evaluator.

Strong guidance falls into the following broad categories: materials, trainings, individual feedback and TA, and requirements.

Materials

Several of these SEAs developed and disseminated materials and tools that support promising evaluation practices to their sub-grantees. In Maryland state evaluators developed a manual for sub-grantees that describes components of evaluation such as stating goals and objectives, options for evaluation design, and data collection and analysis. The Michigan Department of Education and Michigan State University produced a Local Evaluators Guide that provided guidance to local evaluators on federal 21st CCLC outcome measures to assist in structuring evaluation and program plans. This resource provides information on a variety of topics including hiring a local evaluator, program evaluation standards, a job description template for an evaluator, and tips for evaluators and project directors for working with a school system. Missouri provided a set of guidance materials that included report templates, survey instruments, a focus group guide for evaluator and sub-grantees, a self-assessment tool for sub-grantees, and afterschool standards and core competencies. In New York, an evaluation manual was provided to sub-grantees, including a logic model reflecting the goals, objectives and intended outcomes for the state 21st CCLC program, which guided programming and evaluation for sub-grantees. The Texas SEA provided sub-grantees with an Independent Evaluation Guide, which describes how to find and select an independent evaluator, guidelines for evaluation cost, the difference between basic and comprehensive evaluations, and suggested roles and responsibilities of evaluators. The Guide also provides report templates, recommended report elements and guidance about evaluation use. Another example of a useful tool is New Jersey’s modified web-based data collection system (developed by AIR – the SEA evaluator – to obtain center-level

information from the sub-grantees about the characteristics and performance of their programs), which allowed sub-grantees to work with interactive, site-specific report templates that are easily navigated and that provide one-click links to evaluation resources and tools.

TA/Trainings

Another element of strong guidance and technical assistance includes specific types of trainings and individualized TA for sub-grantees. In Maryland, training is provided each year on the evaluation manual that the SEA distributes to sub-grantees. Additionally, networking meetings that each sub-grantee attends include information on evaluation and data collection. In Massachusetts, trainings on the following topics are held annually for sub-grantees in relation to evaluation: an introduction to evaluation for new project directors and center coordinators; SPSS (data analysis software package) training for new project directors and center coordinators; optional supplemental trainings on topics like making use growth model data; and using Survey Monkey to create and administer surveys and longitudinal data analysis. The New Jersey SEA provides webinars, seminars, topical presentations and consultation from the New Jersey School-Age Care Coalition (NJSACC), their contracted TA provider, as well as training by outside consultants on action research plans. Several sub-grantees mentioned that the action research model was particularly useful in meeting program evaluation needs. In South Carolina, the state contractor (SWS) provides tailored and individualized evaluation TA. SWS staff walks program staff through the evaluation process and reviews each evaluation report submitted. SWS staff also helps sub-grantees with their continuation application if needed. Texas Education Agency (TEA) works with a technical assistance consultant to provide multiple trainings for evaluators throughout the state. TEA also brought in Priscilla Little (Harvard Family Research Project) for an orientation on evaluation.

Feedback/Individualized TA

One feature that stood out as representing particularly strong support for local program evaluation was the feedback that SEAs (or their contractors) provide to sub-grantees on their evaluation reports. In Minnesota, annual evaluation reports and the desktop and internal monitoring process provided multiple opportunities for feedback from the SEA on the appropriateness of the evaluation targets set by each grantee. The SEA indicated (and sub-grantees concurred) that helping reduce the “gotcha” mentality and fear of *not* meeting targets built a culture of continuous program improvement. In South Carolina, sub-grantees submit their evaluations to the external evaluator (SWS) who provides feedback and sends them back for revisions if needed. In some cases, sub-grantees needed to do the whole evaluation again, depending on the quality of their evaluations. In Texas, sub-grantees are starting to submit evaluation plans – and TEA is coming up with a checklist to ensure that evaluation plans comply with TEA requirements. TEA is working with an experienced TA consultant on reviewing evaluation plans. This provides quality assurance before programs even begin the evaluation process.

Requirements

Finally, the evaluation team identified several SEA requirements of sub-grantees that reflect strong practices in evaluation. New Jersey requires that sub-grantee evaluations be conducted by a local, external program evaluator, working in conjunction with program staff. As part of their application for new and continuation funding, sub-grantees are required to obtain and submit at least one quote for the external program evaluator with their application for funding. The quote must include information regarding the scope of the evaluation, deliverables and costs.

Minnesota requires that each grantee set their own targets on the following measures: student achievement in math and/or literacy via a measurement of their choosing (grades or test scores), recruitment and retention of students, and targets regarding student engagement. In addition, the SEA requires that each sub-grantee select a method of self assessment, but they do not need to report on the outcomes of the self-assessment. In New York as well, an external evaluator is required. In addition, the SEA requires (starting in 2012-2013) that the “evaluability” of each program be assessed before evaluation even begins. This is a multi-stage process that entails that local program evaluators meet with key stakeholders, visit sites to review early implementation, complete an evaluability checklist and submit it to the SEA. In completing the checklist, evaluators must make an overall determination of whether the program is sufficiently ready to be evaluated. In Texas, because the evaluations that TEA was receiving from sub-grantees varied greatly in quality, TEA decided to stop requiring evaluations for one year and to use that time to develop stronger guidance for sub-grantees; this was followed by a requirement that each sub-grantee submit a plan for evaluation that would be reviewed by TEA for quality.

Sub-Grantee Design in SEAs that Provided Strong Guidance

In theory, the sub-grantees from SEAs that provided particularly strong guidance and TA related to evaluation tended to use more rigorous evaluation designs than sub-grantees in states where the SEA provided little guidance. According to the Chi Square test, there is a significant association between these two variables. As shown in Table 18, 10% of sub-grantees that used rigorous outcome analyses came from SEAs that the team determined to be strong in guidance, compared with 4% of sub-grantees from SEAs that did not provide strong guidance.

Table 18. Sub-Grantee Design within SEAs that provide Strong Guidance**

Sub-Grantee's Evaluation Design	Strong Guidance/TA	
	Yes (N=17 SEAs)	No (N=37 SEAs)
Comparison group is rigorously formed	10%	4%
Comparison to district, state or national average	11%	24%
Single group, multiple points in time	46%	34%
Program comparisons made over time	13%	25%
Single group, single point in time	20%	12%
Total Sub-grantees	114	157
**Pearson Chi Square (4) = 18.5418 Pr = 0.001		

The BPA team also explored whether the SEAs that provided strong guidance tended to be the larger states, and were able to provide more support because they had more funding. Table 19 displays the relationship between SEAs that provided strong guidance and state size.

Table 19. SEAs that Provided Strong Guidance and SEA Size

Level of Guidance	Small	Medium	Large	TOTAL
Provided strong TA/guidance	4	8	6	18
Did not provide strong TA/guidance	20	12	4	36
TOTAL	24	20	10	54

The results show that for the large states, slightly more than half (6 out of 10) provided strong guidance. But for the small and medium states, the majority did not provide strong guidance. For the small states, the difference is large: four were determined to have provided strong guidance, but 20 were not.

Changes in Requirements and Guidance

Throughout interviews and discussions with SEA staff and evaluators over the course of this review, the BPA team gathered information about changes related to evaluation practices over time. This section describes changes specific to SEA evaluation requirements for sub-grantees, and changes related to evaluation guidance and technical assistance provided by SEAs to sub-grantees.

Requirements

States have used their experiences as 21st CCLC grantees to modify the evaluation requirements of sub-grantees. Though at the time of the evaluation interview Tennessee did not set required targets for outcome measures, the state would like to examine student data, and based on how the students are performing, set statewide targets for the following year. Maryland, as a result of an unavailability of data, did not set performance targets. However, with longitudinal data being collected from the preceding two years, the SEA may consider setting SEA performance targets for the evaluation following the monitoring visit in 2012.

Texas was the only state to defer sub-grantee evaluations. Because TEA was receiving sub-grantee evaluations that varied greatly in quality, the SEA decided to suspend the evaluation requirement for one year so that the state could put greater attention to providing guidance on evaluation. In the meantime, the SEA is requiring sub-grantees to submit evaluation plans, including RFPs for independent evaluators, performance indicators to be used or a description of data collection. The SEA is monitoring the progress of sub-grantees developing their evaluation plans and reviewing them upon final submission.

In California, conversely, the SEA recently began requiring all sub-grantees to conduct evaluations. With the most recent RFA, applicants were instructed to describe their plans for local evaluation, and sub-grantees were required to submit evaluation data to the SEA. The requirement for hiring an evaluator was already in place for programs serving high school students, and these sub-grantees can spend up to 6% of the grant amount towards the evaluation. Alaska changed its approach, requiring applicants to provide additional information in the grant proposal (e.g., define a set of measurable objectives or describe an evaluation plan for continuous improvement). The state will also require structured site visits and interviews with sub-grantee program directors, and will have sub-grantees focus on a set of 20 key quality indicators.

Technical Assistance and Guidance

One of the areas in which states reported making changes was the technical assistance and guidance they provided to sub-grantees regarding evaluation. One promising practice in regards to changes in guidance to sub-grantees was in the state of New York, where the SEA developed a comprehensive manual on conducting local evaluation for use in the 2012-2013 academic year. This manual addresses an issue the SEA identified among sub-grantees, which is that many sub-grantees are not clear on what to expect from their local evaluators – for example, the type and quality of work evaluators should do. The SEA felt that many local evaluators took advantage of this situation and were paid too much for very little or inadequate program evaluation services. This manual is prescriptive and outlines the requirements of the evaluation; for example, all sub-grantees must use the Afterschool Quality Self-Assessment as a part of the data collection.

Other instances of changes to technical assistance and guidance were found in states like Idaho and Oregon. Idaho informed the evaluation team that when they were in contract with Creating Changes, the state did provide technical assistance on data collection rather than topics specific to evaluation. The state will be providing more technical assistance on local evaluations. Oregon will also ask sub-grantees with exceptional evaluations to share at conferences and statewide meetings.

Summary

Although many SEAs provided some sort of guidance and technical assistance to sub-grantees regarding evaluation, much of this guidance was related to data entry for compliance and reporting purposes, rather than for comprehensive evaluation used for program improvement. In many cases, monitoring and evaluation overlapped when data collection for both activities are intertwined, making it difficult for sub-grantees to see how local program evaluation could produce meaningful results. Many states use a continuous improvement process (or quality improvement process/QIP) and provided thorough training on these assessments for sub-grantees. While these QIPs can serve as strong tools for assessment of program quality and implementation, they were often divorced from student outcomes. Several SEAs, in addition to data entry instructions and QIP trainings, provided a range of guidance and TA related to evaluation including evaluation manuals and materials, templates and sample measures, guidance on choosing an evaluator, logic models and trainings at conferences. The BPA team identified 17 SEAs that provided particularly strong evaluation guidance and TA. These SEAs were more likely to be larger states, but several small and medium-sized states also fell into this category. Finally, data analysis revealed that SEAs are strengthening guidance and requirements related to evaluation based on input from sub-grantees and lessons learned.

VII. Grantee Evaluation Needs and Challenges

During discussions with respondents and through the review of evaluation documents, SEAs and the BPA team noted a number of challenges that grantees, and in some cases their sub-grantees, faced in conducting program evaluation. Grantees also provided some valuable insights about how the Department might help them address these challenges through guidance and technical assistance. This chapter summarizes challenges, identified areas of needs for support and recommendations of delivering needed TA to SEAs.

Challenges in Conducting State Comprehensive Evaluations

The SEA evaluation challenges fall into three major areas: a perceived lack of clarity or specific guidance from the Department on what quality evaluations should look like; difficulties accessing high quality and complete data; and limitations in evaluation resources.

Guidance from the Department

Several states mentioned that the Evaluation Framework developed by BPA was useful in conducting evaluation and providing evaluation guidance to sub-grantees. For example, West Virginia used the information in their grant application instructions and used the webinar presentation as part of their training for sub-grantees. However, many states still cited a lack of consistent and specific guidance from the Department as a major challenge. Several states reported that they had not been asked to submit evaluation reports until about three years ago, and others reported that the requirements seem to be fluid and changing. At least one SEA respondent indicated that Department guidance on program evaluation has been unclear. Several states mentioned the need for some guidance or best practices on supporting sub-grantees in using evaluation results to inform program improvement.

Data Quality

The most frequently reported challenges fall within the context of limitations in the data being collected at the local level. These include concerns about quality and completeness, and access to the data needed to assess student outcomes.

Data Quality and Completeness

Some grantees, such as DC and Puerto Rico, reported difficulty getting data from all of their sub-grantees. Texas, West Virginia, Washington and Nevada all reported difficulty getting good response rates on surveys. Some states, such as Nevada, have responded to this by establishing response rate targets for their sub-grantees. Indiana found that only about three-fourths of the information requested actually gets into the evaluation reports, with missing information especially significant among the new grantees. Nebraska reported that their sub-grantees seem to need a lot of support and assistance with getting all of their data collected and entered into the data system. While one solution to data quality problems would be to ensure that sub-grantees all use well qualified evaluators, New Jersey pointed out that the SEA is legally prevented from telling sub-grantees that they should contract with a particular local evaluator that is known to the SEA for quality and cost-effectiveness in evaluation.

Data Consistency Across Sub-Grantees

Several states reported difficulty in obtaining consistent data across the state. For example, schools in South Carolina use different grading systems, and the most general of those is only a 3-point scale, which offers very little variation to track over time. The SEA also identified variation in discipline data due to differences in the way the numbers of actions are counted in

the new and old statewide data systems. In Wisconsin, sub-grantees have a choice of four different quality assessment tools, making it difficult to aggregate the data statewide. In Pennsylvania, sub-grantees report aggregate outcomes for students, rather than individual student results. Because each grantee established performance indicators in slightly different ways and are using various methods and instruments, grantees were allowed to report results in the general change categories, having freedom to define how change would be calculated. Minnesota is a local control state with independent school districts, so the SEA is limited in what it can ask local districts to do regarding evaluation. A major challenge is how to get 30 different stories into a statewide evaluation, maintaining the diverse richness of programming while still being able to tell a coherent story. BIE encounters political and jurisdictional issues with the tribally-operated schools that are quite significant and make it difficult to collect consistent data across sub-grantees.

Data Availability

In some cases, it is not the quality of data being reported by the sub-grantees that is the challenge, but rather access to SEA-level data. For example, Pennsylvania has not used comparison data because they have not had non-participant data available, but once the program starts using the SEA's standardized assessment database, they will be able to examine participant data compared with the state as a whole. Missouri pointed out the challenge of figuring out how to collect data from private schools. Delaware mentioned that they had tried to report "State Assessments Cross Year (Disaggregated)" measures, but had a problem with providers getting data from the state on time. In Connecticut, a major challenge was that student outcome data are not available in a realistic timeframe for affecting how programs operate the following year. The SEA has worked on getting their data system to be as useful as possible. The system is aimed at uploading information into PPICS, but the evaluators struggled with getting data to connect the process and outcome data. Washington found it a challenge that participant data is not available in real time.

PPICS

In addition to concerns about the quality of data being reported by sub-grantees, several states also mentioned challenges with the PPICS data system. For example, Florida pointed out that the PPICS is not tailored to the specific needs and terminology used in each state, so it is not user-friendly for sub-grantees. In fact, most states found they needed to provide training and guidance for their sub-grantees beyond that which is currently being provided by Learning Point/AIR. Another limitation mentioned by several grantees was that the PPICS does not provide individual level data. But the most frequently mentioned concern is the turnaround time between reporting the data and being able to get useable reports from the system. Wisconsin also reported they would like to see some updates to the PPICS so that it can better reflect their leading indicators.

Resources

Other frequently mentioned challenges to conducting program evaluation were those that are related to insufficient resources, both financial resources and the availability of expertise. For example, Arizona reported that they did not include quantitative analysis in their recent evaluation report because they simply lacked adequate staff time to produce the report. (Now that an evaluator has been assigned to this task, next year's evaluation report will include both qualitative and quantitative components.) BIE and Hawaii also cited their small staff as their biggest challenge in trying to address all the needs of their program, including program evaluation. California was concerned that the cost of evaluation is very high and it is difficult to

monitor the entire data collection process when it is being handled by an external evaluator. The SEA thought it might be more efficient to do it internally. Minnesota was concerned about the amount of staff time it takes for data collection at the local level. In New Mexico, the SEA Coordinator reported that state policy limits the funding for comprehensive statewide evaluation.

Some states reported challenges with some aspects of the evaluation itself, with limited expertise being the main concern. For example, Arkansas reported a challenge in getting an evaluation report that is actually usable at the state level, and figuring out what information they really need it to contain. California mentioned that using evaluation data to inform programming is a challenging new area for the SEA. Some states mentioned the need to find a new statewide evaluator. Several states, including Hawaii and Nebraska, reported that SEA staff who are not evaluators found it difficult to judge the quality of the evaluation work done by external evaluators.

Geography

Some SEAs reported that their biggest challenge in conducting statewide comprehensive evaluations is the geography of their target area, especially in terms of getting out to their sites to conduct on-site assessments. In Alaska, for example, not only are there large distances to travel to remote sites, but many are inaccessible during much of the year due to the weather. The evaluator selected a sample of sites for evaluation site visits based on their need for technical assistance, and both the SEA and the evaluator are trying to address this challenge through the creative use of technology. BIE struggled with trying to evaluate programs across 14 different states. Nevada and Texas also mentioned the challenge of the evaluator reaching rural sites, and in North Dakota, weather is also a major barrier. North Dakota also pointed out the difficulty of locating and retaining quality staff in rural areas.

Across all of these different types of challenges, a common theme was the desire to learn from other SEAs, not only in terms of promising practices in evaluation methods, but also in terms of some practical ways to address specific challenges. Many SEAs have asked for examples of what evaluation reports should look like and for more structured opportunities to exchange information and ideas with other SEAs.

SEA-Identified Technical Assistance and Support Needs

Mirroring many of the challenges described above, SEAs also described areas where they could use additional technical assistance and support in addressing the program evaluation requirements.

Department Expectations

Several SEAs have specifically reported on the lack of clear guidance and expectations from the Department with regards to evaluation. Aside from the general requirement to conduct evaluation, very little detail is provided to grantees about evaluation planning, which evaluators to work with, evaluation design, minimum requirements, or requirements for sub-grantees. A range of grantees have specifically indicated the need for such guidance, including SEAs that have put fairly extensive evaluation plans into place and those that have not. While many SEAs rely primarily on PPICS data and some kind of self-assessment of their service process, some have moved forward with bringing in qualified evaluators and conducting comprehensive evaluations. Some evaluations, such as those in California and Texas, involved a wide range of data collection activities. For example, an SEA representative from North Carolina emphasized

the importance of knowing what evaluation activities they are expected to engage in and what dimensions of program quality should be assessed among sub-grantees.

Recruiting and Selecting a Qualified Evaluator

As mentioned earlier, a number of SEAs are in the process of seeking a high-quality independent evaluator that can meet the programs' needs. Some are also concerned about helping sub-grantees make a good choice of local evaluator. SEAs identified three different needs when discussing the challenges associated with locating and selecting a qualified evaluator:

- **Criteria for selecting an evaluator.** Several respondents mentioned the need for more information about how to select an evaluator who will meet their needs. While criteria such as training and experience are helpful, it seems that some SEAs could benefit from additional ideas about things to look for when reviewing resumes and conducting interviews, or the kinds of criteria to use when reviewing proposals to identify the evaluator most likely to work with the SEA to ensure that the evaluation meets the agency's need.
- **Developing an evaluation RFP.** SEAs could benefit from an example of a strong evaluation RFP, or at least some guidance on what should be included in an RFP for a competitive bid process for selecting an evaluator, either because they had not selected one competitively before, or because their previous RFPs had lacked specificity. In particular, several respondents were concerned about how much of the evaluation design to lay out in their RFP and how much to expect the potential evaluators to specify in their proposals.
- **Potential sources of qualified evaluators.** Several respondents mentioned that they need to go beyond the evaluators used by the SEA in the past or their local university to look for highly qualified evaluators, but they were not sure where to look and could use some suggestions from the Department.

Performance Measures

SEA staff have varying degrees of experience in developing measurable objectives and performance measures. As one SEA Coordinator put it, "We know what the GPRA measures are and we know we have to provide some information on how programs are stacking up, but we don't know what would be good objectives for sub-grantees to strive toward. It's hard to know what reasonable targets are for sub-grantees." In some cases SEAs had established performance measures to which it was difficult to hold the program accountable. For example, in Hawaii, one of the performance measures was the percentage of schools statewide achieving AYP under No Child Left Behind; however, the 21st CCLC program can really only directly affect the students receiving 21st CCLC-funded services, rather than the state school system as a whole. Many SEAs had performance measures such as "improving student behavior" that did not include specific targets, so they were unable to assess whether these had been achieved or not. The area of developing realistic and meaningful performance measures was one where some SEAs felt it would be valuable to learn from other grantees, either by having the Department compile examples or structure opportunities for the SEAs to share current practice with each other.

Recommendations for Methods of Delivering Technical Assistance to SEAs

In the course of discussing their needs and challenges, grantees made numerous suggestions for the kinds of assistance they would find helpful and methods that the Department might use to address their technical assistance needs.

Specification of Requirements

Several SEAs suggested that written guidance would be extremely useful. Not only would written guidance provide a resource for reference during the evaluation planning process, but it would also be extremely valuable in cases where there is turnover of SEA staff. With written guidance as a reference, a new SEA coordinator would be able to easily see what is required, rather than relying on periodic technical assistance that may or may not be timely during turnover. For example, Nevada suggested that helpful resources could include a guidebook, a manual or some type of legislation that provides more specific guidance on what is expected of a statewide evaluation.

TA Appropriate to Different Levels of Need

A common suggestion among SEA respondents was that the Department look for ways to develop technical assistance and materials that recognize the wide variation in levels of expertise among potential TA recipients, as well as the different stages of development grantees are going through as their evaluation efforts evolve over time. For example, Illinois suggested that TA could be tiered according to universal, targeted and intensive needs.

On-Site Technical Assistance

Several grantees indicated that they would find it useful to have time for specific technical assistance on program evaluation that would go beyond the time available during a monitoring visit. This could include meetings with the SEA coordinator and evaluator to review and discuss current evaluation plans, review of an SEA's evaluation RFP, or even assistance with reviewing and rating proposals from potential evaluators, advice on how to link process and outcome data, and reviewing data to help with interpretation of results. While some grantees saw this type of TA as being specific to evaluation, others saw advantages to combining evaluation TA with consultation on program implementation in general. For example, respondents from the Virgin Islands reported that at the beginning of their grant period, it would have been helpful to have a Department representative conduct a visit to their site to review all aspects of the grant, including evaluation.

Webinars

Several SEAs reported that the Department's online evaluation training (presented with BPA) for state grantees was helpful. For example, Georgia reported that BPA's webinar and evaluation framework was extremely useful in that the SEA pulled out sections of the Framework and put them in the RFP to provide guidance to sub-grantees about aspects of evaluation, such as how to create measurable objectives. West Virginia used the PowerPoint from the webinar in their own training as well. While the Evaluation Framework webinar was seen as beneficial, this particular webinar was very general and served as an introductory session. A number of grantees have indicated that they would like to see more webinars that go into more depth, both for evaluation novices and for more experienced SEA coordinators and evaluators. In addition to providing a step-by-step process for developing evaluation plans, webinars could also be used as a vehicle for idea exchange between grantees who are already implementing comprehensive evaluations to discuss enhancements to methodologies and effective use of results.

Summer Institutes

SEA staff who attended the sessions on evaluation at the Summer Institutes (in 2011 and 2012) noted that these sessions were useful, as did many sub-grantees. Attendees indicated that it would be useful to offer a wide variety of sessions on program evaluation and that they would appreciate structured opportunities to exchange ideas and discuss practical solutions to challenges.

Networking/Sharing with other Grantees/Peer Support

In addition to more specific guidance from the Department, many SEAs expressed the need for more networking and sharing among grantees. For example, Arkansas mentioned that the SEA meetings are very beneficial, especially for those who do not know a lot about evaluation and are not used to operating large programs. There was strong interest among many states, such as Minnesota and Indiana, in seeing what effective evaluation looks like in other states. Some grantees reported they could benefit from networking with other states at a national conference dedicated specifically to evaluation. Kentucky was one of many states that reported that it would be helpful to have a structured process for sharing lessons learned about program evaluation and what works well in the field. Pennsylvania mentioned that open collaboration among grantees can be very helpful. West Virginia participates in and co-hosts a multi-state afterschool conference with neighboring states. The Department could host or support regional conferences of this type with a particular focus on evaluation.

Providing Support to Sub-Grantees

SEAs have expressed some challenges with supporting evaluation practices among sub-grantees and communicating to sub-grantees the enormous potential that evaluation can have for program improvement. For example, Idaho revealed that sub-grantees are anxious about the evaluation activities planned for the upcoming year and are worried that the evaluation will result in punitive action. Sub-grantees in Oregon expressed similar concerns about how evaluation results will be used. Oregon's SEA Coordinator, therefore, reported a need to make clearer the distinction between evaluation and monitoring in order to abate their fears. SEAs expressed the desire for guidance from the Department on how to convey these distinctions to sub-grantees.

In addition, SEAs could use advice from the Department about recommending evaluators to sub-grantees. Oregon does not require sub-grantees to hire external evaluators, because they feel the evaluators' interests may not always align with the interests of the program providers. Therefore, the state would also like guidance on how to shape local evaluations to capture what is actually needed for program improvement and sustainability. A couple of SEAs also mentioned that they are not allowed to recommend or require that sub-grantees contract with specific evaluators and could use some guidance on how to provide support to sub-grantees without violating any policies.

Summary

This chapter has summarized the types of challenges SEAs encountered in carrying out effective evaluations. It has also presented specific areas of need and potential TA opportunities that the Department could provide to bolster support to grantees and sub-grantees regarding program evaluation. The next chapter presents implications for the Department based on all of the data collected and analyzed in this report, and specific recommendations from BPA to the Department regarding evaluation guidance and TA for SEAs.

VIII. Implications for Department of Education: Recommendations for Evaluation TA and Guidance to SEAs

Based on the information gathered from the states and sub-grantees reviewed and the challenges discussed above, several areas of technical assistance (TA) and evaluation guidance for states and sub-grantees have been identified. BPA has used the information gathered about grantee and sub-grantee evaluation practices over the past three years to develop a basic framework for program evaluation that operationalizes the federal requirements and describes five key features of comprehensive evaluation (see Appendix H: *Evaluation Framework for 21st CCLC Programs*). In this section, recommendations for providing TA are organized in terms of these key features of the Framework.

In 2010's Interim Report, BPA recommended that the Department provide specific expectations and criteria for effective evaluation and professional development in program evaluation. BPA became a part of these efforts during 2011 in assisting the Department with developing expectations for program evaluations, and by providing TA and support through dissemination of the Framework, through presentations at the 2011 and 2012 21st CCLC Summer Institutes, and in providing TA to selected states and wider audiences via webinars (hosted by the Department).

Selecting a Qualified Evaluator

Because many SEA 21st CCLC coordinators and sub-grantee program directors lack a strong background in program evaluation, it can be challenging for them to identify well-qualified evaluators and successfully manage their evaluation efforts. Even SEAs with a good understanding of program evaluation have found that evaluation contractors do not necessarily provide sub-grantees with cost-effective services. The Framework provides suggestions of credentials and qualifications to look for in evaluators that address both methodological and content expertise (such as a Master's degree or Ph.D. in education or a social science discipline, training in rigorous evaluation design and using relevant qualitative and quantitative methodologies, experience evaluating other 21st CCLC programs or other school or community programs aimed at increasing student academic achievement, and experience collecting and analyzing student outcome data).

Additional assistance that may benefit grantees and sub-grantees includes suggestions of where to look for evaluators and how to select a qualified trustworthy evaluator. For example, the Department could provide a suggested list of resources such as local universities and local evaluation agencies (some can be found through the American Evaluation Association website), and also connect grantees and sub-grantees with their peers to share ideas. Sub-grantees could also benefit from guidelines related to cost and budgets for local program evaluations. Most of the evaluation contracts BPA reviewed provided little detail about the specific services and cost breakdown for those services, and many sub-grantees' evaluation contracts did not have a clear scope of work that explicitly stated expectations of evaluators, timelines and deliverables. Clear expectations of evaluation services will help sub-grantees budget effectively for evaluation and ensure that limited funds are used efficiently.

Articulating Program Goals and Measurable Objectives

Many of the evaluation reports reviewed did not include well-articulated program goals and measurable objectives. In the Framework, BPA provides examples of both goals and objectives and explains the importance of each. The Department should encourage articulation of

measurable objectives so that programs know when goals and objectives are met. Most of the SEAs required at least one or two types of student academic outcome measures be used in sub-grantee evaluations. However, beyond this requirement, SEAs (or their contractors) often did not provide specific guidance to their sub-grantees on the measures and outcomes they should examine as part of their local program evaluations. SEAs and sub-grantees could also benefit from the use of logic models (or a template/sample of a logic model) that shows programs how goals should link to outcomes, and how and why the data collected as part of the evaluation will feed back into program improvement. Sub-grantee evaluation quality and efforts varied widely within states, which in turn made it more challenging to evaluate the 21st CCLC programs at the state level. Clearly defined goals and objectives will help strengthen statewide and individual sub-grantee evaluation efforts.

Using Designs Appropriate for Measuring Program Quality and Effectiveness

Evaluation design incorporates evaluation questions or objectives, process and outcome measures, rigorous analysis, stakeholder representation, proper documentation, data management and ethical standards. These elements are described in the Framework, and grantees should ensure that their evaluators are familiar with each of them and understand their importance. The Department should encourage SEAs and sub-grantees to examine both process and outcome data as part of their evaluations, in order to understand how implementation of programs may affect student outcomes, and in order to assess what program changes could bring about improvement in outcomes. Evaluations should include attention to sampling, comparison groups (counterfactuals), and the linking of measures and outcomes to program goals.

SEAs might also benefit from technical assistance in the effective use and management of statewide student-level data for both program participants and non-participants. It seems that most SEAs and sub-grantees would benefit from efficient management of student assessment data tracked over time. While access to student data can sometimes be a challenge, many schools, districts and SEAs have a great deal of data available to their evaluators and to 21st CCLC program personnel (especially school and district personnel). Evaluators and program staff could benefit from guidance about how to utilize these data most effectively. It is important to manage and track participant data at the student level so that changes in academic performance, school attendance and other indicators of interest can be monitored longitudinally. It can also be advantageous and strengthen evaluation rigor to include non-participant data for comparison purposes. This information can then feed into the evaluation of the programs at both the sub-grantee and statewide levels. SEAs should ensure they have unique student-level identifiers and are able to match data across program years to measure individual student change over time.

Analysis and Reporting

One of the most concrete recommendations the Department can make to grantees is that they analyze their outcome data rather than merely report descriptive information. Many evaluation reports reviewed by the BPA team used fairly weak designs and did not typically conduct analyses to determine program effects. Many reports presented PPICS data to make generalizations about the 21st CCLC programs. Those that included process data generally did not link the process and outcome data to understand how implementation of programs may affect student outcomes, or to assess what program changes could bring about improvement in outcomes. Often recommendations were made based on the evaluator's suggestions rather the

data. In addition, far too many evaluations of implementation or process relied on anecdotal observations made by a few stakeholders (usually program staff) rather than a systematic collection and analysis of implementation data.

It is also important to ensure that SEAs and sub-grantees thoroughly describe all pertinent information in their evaluation reports. This information may include: description of their sample and/or program participants, description of evaluation participants (teachers who were surveyed), response rates, data collection procedures and analysis methods. Only by including documentation of the evaluation methods is it possible to assess the quality of the data and the evaluation effort.

Use of Evaluation Results

Grantees and sub-grantees should be encouraged to use evaluation results in a meaningful way and disseminate evaluation results widely among stakeholders. Evaluations can be used for a variety of purposes including: to provide evidence of program effectiveness and procure additional funding; to market and publicize a positive program throughout the community and increase community participation; to modify programs by increasing successful activities and eliminating ineffective activities; to modify RFPs; and to document successes, challenges and progress toward long-term goals. BPA's analysis of evaluation practices showed that grantees and sub-grantees used evaluations for program improvement to some extent. However, given that evaluations were often based on anecdotal evidence rather than systematic data analysis, this is an area that requires more attention.

Providing concrete examples of effective evaluation practices to program personnel can be an effective way of relaying the importance of these elements. Sample evaluation reports and plans can be helpful, bearing in mind that evaluations of individual programs and sub-grantees will vary based on each program's unique objectives, capacity and context. It is important that program operators understand evaluation basics even (or especially) if they hire an external evaluator, so that the quality of the work of the hired contractor can be monitored and most appropriately reflect the 21st CCLC program under study.

BPA developed the Evaluation Framework (Appendix H), which highlights the five key features discussed above, to address these concerns and assist programs and evaluators in planning and implementing evaluations. The Framework was created with both program staff and evaluators in mind as the audience. However, the Framework is quite general, and while it lays out a structure that can be used to build stronger evaluations, it does not provide very much detail. Provision of technical assistance materials, state-specific consultation, and convening program evaluators to discuss evaluation concepts and share promising practices would all be valuable investments in the improvement of the overall quality of evaluation efforts.

IX. Summary and Conclusions

This final report, *Review of 21st Century Community Learning Centers' Grantee Evaluation Practices*, presents the culmination of three years of data collection and analysis by the BPA research team. The team has reviewed thousands of evaluation related documents and conducted hundreds of interviews, both on-site during monitoring visits and via telephone. These data represent information from all 54 SEAs and a representative sample of their sub-grantees. The BPA team, in working with the Department of Education on this important endeavor, has learned a great deal about the types of evaluation approaches that 21st CCLC programs are using and the areas of support that can be addressed through additional technical assistance and guidance. This chapter presents a summary of these evaluation approaches, key evaluation findings, guidance provided by SEAs to sub-grantees, the ways in which SEAs and sub-grantees use and disseminate evaluation results, the evaluation challenges programs are experiencing, and BPA's recommendations for the Department regarding evaluation support to grantees.

The key questions guiding this evaluation review were:

1. To what extent are SEAs and sub-grantees conducting evaluations of their 21st CCLC programs?
2. What is the range of approaches that are being used to evaluate 21st CCLC programs?
3. What research designs do programs use to evaluate student outcomes?
 - a. How does research design vary by selected SEA characteristics?
 - b. How have evaluation practices been changing over time?
 - c. What are promising practices in evaluation among 21st CCLC grantees?
4. What are the key findings from 21st CCLC program evaluations?
5. What guidance and technical assistance are SEAs providing to sub-grantees regarding program evaluation and what topics do they cover?
6. How do programs use evaluation and disseminate evaluation results?
7. What information or support do grantees need to conduct more rigorous evaluations?

A brief summary of the results of the review in addressing each of the questions is presented below.

Extent to Which Grantees and Sub-Grantees Conduct Evaluation

Over the three years, documents were collected from 54 SEAs and a sample of their sub-grantees. Although 48 (out of 54) SEAs submitted evaluation reports, the BPA team gathered information through interviews suggesting that *all* SEAs are engaged in some sort of statewide evaluation activity. Some are in the process of developing evaluation designs, changing evaluation plans or evaluators, or having their reports written or revised, and therefore did not have reports available for submission at the time of data collection.

On the other hand, the team found that many sub-grantees did not conduct local program evaluations. Evaluation documents were requested from a total sample of 595 sub-grantees (about 20 percent of the sub-grantees in the selected states), representing about 20 percent of the active centers in those states. Although 78 percent of the sample submitted evaluation reports,

these documents did not all represent formal evaluations. Some of these documents included monitoring reports, raw data in the form of tables or lists, self-assessments, peer observations and perceptions of program effectiveness based on anecdotes and program staff members' opinions.

Evaluation Approaches

Frequency and Type of Evaluator

Most SEAs and sub-grantees conduct program evaluations annually, with some sub-grantees required to submit multiple progress reports per year. SEAs and sub-grantees varied widely with regard to program evaluation approaches. Most SEAs contracted with research agencies or academic institutions to evaluate their statewide programs. Almost a quarter of sub-grantees used research agencies, while 22 percent used their own program staff to evaluate their programs. The evaluator that a program works with can present benefits as well as challenges. Several program staff praised their evaluators for their expertise, good collaboration skills and ability to translate data into useable information for program improvement. Some SEAs and program directors experienced challenges with evaluators who they perceived to be highly costly in comparison to the value of the services received.

Measures

Programs used a variety of measures to assess program quality and outcomes. The most common measures of student outcomes were state standardized assessments in math and reading/English Language Arts. Many programs use these data because they also fulfill a requirement of the Government Performance and Results Act (GPRA) measures that states must report to the federal government. Other commonly used measures included school or classroom grades and the teacher survey designed by AIR that assesses changes in student behavior and academic performance. Non-academic outcome measures included behavior, homework completion, school attendance and disciplinary incidents. In examining program implementation and quality, many SEAs required their sub-grantees to participate in a quality improvement process (QIP) or continuous improvement process (CIP) (e.g., YPQA). In some states, the QIP serves as the primary evaluation component for sub-grantees. The most common data sources for assessing process measures were PPICS or other types of program records. In some cases, student, parent and staff satisfaction surveys were also used to examine program quality.

Research Designs

Evaluators used a range of research designs to examine student outcomes. The most common research designs among SEAs and sub-grantees were "single group, multiple points in time" and "program comparisons made over time". In the first case, programs examined student outcomes prior to program implementation and then at one or more times after implementation began. This method required access to data at the individual student level, so that progress for individual students could be measured and then aggregated across the program. For program comparisons made over time, evaluators examined student data for different cohorts (different groups of students) over time; this approach serves to compare the program as a whole rather than the progress of specific students. Overall, sub-grantee evaluations were much weaker and less rigorous than state level evaluations. In addition, sub-grantee designs and evaluation reports varied greatly within and between states.

The BPA team found significant relationships between sub-grantee design and selected SEA characteristics. State size appears to be related to sub-grantee design, in that medium-sized states

seemed to have the most sub-grantees using the most rigorous design (using a comparison group that is rigorously formed using matching techniques). In addition, SEAs that used the most rigorous design were most likely to have sub-grantees that used this design as well. In another analysis, the team found that sub-grantees in states that required specific targets be used in sub-grantee evaluations were somewhat more likely to use rigorous designs (7%) compared to those in states that did not require targets, which did not include any sub-grantees using rigorous designs.

Promising Practices

The evaluation team observed promising practices in design, data collection, evaluation procedures, reporting and evaluator expertise at the state and sub-grantee level that may benefit the evaluation and implementation of the 21st CCLC program. Some SEAs and sub-grantees revealed a more rigorous evaluation design, such as conducting a longitudinal study of students' academic performance, or implementing a quasi-experimental assessment between program participants' academic outcomes compared to those of non-program participants. Several SEAs and sub-grantees also developed personalized data collection tools to fit the needs of specific programs. Other promising practices observed among both SEAs and sub-grantees included improved procedures for data collection, including innovative methods to increase survey response rates or obtain student-level data from schools. Several SEAs and sub-grantees also reported confidence in and satisfaction with their evaluators. The research team also identified several instances of high-quality evaluation reporting among SEAs and sub-grantees that provided comprehensive and useful information about evaluation goals, data collection procedures, evaluation methodology, recommendations or use of evaluation results. Finally, the research team identified a number of examples of well-developed in-depth guidance materials used by SEAs to promote high quality evaluation efforts among their sub-grantees.

Findings of State and Local Program Evaluations

State-level evaluation reports generally presented more robust findings than the evaluation reports of sub-grantees. State-level evaluations focused on student outcomes (particularly academic outcomes) provide context and support findings with evidence more so than sub-grantee evaluations. Overall, the findings from states and sub-grantees were largely positive, although there were cases where programs did not meet their objectives. There were also sometimes inconsistent findings observed among sub-grantees within a state. This could be attributed to various factors, such as some sub-grantees having more effective programs than others or simply having constructed a better evaluation approach.

SEA Guidance and Technical Assistance to Sub-Grantees

The guidance and training regarding evaluation provided to sub-grantees by SEAs (or their contractors) included materials, data entry and compliance reporting, training on some aspect of evaluation (usually a quality improvement processes or program self-assessments), and training delivered at conferences and meetings. Although most SEAs provided some sort of guidance and technical assistance to sub-grantees regarding evaluation, much of this guidance is limited to data entry for compliance and reporting purposes, rather than focusing on program evaluation designed to support program improvement. In many cases, monitoring and evaluation overlapped when data collection for both activities were intertwined, making it difficult for sub-grantees to see how local program evaluation can produce meaningful results.

Many states used a continuous quality improvement process (QIP) and provided thorough training on these assessments for sub-grantees. While these QIPs can serve as strong tools for assessment of program quality and implementation, they were often divorced from student outcomes.

About one third of SEAs provided comprehensive and high quality guidance on evaluation, including support for programs with regard to: choosing an evaluator, developing measurable objectives, collecting and analyzing data, and including both process and outcome measures. They have developed evaluation manuals and materials, templates and sample measures, sample logic models and training materials for statewide evaluation trainings. The BPA team identified 17 SEAs it considers to have provided particularly strong evaluation guidance and TA. These SEAs were more likely to be larger states, but several small and medium sized states also fell into this category. SEAs are also strengthening guidance and requirements related to evaluation based on input from sub-grantees and lessons learned.

Use and Dissemination of Evaluation Results

SEAs and sub-grantees for the most part are fulfilling the requirements to use evaluation results towards program improvement, as well as making the evaluation results available to the public. Some respondents noted the perception that program evaluation has positively shifted from being seen as merely a requirement to receive grant funding to being considered a tool to make program improvements. The evaluation team identified common ways in which SEAs and sub-grantees used evaluation results. SEAs and sub-grantees both reported similar uses of evaluation results, including identifying technical assistance needs, improving evaluation methods, garnering funding or advocacy, and program improvement in general. The evaluation team also identified various methods of disseminating evaluation results among SEAs and sub-grantees. SEAs disseminated the evaluation results primarily through state websites, at state conferences, to other state and government agencies, to sub-grantees during trainings, as well as to the public upon request. Sub-grantees tended to concentrate dissemination efforts of evaluation results on a small number of target audiences by posting evaluation results on the sub-grantee's website or on their SEA's website, sharing findings at meetings with key stakeholders, and sharing key results through school newsletters.

Challenges and Needs

SEAs and sub-grantees reported both challenges and needs in regards to conducting program evaluation. One challenge that filters from the SEA level down to the sub-grantee level is the overlap in many states between monitoring, quality assessments and evaluation, and lack of clarity about the definition of program evaluation. For example, per some SEA requirements, some sub-grantees equated uploading program data to conducting evaluation. Another common challenge among SEAs involved data collection; SEAs noted that it was difficult to get complete and quality data as well as reconcile sub-grantee-level data that are collected through different tools or methods. In addition to challenges in collecting data, SEAs identified a number of limitations in the PPICS data system, including a significantly delayed turnaround time for getting reports, the lack of student-level data, it being generally not user-friendly, and the fact that it is not tailored to the specific needs or terminology used in each state. SEAs also noted the general lack of resources available when conducting an evaluation, including being limited by the geographic spread of program sites, as well as a lack of expertise in evaluation. Finally, another frequently reported challenge was the Department's lack of consistent and specific

guidance, with some SEAs noting that reporting requirements seem to be fluid and changing. To address these challenges and concerns, SEAs identified key areas where they would like more guidance from the Department: sources for qualified evaluators, selecting a qualified evaluator, developing an evaluation RFP, developing performance measures and clear expectations from the Department regarding evaluation.

BPA Recommendations

BPA makes the following recommendations to the Department regarding evaluation technical assistance and guidance, based on the cumulative review of SEA and sub-grantee evaluations from 2010-2012:

1. Provide clear expectations for what is required of SEA and sub-grantee evaluations. This could include some minimum requirements for evaluation (the BPA Framework could be used as a starting point), and some clarity around the distinction between monitoring and evaluation. This guidance could be incorporated into the Non-Regulatory Guidance that grantees currently receive.
2. Provide more guidance on selecting a qualified evaluator. This includes what constitutes a quality evaluator, advice about where to find potential evaluators, how to conduct an effective competitive contracting process (including developing a strong RFP and good criteria for selection), and parameters on how much to spend on evaluation services.
3. Encourage SEAs and sub-grantees to work with evaluators who can increase the rigor of their evaluation research designs. Solid evaluation designs begin with clearly articulated program goals and measurable objectives and include specific evaluation questions or objectives, assessing both outcome and process measures, managing data, rigorous analysis, involving stakeholder representation, and proper reporting and documentation.
4. BPA encourages the Department to promote a greater investment in evaluation design and analysis to provide evaluation results that can be used by SEAs and sub-grantees in a meaningful way.
5. Finally, BPA encourages the Department to provide SEAs with networking and professional development opportunities focused on evaluation. Suggested venues for sharing promising practices and resources include webinars, regional conferences, national evaluation conferences and on-site evaluation technical assistance visits.

References

- Government Accountability Office (GAO). (2009). *Program Evaluation: A Variety of Rigorous Methods Can Help Identify Effective Interventions. Report to Congressional Requesters.* GAO-10-30. Washington, DC.
- Joint Committee on Standards for Educational Evaluation. (2011). *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users.* 3rd Ed. Yarbrough, D.B., Shulha, L.M., Hopson, R.K., & Caruthers, F.A. (Eds). Thousand Oaks , CA : Sage Publications.
- McDavid, J. & Hawthorn, L. (2006). *Program Evaluation and Performance Measurement.* Thousand Oaks , CA : Sage Publications.

Appendices

- Appendix A: Detailed Evaluation Review Methods and Procedures
- Appendix B: Analysis Plan
- Appendix C: Evaluation Review Documents Requested
- Appendix D: Interview Guide
- Appendix E: Codebook for Reviewing Documents
- Appendix F: Document Review Instructions and Guidance to Reviewers
- Appendix G: Detailed Information on 21st CCLC Evaluation Measures Used by SEAs and Sub-Grantees
- Appendix H: Evaluation Framework for 21st CCLC Programs

Appendix A

Detailed Evaluation Review Methods and Procedures

Appendix A Detailed Methodology

This Appendix details the methods and procedures used for the Review of 21st CCLC Grantee Evaluation Practices. The first three sections describe the SEA selection methods and identify the SEAs undergoing evaluation reviews in each year of the study. The fourth section describes the sub-grantee sampling method. The last several sections describe the document review process with detailed document information, the interview process and the desk monitoring calls, and include a summary table of data collected for each SEA over the three years of this review.

Selection of SEAs for Review during 2009-2010

During 2009-2010, BPA selected a representative random sample of 18 states stratified by program size and federal region. Program size was determined according to the following criteria:

1. Small states - receiving less than \$1 million in 21st CCLC grant funds in 2009.
2. Medium states - receiving \$1 million to \$10 million in 21st CCLC grant funds in 2009.
3. Large states - receiving over \$10 million in 21st CCLC grant funds in 2009.

Federal regions were chosen to represent geographic distribution because they are familiar and well understood by administrators and policy-makers, and because the broader regions often used in describing the U.S. (Northeast, Southeast, Midwest, Southwest, and West) may be too large to capture regional variation in state-level policies regarding evaluation practices.¹

The resulting sample of states reviewed included 11 large states, 6 smaller states and one special jurisdiction, evenly spread across the federal regions:

Arizona	Louisiana	New Hampshire	Puerto Rico
Colorado	Massachusetts	New Mexico	South Dakota
Delaware	Missouri	New York	Texas
Idaho	Mississippi	Oregon	
Illinois	Nebraska	Pennsylvania	

Selection of SEAs for Review during 2010-2011

For 2011, the Department and the BPA project team decided to select the Evaluation Review sample using a different method. States that were monitored in either 2010 or 2011 *and* had not been part of the evaluation review sample in 2010 were automatically included in the evaluation review sample for 2011. The resulting sample for 2011 included the following 20 states and special jurisdictions:

Alabama	Indiana	North Carolina	Utah
Alaska	Iowa	Ohio	Vermont
Bureau of Indian Affairs (BIE)	Kentucky	Oklahoma	Virgin Islands
Connecticut	Montana	Rhode Island	Virginia
Hawaii	Nevada	South Carolina	Wyoming

¹ Census divisions were also considered, because the 10 divisions break the country down into smaller regions, and because they treat the special jurisdictions as a separate group rather than including them in regions with U.S. states. However, by pulling the special jurisdictions out of the regional strata, we can accomplish the same thing while still maintaining the use of the more familiar (and potentially more policy relevant) federal regions.

The evaluation review team collected evaluation-related documents from all of these states (except for Utah, which had not conducted program evaluation for their 21st CCLC program). In addition, the team collected evaluation-related documents from a sample of sub-grantees from each state² (further described below). The evaluation review team also collected both updated evaluation documents and interview data from states that were part of the 2011 monitoring visit sample and had been included in the evaluation review sample in 2010 (*updated 2010 sample*). In addition, part of the monitoring visits entailed that the Evaluation Lead collect current evaluation related documents. The states included in the *updated 2010 sample* are:

Colorado	Massachusetts	Nebraska	Puerto Rico
Delaware	Missouri	Pennsylvania	Texas

Selection of States for Review during 2011-2012

For 2012, states that had not been part of the evaluation review sample in 2010 or 2011 were included in the evaluation review sample for 2012. The resulting sample for 2012 included the following 16 states and special jurisdictions:

Arkansas	Georgia	Michigan	Tennessee
California	Kansas	Minnesota	Washington
District of Columbia	Maryland	North Dakota	West Virginia
Florida	Maine	New Jersey	Wisconsin

The evaluation review team collected evaluation-related documents from all of these states as well as a sample of sub-grantees from each state (further described below). The evaluation review team also requested updated evaluation documents from 10 states that were monitored and included in the evaluation review sample in 2010. Seven of the 10 states submitted updated documents to the evaluation review team. In addition, the Evaluation Lead collected current evaluation documents from New Mexico as part of the monitoring visit (New Mexico was the only state monitored in 2012 that was part of the Evaluation Review sample in 2010). The states included in the updated sample are:

Arizona	Louisiana	New Mexico	South Dakota
Idaho	Mississippi	New York	
Illinois	New Hampshire	Oregon	

Selection of Sub-Grantees for Review

For each SEA, a representative sample of sub-grantees was selected using a random sampling method stratified by funding amount and type of sub-grantee (e.g., school, district or other type of agency). Only sub-grantees active at the dates of sampling with grant award dates beginning during specific time periods were included in sampling. These time periods are as follows: September 2005 and January 2009 for Base Year; September 2005 and October 2009 for Option Year 1; and September 2005 and October 2010 for Option Year 2³. For each state, the sub-grantees were split into three groups (small, medium and large sub-grantees) based on the funding amount. An additional group for the largest sub-

² There was one exception: BPA did not receive sub-grantee documents from Iowa.

³ The team used this cutoff for Option Year 2 because APR data for 2010 was not finalized until early 2011, and also because more recent sub-grantees were not likely to have conducted an evaluation if they were awarded a grant less than one year before.

grantee(s) was excluded from these three groups. If a state contained a single largest sub-grantee, this sub-grantee was automatically included in the sample. If the second largest sub-grantee funding amount was very close to the amount of the largest sub-grantee, the largest sub-grantee was included with the “large” group. If there were multiple sub-grantees that had the identical highest funding amount, these sub-grantees were grouped together to become a fourth group for sampling. In the end, a random sample of at least 10 sub-grantees was selected in each state, or 10 percent of the sub-grantees in states with more than 100 total sub-grantees. The sample number from each group was proportional to the size of each group. In addition, the number of school-based and non-school-based sub-grantees sampled was proportional to the number within each group. Table A1 shows the breakdown of the total sub-grantees and the total number sampled.

Table A1. Sub-Grantee Sample

	Total Sub-Grantees	Total Sub-Grantees in Sample	Total Centers	Total Active Centers	% of Active Centers	Sub-Grantees Submitting documents	Sub-Grantees Not Submitting Documents
Base Year	1385	213	4011	793	20%	177	36
OY1	560	215	1236	287	23%	155	60
OY2	975	167	3031	540	18%	126	41
TOTAL	2920	595	8278	1620	20%	469	126

Document Review

One of the primary sources of information regarding evaluation practices was the review of state and sub-grantee evaluation-related documents. The full list of documents requested of SEAs and sub-grantees is included as Appendix C. In general, the types of documents reviewed included:

- Evaluation reports;
- SEA requests for proposals (RFPs) or grant solicitation packages;
- Scopes of work for evaluators;
- Guidance materials and documentation of technical assistance provided by SEAs to sub-grantees; and
- Evidence of use and dissemination of evaluation results.

Table A2 summarizes the numbers and types of documents reviewed over all three years. While 2,272 documents were reviewed, in many cases several related documents were grouped together and assigned a single identification number for coding purposes. For example, if a sub-grantee sent multiple examples of dissemination materials, these were grouped together for the purposes of coding dissemination practices. Table A2 reflects the total number of individual documents reviewed over the three years.

Table A2. Grantee Evaluation Documents Reviewed

Type of Documents	2010	2011	2012	Total
State evaluation report	32	42	27	101
Sub-grantee evaluation report	320	367	227	914
State RFP (Grant Solicitation Package)	35	29	36	100
Guidance material	60	33	52	145
State evaluation contract	20	35	34	89
Sub-grantee evaluation contract	92	38	51	181
Documentation of technical assistance provided	89	56	49	194
Documentation of evaluation use	169	39	58	266
Documentation of dissemination of results	124	68	45	237
Other	45			45
Total Documents Reviewed	986	707	579	2272
<i>Not reviewed*</i>	505	334	336	1175
Total Documents Collected	1491	1,041	915	3447

*The documents labeled “not reviewed” refer to materials that were collected and screened but deemed unnecessary to review and code (e.g., newsletter article advertising program parents, an email or memo containing no relevant evaluation-related information).

Detailed Document Collection and Review Procedures

BPA obtained copies of the most current state and sub-grantee evaluation materials using the following processes:

- In 2010 and 2011, SEAs that were receiving on-site monitoring visits in the same year they were part of the evaluation review sample submitted their most recent evaluation report, grant application, contract with the statewide evaluator and evaluation reports from the sub-grantees selected to be visited during the monitoring trip directly to monitoring teams. SEA Coordinators submitted other state-level evaluation documents and relevant evaluation documents from sub-grantees in the evaluation review sample to the evaluation review team. Since these states were being asked to submit documents prior to their monitoring visits, they were asked to submit evaluation documents to BPA’s evaluation review team at the same time. These documents were submitted on a rolling basis throughout the year.
- In 2010 and 2011, SEA coordinators in states that received monitoring visits in different years than the year they were part of the evaluation review sample submitted state-level evaluation documents and evaluation documents from sub-grantees in the evaluation review sample directly to the evaluation review team.
- In 2012, SEA coordinators in all states that were in the evaluation review sample and/or received on-site monitoring visits submitted documents directly to the evaluation review team.

Once documents were received, a team of BPA and AIR analysts examined state and sub-grantee evaluation materials. The team screened all incoming documents and set aside those documents deemed irrelevant to the task. Relevant documents were categorized, given IDs, and then assigned to team members for coding. The assigned team members coded the evaluation documents, entering all results into an electronic document review database. Reviewers were instructed to code various

aspects of the documents including: type of evaluator, measures used, research design used, description of sample, evidence of use, etc. The data entry system used both closed-ended variables (e.g., “Does the SEA require that specific performance measures are used in the sub-grantee evaluations?”), check boxes for specific items (e.g., evaluation measures include: math achievement using state standardized tests), and fields for open-ended comments (e.g., “According to the document, how were evaluation results used?”). The instructions and codebook in Appendices E and F describe the procedures for coding and entering data. Table A3 displays the number of documents reviewed from each state or SEA during all three years of the contract.

Table A3. SEA and Sub-grantee Evaluation Documents Reviewed

	State	Number of State Documents Reviewed	Number of Sub-Grantee Documents Reviewed	Number of Updated State Documents Reviewed	Totals
BASE YEAR	Arizona	12	17		29
	Colorado	31	114	3	148
	Delaware	11	33	4	48
	Idaho	14	48		62
	Illinois	2	45	12	59
	Louisiana	3	17	5	25
	Massachusetts	11	24	4	39
	Mississippi	9	15		24
	Missouri	7	19	4	30
	Nebraska	7	40	1	48
	New Hampshire	9	38	4	51
	New Mexico	9	61	2	72
	New York	30	84	5	119
	Oregon	20	20	3	43
	Pennsylvania	22	70	3	95
	Puerto Rico	41	16	1	58
	South Dakota	8	2	4	14
	Texas	43	34	5	82
	OPTION YEAR 1	Alabama	24	39	
Alaska		11	16		27
Bureau of Indian Education		4	10		14
Connecticut		4	1		5
Hawaii		9	28		37
Indiana		21	14		35
Iowa		5	0		5
Kentucky		17	36		53
Montana		4	18		22
Nevada		3	12		15
North Carolina		4	16		20
Ohio		18	49		67
Oklahoma		4	10		14
Rhode Island		14	75		89
South Carolina		12	10		22
Vermont		10	43		53
Virgin Islands		1	0		1
Virginia		18	37		55
Wyoming		8	77		85

	State	Number of State Documents Reviewed	Number of Sub-Grantee Documents Reviewed	Number of Updated State Documents Reviewed	Totals
OPTION YEAR 2	Arkansas	11	19		30
	California	6	5		11
	District of Columbia	6	11		17
	Florida	11	20		31
	Georgia	12	43		55
	Kansas	12	26		38
	Maryland	12	31		43
	Maine	8	10		18
	Michigan	9	40		49
	Minnesota	3	10		13
	North Dakota	11	0		11
	New Jersey	10	36		46
	Tennessee	12	10		22
	Washington	28	54		82
	West Virginia	10	20		30
	Wisconsin	24	24		48
		TOTAL	665	1547	60

To help ensure consistency across document reviewers, the evaluation review team performed inter-rater reliability checks. Two reviewers were assigned to each state and instructed to independently code three documents. Upon completion, they came together to compare notes and address any confusion or inconsistencies that arose during coding. Reviewers documented their ratings, discrepancies and agreements reached in a spreadsheet that was sent to the BPA project manager. In 2010, inter-rater reliability was performed after all documents had been reviewed and coded. Senior members of the evaluation review team used information gathered from reviewing inter-rater reliability data and feedback from reviewers to refine the codebook and coding procedures. In 2011 and 2012, analysts performed the inter-rater reliability at the beginning of the coding process before the entire set of documents was reviewed for each state. This process helped reviewers come to a shared understanding of coding procedures and ensured increased consistency throughout the document review. It also ensured that both reviewers had the basic background information about the state’s evaluation process before coding the rest of the state’s documents, as each reviewer had reviewed the state evaluation report, state guidance to their sub-grantees on evaluation, and a sample sub-grantee evaluation report.

Interviews

Interviews were conducted with SEA and sub-grantee representatives and evaluators for all 54 states and special jurisdictions that received on-site monitoring visits from the Department and BPA’s monitoring team over the three years of the contract. During these visits, the designated Evaluation Lead from the BPA team gathered information from key respondents about SEA and sub-grantee evaluation practices (see Appendix D for Interview Guides). The information gathered on-site provided valuable context and background for understanding the documents being reviewed. To take advantage of this knowledge, document review teams were assigned – to the extent possible -such that site visit Evaluation Leads were also involved in document review for the state(s) they visited. In 2011 and 2012, Evaluation Leads also submitted formal written summaries (using the Interview Guides) of these conversations to the evaluation review team, so that throughout the data analysis process all evaluation review team members had access to the knowledge gained on-site. These written summaries were also

entered into the team’s qualitative data analysis software package, Nvivo, for coding, analyzing and drafting sections of the 2011 Interim Report and the 2012 Final Report.

Desk Monitoring Calls

In 2010, BPA participated in the Department’s desk monitoring calls with SEA coordinators in states that were included in the 2010 evaluation sample but did not receive on-site monitoring visits that year, in order to help provide context for the document review. During these calls, BPA staff asked SEA coordinators about their state evaluation efforts. Questions included:

- When was the most recent state evaluation conducted?
- Who conducted the evaluation?
- How much did the evaluation cost?
- How many of the states’ sub-grantees conduct evaluations and submit evaluation reports to the SEA?
- What guidance and/or technical assistance on program evaluation does the state provide to sub-grantees?
- How does the state use evaluation results for program improvement?
- How do sub-grantees use their evaluations for program improvement?
- How are evaluation results disseminated?
- What are the biggest challenges facing the state and sub-grantees in terms of conducting evaluation activities?

In 2012, BPA participated in the Department’s desk monitoring calls with SEA coordinators in states that were in the evaluation review sample and received on-site monitoring visits in 2010. BPA made requests to federal program officers to participate in desk monitoring calls with nine states and interviewed SEA coordinators in five states. The purpose of participating in the calls was to gather information about any relevant changes in evaluation practices, as the evaluation review team did not gather any document review or interview data from these states in 2011 or 2012.

Data sources for each SEA

Table A4 describes the sources of data the BPA team collected for each SEA included in the evaluation review task.

Table A4. Data Sources for Evaluation Review Task 2010-2012

State	State Documents Reviewed	Sub-Grantee Documents Reviewed	Interviews	Updated State Level Documents	Desk Monitoring Calls
Alabama	2011	2011	2010	-	-
Alaska	2011	2011	2011	-	-
Arizona	2010	2010	2010	-	-
Arkansas	2012	2012	2012	-	-
Bureau of Indian Education	2011	2011	2011	-	-
California	2012	2012	2012	-	-

State	State Documents Reviewed	Sub-Grantee Documents Reviewed	Interviews	Updated State Level Documents	Desk Monitoring Calls
Colorado	2010	2010	2011	2011	2010
Connecticut	2011	2011	2011	-	-
D.C.	2012	2012	2012	-	-
Delaware	2010	2010	2011	2011	2010
Florida	2012	2012	2012	-	-
Georgia	2012	2012	2012	-	-
Hawaii	2011	2011	2011	-	-
Idaho	2010	2010	2010	-	2012
Illinois	2010	2010	2010	2012	2012
Indiana	2011	2011	2011	-	-
Iowa	2011	-	2010	-	-
Kansas	2012	2012	2012	-	-
Kentucky	2011	2011	2011	-	-
Louisiana	2010	2010	2011	2012	2010
Maine	2012	2012	2012	-	-
Maryland	2012	2012	2012	-	-
Massachusetts	2010	2010	2011	2011	2010
Michigan	2012	2012	2012	-	-
Minnesota	2012	2012	2012	-	-
Mississippi	2010	2010	2010	-	-
Missouri	2010	2010	2011	2011	2010
Montana	2011	2011	2010	-	-
Nebraska	2010	2010	2011	2011	2010
Nevada	2011	2011	2011	-	-
New Hampshire	2010	2010	2010	2012	2012
New Mexico	2010	2010	2012	2012	2010
New Jersey	2012	2012	2012	-	-
New York	2010	2010	2010	2012	2012
North Carolina	2011	2011	2011	-	-
North Dakota	2012	2012	2012	-	-
Ohio	2011	2011	2010	-	-
Oklahoma	2011	2011	2010	-	-
Oregon	2010	2010	2010	2012	2012
Pennsylvania	2010	2010	2011	2011	2010
Puerto Rico	2010	2010	2011	2011	2010
Rhode Island	2011	2011	2010	-	-
South Carolina	2011	2011	2011	-	-
South Dakota	2010	2010	2010	2012	-
Tennessee	2012	2012	2012	-	-

State	State Documents Reviewed	Sub-Grantee Documents Reviewed	Interviews	Updated State Level Documents	Desk Monitoring Calls
Texas	2010	2010	2011	2011	2010
Utah	-	-	2010	-	-
Vermont	2011	2011	2010	-	-
Virgin Islands	2011	2011	2011	-	-
Virginia	2011	2011	2011	-	-
Washington	2012	2012	2012	-	-
West Virginia	2012	2012	2012	-	-
Wisconsin	2012	2012	2012	-	-
Wyoming	2011	2011	2010	-	-
TOTAL SEAs	53	52	54	15	15

Appendix B

Analysis Plan

Research Questions	Source	Unit of Analysis	Year(s)	Variables/Nodes	Analysis
<i>a. What is the range of approaches that are being used to evaluate 21st CCLC programs?</i>					
a.1i. What percentage of states conduct program evaluations/assessments?	Stata	SEA	All	A8	Frequency
a.1ii. What percentage of sub-grantees conduct program evaluations/assessments?	Stata	Sub-Grantee	All	A8	Frequency/ percentage
a.2i. How often are state evaluations conducted?	Nvivo	SEA	All	Frequency of Evaluation node	Summarize info in Frequency of Evaluation node; Query state sources
a.2ii. How often are sub-grantee evaluations conducted?	Nvivo	Sub-Grantee	All	Frequency of Evaluation node	Summarize info in Frequency of Evaluation node; Query sources from sub-grantee in the evaluation review sample
a.3 i. What evaluation designs are state programs using to evaluate/assess their 21st CCLC programs (types of research design and analysis)?	Stata	SEA	All	F2	Frequency of F2
a.3 ii. What evaluation designs are state programs using to evaluate/assess their 21st CCLC programs (types of research design and analysis)?	Nvivo	SEA	All	Evaluation Design & Analysis node (and sub-nodes)	Use info from Nvivo to identify examples of each design and of 'other' designs that do not fit into F2 categories; Query state sources
a.3.ii-a. Are states articulating key research/ evaluation questions? What are some examples?	Nvivo	SEA	All	Research/ Evaluation Questions node	Summarize info in the Research/Evaluation Questions node; Query state sources
a.3 iii. What evaluation designs are sub-grantee programs using to evaluate/assess their 21st CCLC programs (types of research design and analysis)?	Stata	Sub-Grantee	All	F2	Frequency of F2
a.3 iv. What evaluation designs are sub-grantee programs using to evaluate/assess their 21st CCLC programs (types of research design and analysis)?	Nvivo	Sub-Grantee	All	Evaluation Design & Analysis node (and sub- nodes)	Use info from Nvivo to identify examples of designs and of other designs that do not fit into F2 categories; Query sources from sub-grantee in the evaluation review sample
a.3. iv-a. Are sub-grantees articulating key research/ evaluation questions? What are some examples?	Nvivo	Sub-Grantee	All	Research/ Evaluation Questions node	Summarize info in the Research/Evaluation Questions node; Query sources from sub-grantee in the evaluation review sample
a.3v. What types of process/implementation measures are programs using?	Stata	SEA	All	D19-D34	Frequencies of measures for all SEAs

Research Questions	Source	Unit of Analysis	Year(s)	Variables/Nodes	Analysis
a.3v. What types of process/implementation measures are programs using?	Nvivo	SEA	All	Process Measures node	Use info from Nvivo to identify examples of Other process measures; summarize info about process measures provided in Nvivo; Query state sources
a.3vi. What types of outcome measures are programs using?	Stata	SEA	All	D1-D16	Frequencies of measures for all SEAs
a.3vi. What types of outcome measures are programs using?	Nvivo	SEA	All	Family/parent outcome measures, Other Outcome measures, Student Academic outcome measures, & student non-academic outcome measures nodes	Use info from NVivo to identify examples of different types of outcome measures; summarize information about outcome measures provided NVivo; Query state sources
a.3vii. What types of process/implementation measures are programs using?	Stata	Sub-Grantee	All	D19-D34	Frequencies of measures for all Sub-Grantees
a.3vii. What types of process/implementation measures are programs using?	Nvivo	Sub-Grantee	All	Process Measures node	Use info from Nvivo to identify examples of Other process measures; summarize info about process measures provided in Nvivo; Query sources from sub-grantee in the evaluation review sample
a.3viii. What types of outcome measures are programs using?	Stata	Sub-Grantee	All	D1-D16	Frequencies of measures for all Sub-Grantees

Research Questions	Source	Unit of Analysis	Year(s)	Variables/Nodes	Analysis
a.3viii. What types of outcome measures are programs using?	Nvivo	Sub-Grantee	All	Family/parent outcome measures, Other Outcome measures, Student Academic outcome measures, & student non-academic outcome measures nodes	Use info from NVivo to identify examples of different types of outcome measures; summarize information about outcome measures provided NVivo; Query sources from sub-grantees in the evaluation review sample
a.4. How do state evaluation practices vary by size of states?	Stata	SEA	All	F2, Size variable	Crosstab: F2 x Size variable (small, medium, large states)
a.4. How does sub-grantee evaluation design vary by size of states?	Stata	SEA, Sub-Grantee	All	F2 (for Sub-Grantees) and Size variable (for States)	Crosstabs: % of sub-grantees using particular design x State size: small, medium, large; chi-square test of association
a. 5. To what extent does sub-grantee evaluation design vary within states?	Stata	SEA, Sub-Grantee	All	F2	Divide SEAs along categories: 1) all sub-grantees use same design; 2) all sub-grantees except one use same design; 2) sub-grantees use one of two designs, 4) sub-grantees use one of three (or more) designs.
a.6. What is the relationship between design of state evaluations and design of its sub-grantee evaluations?	Stata	SEA, Sub-Grantee	All	F2	Crosstab: % of sub-grantees using particular design x type of state evaluation design; Divide states by design of state level evaluation; tabulate frequency of various sub-grantee designs within each group.
a.7. How many states specify/require specific performance measures of their sub-grantees?	Stata	SEA	All	C1	Frequencies of C1; gather data from RFPs or other guidance documents (if "Yes" in any of these docs)
a.7.i. Do these states produce more rigorous evaluations than others?	Stata	SEA	All	C1, F2	Crosstabs: F2 x C1
a.7.i. Do sub-grantees in these states produce more rigorous evaluations than others?	Stata	SEA, Sub-Grantee	All	C1, F2	Crosstab: Requirement of Performance Measures (C1) x % of Sub-grantee evaluation design types (F2)

Research Questions	Source	Unit of Analysis	Year(s)	Variables/Nodes	Analysis
a. 7.i.-a Examples of required measures	Nvivo	SEA, Sub-Grantee	All	State Requirements/Measures node	Use info from NVivo to identify examples of state requirements
a.8. How many states specify outcome targets for their sub-grantees?	Stata	SEA	All	C4	Frequencies of C4
a.8.i.Do these states produce more rigorous evaluations than others?	Stata	SEA	All	C4, F2	Crosstab: F2 x C4
a.8.i.Do sub-grantees in these states produce more rigorous evaluations than others?	Stata	SEA, Sub-Grantee	All	C4, F2	Crosstab: C4 x % of sub-grantee design types (F2)
a. 8.i.-a. Examples of targets	Nvivo	SEA, Sub-Grantee	All	State Requirements/Targets node	Use info from NVivo to identify examples of state targets
a.9. Who conducts the evaluations at state level?	Stata	SEA	All	B3	Frequencies of type of evaluator for all SEAs
a.9. Who conducts the evaluations at the state level?	Nvivo	SEA	All	Evaluator & Evaluator Selection nodes	Summarize info in Evaluator node; Use NVivo to identify examples of different types of evaluators and other info; Query state sources
a.9.i. Who conducts the evaluations at sub-grantee level?	Stata	Sub-Grantee	All	B3	Frequencies of type of evaluator for all Sub-Grantees
a.9.i. Who conducts the evaluations at sub-grantee level?	Nvivo	Sub-Grantee	All	Evaluator & Evaluator Selection nodes	Summarize info in Evaluator node; Use NVivo to identify examples of different types of evaluators and other info; Query sources from sub-grantees in the evaluation review sample
a.9.ii. What is the relationship between the type of evaluator and the rigor of the evaluation designs (state	Stata	SEA	All	B3, F2	Crosstab: B3 x F2
a.9.ii. What is the relationship between the type of evaluator and the rigor of the evaluation designs (sub-grantee level)?	Stata	Sub-Grantee	All	B3, F2	Crosstab: B3 x F2
a.10. How much funds do states spend on evaluation/assessments?	Nvivo	SEA	All	Cost node	Summarize info in Cost node; Query state sources
a.10. How much funds do sub-grantees spend on evaluation/assessments?	Nvivo	Sub-Grantee	All	Cost node	Summarize info in Cost node; Query sources from sub-grantees in the evaluation review sample

Research Questions	Source	Unit of Analysis	Year(s)	Variables/Nodes	Analysis
<i>Key Findings</i>					
b. What are the key findings from 21 st CCLC program evaluations (state level)?	Nvivo	SEA	All	Findings node	Summarize information/ key findings in Findings node; Query state sources
b. What are the key findings from 21 st CCLC program evaluations (sub-grantee level)?	Nvivo	Sub-Grantee	All	Findings node	Summarize information/ key findings in Findings node; Query sources from sub-grantees in evaluation review sample
<i>Guidance and TA</i>					
c. What guidance and TA are states providing to sub-grantees regarding program evaluation and what topics do they cover?	Nvivo	SEA, Sub-Grantee	All	Guidance provided by state & TA provided by state nodes	Summarize information from Guidance provided by state node and TA provided by state nodes
c. 1. How satisfied are sub-grantees with this guidance and TA?	Nvivo	SEA, Sub-Grantee	All	Sub-grantee TA needs & Challenges faced by sub-grantees nodes	Summarize information from Sub-grantee TA needs and Challenges faced by sub-grantees nodes
c. 2. Do states that provide "strong" evaluation guidance materials and technical assistance have sub-grantees that use more rigorous designs?	Nvivo, Stata	SEA, Sub-Grantee	All	Guidance provided by state, TA provided by state & Promising practices nodes, F2 (for sub-grantees)	Use information from NVivo to identify states with "strong TA/guidance" and create new variable in stata (yes/no); crosstab "strong TA/guidance" variable with F2 for sub-grantees
c.3 How does the quality of evaluation guidance materials and TA vary by size of state?	Nvivo, Stata	SEA, Sub-Grantee	All	Guidance provided by state, TA provided by state, & Promising practices, F2 (for sub-grantees)	crosstab size of state by new strong TA/guidance variable
<i>Use and dissemination</i>					
d. How do programs use evaluation results (states)?	Nvivo	SEA	All	Use node	Summarize information in Use node; Query state sources

Research Questions	Source	Unit of Analysis	Year(s)	Variables/Nodes	Analysis
d. How do programs use evaluation results (sub-grantees)?	Nvivo	Sub-Grantee	All	Use node	Summarize information in Use node; Query sources from sub-grantees in the evaluation review sample
e. How do programs disseminate evaluation results (states)?	Nvivo	SEA	All	Dissemination node	Summarize information in Dissemination node; Query state sources
e. How do programs disseminate evaluation results (sub-grantees)?	Nvivo	Sub-Grantee	All	Dissemination node	Summarize information in Dissemination node; Query sources from sub-grantees in the evaluation review sample
<i>Promising Practices</i>					
f. What are promising practices in evaluation among programs receiving 21st CCLC grants (states)?	Nvivo	SEA	All	Promising Practices node	Summarize information in Promising Practices node; Query state sources
f. What are promising practices in evaluation among programs receiving 21st CCLC grants (sub-grantees)?	Nvivo	Sub-Grantee	All	Promising Practices node	Summarize information in Promising Practices node; Query sources from sub-grantees in the evaluation review sample
g. What information or support via professional development do states need to conduct more rigorous evaluations?	Nvivo	SEA	All	State TA needs & Challenges faced by State node	Summarize info in the State TA needs & Challenges faced by State nodes
<i>Additional questions for BY states with updated info:</i>					
How have evaluation practices changed/improved over the last two years?		SEA	OY2		Summarize info from desk monitoring calls
How have evaluators changed?		SEA	OY2		Summarize info from desk monitoring calls
How have evaluation designs changed?		SEA	OY2		Summarize info from desk monitoring calls
How has guidance to sub-grantees changed?		SEA	OY2		Summarize info from desk monitoring calls
How has use changed?		SEA	OY2		Summarize info from desk monitoring calls

Appendix C

Evaluation Review Documents Requested



21st Century Community Learning Centers

Evaluation Review – Document Submission Checklist



Please provide the specific evaluation-related documents listed below *if they are available*. The documents should pertain to the evaluation of your state’s 21st Century Community Learning Center program and its sub-grantee programs. Use the checklist below to itemize the documents you are submitting in each category. We understand that not every SEA or sub-grantee will have all of the requested documents. If this is the case, please place an “N” in the Submitted column so that we know that no further follow-up is needed. Note that we request the **most recent** versions of documents. For example, if your program produces evaluation reports annually, send **only** the most recent evaluation report produced.

Please submit this completed form and the documents listed below to: Vincent@bpacal.com. Submit your electronic documents as email attachments, if possible. If you do not have electronic versions of documents, we would appreciate your scanning the document so you can send it electronically. If the volume of documents is too large, you can mail a CD to:

21st CCLC Evaluation Review
 c/o Berkeley Policy Associates
 440 Grand Avenue, Suite 500
 Berkeley, CA 94610
 (510) 465-7884

If submitting electronic documents is not feasible, you can mail hard copies to the address above.

If you have any questions, please contact:

- Your U.S. Department of Education Program Officer;
- Linda Toms Barker, Project Director, Berkeley Policy Associates (Linda@bpacal.com, 808-934-9297); or
- Dr. Nada Rayyes, Evaluation Review Task Leader, Berkeley Policy Associates (Nada@bpacal.com, 510-465-7884)

Documentation	Submitted (Y/N)	Date Submitted
State Level Documents		
1. Most recent grant solicitation package, RFP, or RFA , including description of evaluation requirements of applicants		
Title/description:		
2. Most recent state evaluation report		
Title/description:		
3. Scope of work for evaluator(s) or evaluation contract		
Title/description:		
4. Documentation of technical assistance on evaluation provided to sub-grantees (e.g., presentation of training on program evaluation)		
Title/description:		

5. Guidance materials provided to sub-grantees on conducting evaluation (e.g., required components and reporting for sub-grantee evaluation, templates, data collection instruments, performance measures/targets)		
Title/description:		
6. Examples of use of evaluation results for program improvement* (e.g., modifications to RFP based on evaluation results, recommendations to state policymakers on afterschool programming funding, development of TA support for sub-grantees based on results)		
Title/description:		
7. Documentation of dissemination of evaluation results* to stakeholders and public (e.g., snapshot of evaluation results posted on state website, distribution of report to state policymakers, discussion in media on statewide results)		
Title/description:		
Sub-Grantee Level Documents (Provide for each sub-grantee in the sample)		
1. Most recent sub-grantee evaluation report		
Title/description:		
2. Scope of work for evaluator(s) or evaluation contract		
Title/description:		
3. Examples of use of evaluation results for program improvement* (e.g., action plan including program modifications, based specifically on evaluation results)		
Title/description:		
4. Documentation of dissemination of evaluation results*		
Title/description:		

*Note: If you do not have any documents illustrating use and dissemination of evaluation results, you are not required to write a narrative in order to fulfill this request. **Dissemination** does not include program announcements in newsletters or press releases unless these items specifically discuss evaluation and include evaluation results.

Please provide the name, title, email address and phone number of the best person to contact for questions or additional information.

Name: _____ Title: _____

Email address: _____



21st Century Community Learning Centers



Evaluation Review – Document Submission Checklist

Please provide the specific evaluation-related documents listed below *if they have been changed or updated since the 2009-2010 year*. The documents should pertain to the evaluation of your state’s 21st Century Community Learning Center program and its sub-grantee programs. Use the checklist below to itemize the documents you are submitting in each category. If there are no documents to submit, please place an “N” in the Submitted column so that we know that no follow-up is needed.

Please submit this completed form and the documents listed below to: Vincent@bpacal.com. Submit your electronic documents as email attachments, if possible. If you do not have electronic versions of documents, we would appreciate your scanning the document so you can send it electronically. If the volume of documents is too large, you can mail a CD to:

21st CCLC Evaluation Review
 c/o Berkeley Policy Associates
 440 Grand Avenue, Suite 500
 Berkeley, CA 94610
 (510) 465-7884

If submitting electronic documents is not feasible, you can mail hard copies to the address above.

If you have any questions, please contact:

- Your U.S. Department of Education Program Officer;
- Linda Toms Barker, Project Director, Berkeley Policy Associates (Linda@bpacal.com, 808-934-9297); or
- Dr. Nada Rayyes, Evaluation Review Task Leader, Berkeley Policy Associates (Nada@bpacal.com, 510-465-7884)

Documentation	Submitted (Y/N)	Date Submitted
State Level Documents – Only submit UPDATED/NEW documents		
1. Most recent grant solicitation package, RFP, or RFA , including description of evaluation requirements of applicants		
Title/description:		
2. Most recent state evaluation report		
Title/description:		
3. Scope of work for evaluator(s) or evaluation contract		
Title/description:		
4. Documentation of technical assistance on evaluation provided to sub-grantees (e.g., presentation of training on program evaluation)		
Title/description:		

5. Guidance materials provided to sub-grantees on conducting evaluation (e.g., required components and reporting for sub-grantee evaluation, templates, data collection instruments, performance measures/targets)		
Title/description:		
6. Examples of use of evaluation results for program improvement (e.g., modifications to RFP based on evaluation results, recommendations to state policymakers on afterschool programming funding, development of TA support for sub-grantees based on results)		
Title/description:		
7. Documentation of dissemination of evaluation results to stakeholders and public (e.g., snapshot of evaluation results posted on state website, distribution of report to state policymakers, discussion in media on statewide results)		
Title/description:		

Please provide the name, title, email address and phone number of the best person to contact for questions or additional information.

Name: _____ Title: _____

Email address: _____

Appendix D

Interview Guide

**21st CCLC Evaluation Review
 Interview Guide for the Evaluation Lead
 State Respondent: SEA 21st CCLC Coordinator, State-level Evaluator**

Date:	SEA:
Respondent Name/Title:	
Respondent Organization/Department:	
Respondent Name/Title:	
Respondent Organization/Department:	
Respondent Name/Title:	
Respondent Organization/Department:	

1. How often is your program evaluated? When was the most recent statewide evaluation?

2. Who conducted your most recent statewide evaluation? What was the process for selecting your evaluator? *[If evaluator is SEA employee, what department does he/she work in? If not, what organization does he/she work with?]*

3. How much did the most recent evaluation cost? *[If exact amount is unknown, describe how the state has budgeted for evaluation activities.]* How was the evaluation funded? *[If with 21st CCLC grant, were evaluation costs allocated to the allowed 3% State Activities.]*

4. Please describe the evaluation design, including measures, how often data is collected, by what procedures, etc.
 - a. Describe **process** measures (e.g. hours of operation, program attendance, observations of activities, interviews with staff and students) used to evaluate the implementation of your program:

 - b. Does your state set targets for process measures (e.g., “increase program attendance by 30 students per semester next year”)? If so, what are they?

 - c. Describe **outcome** measures used in the comprehensive state evaluation, and any targets that are set?

 - d. Does the state set targets for outcome measures?

 - e. Describe any other measures or data collection not mentioned?

5. Could you describe your sampling strategy, if applicable?

6. Do you use comparison groups? How are they selected?

7. How does the state use evaluation results (e.g., to inform program improvement, fund allocation decisions, revise RFP, set funding priorities, technical assistance for sub-grantees)?

8. How are the state's evaluation results disseminated (e.g., web postings, media outlets, educational board meetings, conference presentations)? To whom (e.g., State education officials, local education agencies, national education forums, State policymakers, advocacy groups, media)?

9. What does the SEA require of sub-grantees in terms of evaluation?
 - a. Are all sub-grantees required to conduct evaluations?

 - b. Does the SEA require sub-grantees to submit their evaluation reports to the State? Approximately what percentage of sub-grantees submit evaluation reports?

 - c. How often are sub-grantees required to conduct evaluations of their 21st CCLC program (e.g., annually, once per 3-5 year grant period, end of grant period)?

 - d. What data are sub-grantees required to collect? These could include performance measures (e.g., grades, standardized tests, behavioral outcomes, parent outcomes, progress toward goals)? Are there any targets associated with these indicators (e.g., 10 percent improvement in test scores)?

 - e. Does your 21st CCLC program use a **formal** quality improvement process (QIP) (or continuous improvement process/CIP)? **[If No, skip to #10]**
 - i. Who conducts the QIP?

 - ii. Does the QIP include both process and outcome measures?

- iii. How does the state use the QIP information to inform program improvements and/or TA provided to sub-grantees?

 - iv. How is the QIP used by sub-grantees?

 - f. To what extent/in what ways do you integrate the QIP, PPICS and any other components of your evaluation effort?
10. What guidance and/or technical assistance on **program evaluation** does the state provide to sub-grantees (*e.g., guidance on selecting an evaluator, trainings about evaluation, manuals or tools/surveys to use, etc.*)?
11. How do you assess (or monitor) the quality of sub-grantee evaluations?
12. What are the biggest challenges facing the state in terms of conducting statewide evaluation of its 21st CCLC program?
13. What are the biggest challenges facing **sub-grantees** in terms of conducting evaluation of their 21st CCLC programs? What do you see as the needs of sub-grantees regarding evaluation?
14. What TA/guidance does the state receive from the Department regarding evaluation? What are your state's TA/support needs regarding evaluation?
15. Could you share any practices regarding your state's evaluation that you think are working well? [*e.g., regarding type of evaluation design, internal/external evaluator, cost-effective strategies, gathering useful information, etc*]

Thank you for your time!

Other

16. **Other: Please include any information gathered about the state's evaluation practices that is not covered in the responses above.**

**21st CCLC Evaluation Review 2012
 Evaluation Lead Interview Guide for *Sub-Grantees***

Date:	SEA:	Sub-grantee:
<i>Part of Evaluation Review Sample?</i>		<i>Check here:</i>
Respondent Name/Title:		
Respondent Organization/Department:		
Respondent Name/Title:		
Respondent Organization/Department:		
Respondent Name/Title:		
Respondent Organization/Department:		

1. When was your most recent evaluation conducted, and by whom? How often do you conduct evaluation? *[Confirm what is in the report. If you do not have a copy of this report, request to have one emailed to you or if you can have a hard copy,]*
2. How long have you been using this evaluator? How did you select this evaluator? Do you expect to continue to use the same evaluator for the next year/report?
3. *[If it is not obvious from report, ask about evaluation data collection and analysis.]* Can you provide a brief overview of what data is/was collected for your evaluation and how it is analyzed? *[Briefly describe data collection and analysis here]*
4. How much did the most recent evaluation cost?
5. What requirements, if any, does the SEA have for you regarding evaluation? *(e.g., frequency, performance measures/targets, RFP-listed requirements)*
6. What technical assistance and/or guidance about evaluation have you received from the State *(e.g. one-on-one trainings, conferences, webinars, sample documents, referrals to evaluators)*? Who provides this support? What kind of guidance materials or technical assistance on evaluation would be helpful to you that you aren't already receiving?

7. How have you used evaluation results? (e.g., staff feedback, partnership development, fundraising efforts, accountability with board, recruiting student participants, staff professional development)?
8. How are evaluation results disseminated (e.g., web postings, media outlets, conferences, education board meetings, partner networks)? To whom (e.g., program staff, advisory boards, students/parents, partner schools, other community partners, general public, SEA, state policy makers)?
9. What challenges have you experienced in conducting evaluation activities or working with your evaluator?
10. Could you share any practices regarding your evaluation that you think are working well? [*e.g., regarding type of evaluation design, internal/external evaluator, cost-effective strategies, gathering useful information, etc*]

Thank you for your time!

Other

11. **Other: Please include any information gathered about the sub-grantee's evaluation practices that is not covered in the responses above.**

Appendix E

Codebook for Reviewing Documents

21st CCLC Evaluation Review Codebook and Instructions

The Access Document Review database includes the following sections:

Data entry section	Documents related to section
A. Document Attributes	All docs
B. Attributes of Evaluation	Evaluation reports
C. Overview of Performance Measures	Evaluation reports, State RFP, Guidance Materials
D. Outcomes and Process Measures	Evaluation reports, Guidance Materials
E. Sample and Data Collection	Evaluation reports
F. Research Questions and Analysis of Student Outcomes	Evaluation reports
G. Presentation of Findings and Recommendations	Evaluation reports
H. Evaluation Guidance Materials	State RFP, Guidance Materials
I. Other Documentation	State RFP, TA, Use, Dissemination, SOW
J. General Impressions or Special Issues	All docs

1. Please complete the Access file for the documents you are coding using the *Access database Instructions* and *21st CCLC OY2 Doc. Coding Instructions*.
2. There are two tabs; be sure to complete all relevant sections.

Email file to Vincent (Vincent@bpacal.com) when complete
Contact Nada (Nada@bpacal.com) if you have coding/substantive questions

***Begin with Question A8 then complete all applicable (enabled) fields.**

A. Document Attributes	
A1.	Name: _____ (choose from drop down)
A2.	Enter document ID: _____
A2a.	Re-enter document ID: _____ [double enter doc ID to ensure accuracy]
Agency	
A3.	State: _____ (drop down)
A4.	Sub-grantee: _____ (drop down) [Skip if state-level document]
Dates	
A5.	Publication date: month _____ [If multiple docs, enter the date that appears on the first document alphabetically (a, b, c, etc.)]
A6.	Publication date: year _____
A7.	Academic Year(s) document pertains to: year(s) _____ [Enter year(s) such as 2008-09 or 2008-2010. If no date, leave blank. Enter comment or explanation if necessary].
*A8.	Type of Document(s) (drop down) [Documents have been grouped and/or categorized; coder should choose appropriate document type. If there is a question about document type, coder should discuss with Task Leader]
	State evaluation report (Go to sections A-G, J)
	Sub-grantee evaluation report (Go to sections A-G, J)
	State RFP (grant solicitation package) (Go to sections A, C, H, I, J)
	Guidance material (Go to sections A, C, D, H, J)
	State evaluation contract, scope of work (contract with outside evaluator, RFP for evaluator) (Go to sections A, I, J)
	Sub-grantee evaluation contract (contract with outside evaluator, RFP for evaluator) (Go to sections A, I, J)
	Documentation of technical assistance provided (Go to sections A, I, J)
	Documentation of evaluation use (Go to sections A, I, J)
	Documentation of dissemination of results (Go to sections A, I, J)
	Other (specify) : _____ (Go to sections A, J)
A9.	State-level document ____ [check if coding state level document(s)]

EVALUATION REPORTS

B. Attributes of Evaluation	
For B1-B2a: Select 'All', 'DK' or 'Other' and enter percentage and/or explanation in text field. Provide percentage if available; if not, indicate that information is missing.	
B1.	For STATE evaluation reports, what proportion of sub-grantees are included in the evaluation? ___All___ Don't Know ___Other
B1a.	Percentage/ Explanation: [Percentage?]
B2.	For SUB-GRANTEE evaluation reports, what proportion of centers/sites are included in the evaluation? ___All___ Don't Know ___Other
B2a.	Percentage/ Explanation: [Percentage?]
B3.	What type of evaluator conducted the evaluation? [Choose one. If you don't know or it's unclear, leave it blank]
	Program staff [e.g. program coordinator or personnel working for 21 st CCLC program or employed by grant]
	State or district personnel, but not program staff [e.g., state department of education research office]
	Academic institution [professor/researcher or group of researchers working for and representing an institution of higher education; research center based at university]
	Independent consultant [individual evaluator]
	Research agency/firm [e.g., BPA, WestEd]
B4.	Name of evaluator and/or agency:

C. Overview of Performance Measures	
C1.	Does SEA require that specific measures be included in the SUB-GRANTEE evaluation? ___ Yes ___ No ___ DK [most likely will be a Yes or Don't Know] (If "NO" or "DK", skip to C3.)
C1a.	If Yes, what are the required measures? _____
C1b.	Are these measures used in the evaluation (if applicable)? ___ Yes ___ No ___ DK
C2.	Are <i>additional</i> measures also included in the evaluation? ___ Yes ___ No ___ DK [most likely will be a Don't Know]
C3.	Does SEA set performance targets for statewide evaluation? ___ Y ___ N ___ DK [a target uses a specific numerical benchmark; see examples in C3a and C4a below]
C3a.	What are the STATEWIDE targets? (e.g., increase student achievement on reading assessment by 10 points annually):
C4.	Does SEA set targets for sub-grantees ___ Y ___ N ___ DK
C5.	Does sub-grantee set performance targets for themselves? ___ Yes ___ No ___ DK
C5a.	What are the SUB-GRANTEE targets? _____ [unlimited text box]

D. Outcome and Process Measures	
Student Academic Outcome Measures [For evaluation reports, indicate measure and source if data is actually collected and tabulated as part of the evaluation; not if measure is mentioned or planned for use in the future. For guidance materials, indicate measure and source if guidance is provided specifically about the measure. Leave blank if source or measure is not listed. For each measure, check appropriate source(s).]	
Measure [check if the measure is used; if measure is not used, skip it]	Source Options [for each measure, check source(s) used. If the source is unknown, check 'other' and comment]
D1. <i>Math achievement</i>	<i>State standardized test</i>
	<i>School/classroom grades</i>
	<i>Teacher assessment (e.g. teacher survey or questionnaire)</i>
	<i>Other (including local and formative assessments, student survey)</i>
D2. <i>Reading/ELA achievement</i>	<i>State standardized test</i>
	<i>School/classroom grades</i>
	<i>Teacher assessment (e.g. teacher survey or questionnaire)</i>
	<i>Other (including local and formative assessments, student survey)</i>
D3. <i>Other content area achievement</i>	<i>State standardized test</i>
	<i>School/classroom grades</i>
	<i>Teacher assessment (e.g. teacher survey or questionnaire)</i>
	<i>Other (including local and formative assessments, student survey)</i>
D4. <i>Other (could include grades but content not specified; and AYP)</i>	<i>State standardized test</i>
	<i>School/classroom grades</i>
	<i>Teacher assessment (e.g. teacher survey or questionnaire)</i>
	<i>Other (including local and formative assessments, student survey)</i>
D5.	Comments: explain 'Other' options here; name of state standardized tests (please don't write an acronym for a state test that may be unfamiliar)
Student Non-academic Outcome Measures	
Measure [check if the measure is used; if measure is not used, skip it]	Source Options [for each measure, check source(s) used]
D6. <i>Behavior</i>	<i>School records</i>
	<i>Teacher assessment (e.g. teacher survey)</i>

D6.	<i>Behavior (contd.)</i>	Standardized instrument (e.g., Social Skills Rating Scale)
		Student assessment (e.g. student survey)
		Parent assessment (e.g. parent survey)
		PPICs
		Other
D7.	Disciplinary incidents(not just "discipline in general"; but a collection of data - and analysis - of actual incidents)	School records
		Teacher assessment (e.g. teacher survey)
		Student assessment (e.g. student survey)
		Parent assessment (e.g. parent survey)
		PPICs
D8.	School/classroom (not program) attendance	School records
		Teacher assessment (e.g. teacher survey)
		Student assessment (e.g. student survey)
		Parent assessment (e.g. parent survey)
		PPICs
D9.	<i>Satisfaction/attitude toward school</i>	Teacher assessment (e.g. teacher survey)
		Student assessment (e.g. student survey)
		Parent assessment (e.g. parent survey)
		PPICs
		Other
D10.	<i>Homework Completion</i>	School records
		Teacher assessment (e.g. teacher survey)
		Student assessment (e.g. student survey)
		Parent assessment (e.g. parent survey)
		PPICs
D11.	<i>Other</i>	School records
		Teacher assessment (e.g. teacher survey)
		Standardized instrument (e.g., Social Skills Rating Scale)
		Student assessment (e.g. student survey)
		Parent assessment (e.g. parent survey)
D12.	<i>Comments: [explain 'Other'] options here</i>	PPICs
		Other
Parent/Family Outcomes		
	Measure [check if the measure is used; if measure is not used, skip it]	Source Options [for each measure, check source(s) used]
D13.	<i>Satisfaction with child's school</i>	Parent assessment (e.g. parent survey)
		Teacher assessment (e.g. teacher survey)
		PPICs
		Other
D14.	<i>Parent involvement in school/classroom</i>	Parent assessment (e.g. parent survey)
		Teacher assessment (e.g. teacher survey)
		School records
		PPICs

D14.	<i>Parent involvement in school/classroom (contd.)</i>	<i>Other</i>
D15.	<i>Parent employment</i>	<i>Parent assessment (e.g. parent survey)</i>
		<i>Teacher assessment (e.g. teacher survey)</i>
		<i>School records</i>
		<i>PPICs</i>
D16.	<i>Other</i>	<i>Parent assessment (e.g. parent survey)</i>
		<i>Teacher assessment (e.g. teacher survey)</i>
		<i>School records</i>
		<i>PPICs</i>
D17.	<i>Comments: [explain 'Other'] options here</i>	
Schoolwide Safety/Climate Outcome Measures		
D18.	List measure and source/method: [These are rare. If they are using a school-wide outcome measure, please describe it here]	
Process Measures [For evaluation reports, indicate measure and source if data is actually collected and changes are described part of the evaluation; not if measure is mentioned or planned for use in the future. For guidance materials, indicate measure and source if guidance is provided specifically about the measure. Leave blank if source or measure is not listed. For each measure, check appropriate source(s).]		
Measure [check if the measure is used; if measure is not used, skip it]		Source Options [for each measure, check source(s) used]
D19.	<i>Adult-to-student ratio</i>	<i>Teacher assessment (e.g. teacher survey)</i>
D20.	<i>Community partnerships</i>	<i>Parent assessment (e.g. parent survey)</i>
D21.	<i>Core academic activities</i>	<i>Student assessment (e.g. student survey)</i>
D22.	<i>Recreational and enrichment activities</i>	<i>Staff assessment (e.g. Survey)</i>
D23.	<i>Links to school day</i>	<i>School records</i>
D24.	<i>Service hours provided</i>	<i>Focus groups</i>
D25.	<i>Counseling and mentoring to students</i>	<i>Interviews</i>
D26.	<i>Program attendance</i>	<i>Observations</i>
D27.	<i>Services to adults</i>	<i>Program tracking/ attendance records/PPICs</i>
D28.	<i>Communications with parents</i>	<i>Other</i>
D29.	<i>Parent involvement in 21st Century program</i>	
D30.	<i>Parent satisfaction with 21st Century program</i>	
D31.	<i>Teacher/administrator satisfaction with 21st Century program</i>	
D32.	<i>Student satisfaction with 21st Century program</i>	
D33.	<i>Program implementation issues</i>	
D34.	<i>Other</i>	
D35.	<i>Comments: [explain 'Other'] options here</i>	

E. Sample and Data Collection	
E2.	Characteristics of the evaluation sample and/or participants included in evaluation are described? ___Yes ___ No
E2a.	If yes, which characteristics are described (check all that apply): ___race ___gender ___grade levels ___free/reduced lunch ___ELL ___ Students with disabilities ___ Other
E2b.	Other: Specify.

E3.	Does the evaluation sample include representative a group of participants? Please explain: [“representative” means that the students included in the evaluation represent all of the students served by the program in terms of grade levels served, race/ethnic group served, students participating in free/reduced lunch, etc. E.g., if the program serves K-5 students, and only 5th graders are interviewed for the evaluation, this is not a representative sample. Indicate if all participants are included in evaluation.]
Data Collection and Response Rates	
E4.	Describe data collection; was it systematic and unbiased? Are procedures thoroughly reported? Are response rates reported?
E4a.	Are all relevant data collection procedures are described ___ Yes [all data collection procedures must be described; and all relevant response rates reported.]

F. Research Questions and Analysis of Student Outcomes

F1.	What are the evaluation’s key research questions and/or objectives, as stated in the document ? (These should be explicitly stated, e.g., “How effective was the 21 st CCLC program in raising math test scores among participants?” Or “The objective of the evaluation is 1) to understand how program implementation can be improved, 2) to examine the impact of the program on students’ reading assessment scores, and 3) to determine whether student satisfaction with the program has increased from the previous year.” Reviewer should not have to interpret information implicitly. If not explicitly stated, leave blank. If unsure, explain here.)
F2.	Please indicate the research design used to analyze student outcomes : [Choose the design used for the evaluation. If more than one design is used, choose the most rigorous and describe in space below. If it is unclear what design is used, explain in the space below].
	Experimental: Random assignment to treatment and control groups ___ [rare]
	Comparison group is formed , using matching or controlling technique ___
	Comparison to district, state, or national average; may or may not include pre-intervention measure ___
	Single group (no comparison group) of students, pre-post comparison ___
	Single group of students, multiple points in time (no pre-test) (e.g., averages for program participants are presented) ___
	Program comparisons are made over time (different groups of students; multiple points in time)
	Single group of students, single point in time ___
F2a.	Please describe the design and analysis, given the information provided. Does the design address the research questions? Does the design include following the same group of students over time? [regardless of the selection above, please briefly explain the design or analysis conducted, providing necessary details.]

G. Presentation of Findings and Recommendations

G1.	Please share key findings/results from the evaluation (focus on student academic outcomes/ math and reading -- e.g., 75% of participants’ math grades improved compared to 50% of general district population):
G2.	Do findings address research questions/ evaluation objectives? Why or why not?
G3.	Are findings supported by data and evidence? Explain:
G4.	Are recommendations presented? Explain/describe:

EVALUATION GUIDANCE MATERIALS

H. RFP/Grant Solicitations, Evaluation Manuals/Instructions	
Guidance Provided	
H1.	What type(s) of guidance, <i>about evaluation</i> , is provided in the document(s)? (e.g., data collection instruments, evaluation time line/calendar, how to use results, evaluation manual, data collection procedures):
H2.	Does the document include performance measures required by the state? <input type="checkbox"/> Yes [check yes if the state explicitly requires certain measures be used, e.g., state assessment scores in reading and math must be collected for all participants]
H2a.	If Yes, what guidance is provided about these measures?

OTHER DOCUMENTATION

I. Other Documentation	
Documentation of TA or professional development delivered on Evaluation:	
I1.	According to the document(s), what types of technical assistance or professional development (trainings, presentations at conferences, etc.) were provided <i>about evaluation</i> ? Describe what topics were covered by the TA provided: (e.g., "regional trainings about evaluation requirements are provided by the state every fall. Program coordinators and local evaluators participate in these one-day trainings and receive guidelines about required measures, evaluation time line and budgets.")
Documentation of Evaluation Use	
I2.	According to the document(s), how were evaluation results used? Were they used for program improvement? Explain: [If guidance materials] According to the document, how should evaluation results be used?
Documentation of Dissemination	
I3.	According to the document(s), how were evaluation results disseminated? What methods of dissemination are discussed, and to what types of audiences? [If guidance materials] According to the document, how should results be disseminated?
Scope of Work, Evaluation Contract, RFP for Evaluator	
I4.	Are clear expectations for evaluation services articulated in the document? Explain:
I5.	Are costs broken down by item/ activity? Is evaluation budget sufficiently detailed? Explain:

GENERAL IMPRESSION

J. General Impression or Special Issues	
J.	Please briefly describe your general impression of the document. Include any promising practices and/or weaknesses. Please also include any information about special issues that we should be aware of with this document that you have not already documented in your codes.

Appendix F

Document Review Instructions and Guidance to Reviewers

**21st Century Community Learning Centers Project
Evaluation Review Task 2.3 Option Year 2 (2011-2012)
Evaluation Document Review and Coding**

Instructions

Notes:

- All necessary materials will be sent to you in an email from Task Leader. Save this email!
- Follow these instructions, step by step, for each state you are assigned.

Prepare Materials:

1. Save a version of the Access database for each state you are coding. Use the following convention: “full state name_your initials”, e.g. “Colorado_AB”.
2. Open the Google Docs spreadsheet for the state you are coding.
3. Open the Dropbox folder that contains the documents you need to code.
4. In your Dropbox folder you may see sub-folders named “DO NOT CODE”. You do not need to do anything with these. As the title suggests, you do not need to code these, but we want to keep them in our records since they were sent to us by the SEAs.
5. Each document or set of documents has an ID number. At the top of the Google Docs spreadsheet, you will see ‘Reliability Documents’ listed. *Code these 3 documents first.*
6. Open the Access database for the state you are coding. In order to do this, open the document titled *Access database instructions 01-20-12.docx* and follow the instructions.
7. Save the Reliability Database (Excel file). You will need it after you have coded the reliability documents.
8. It may be helpful to open the *21st CCLC Revised Codebook* (pdf) and have it handy as you are coding.

Entering Data into Access Database:

1. Open necessary documents from the Dropbox folder.
2. Enter data into your Access database as instructed, **starting with A8** for each document or set of documents.
 - a. Note: A set of documents can also be referred to as a ‘record’. Each record has an ID # and may be comprised of several documents. For example, if we receive three documents related to guidance materials, the coder will be able to code them together in **one** record (using one ID number, e.g. 10012). In the Google Docs spreadsheet, each of the separate documents would be listed individually and ID’d using this format: 10012a, 10012b, 10012c.

First: Conduct Reliability Coding:

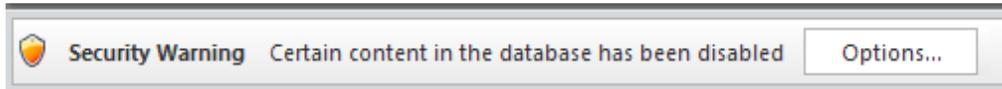
1. Enter data for reliability documents first.
2. Arrange a time to meet with your partner and discuss inter-rater reliability.
3. The **Co-reviewer** should record disagreements and explanations in the reliability spreadsheet (Excel File). Instructions are included in the spreadsheet. There is a worksheet corresponding to each type of document. Please use the appropriate worksheet for each of the three documents.
4. Be sure to enter the document ID # at the top of each worksheet, in the space provided.
5. Through discussion about the discrepant items, the reviewers may reach consensus on a coding decision. The **Lead** reviewer should record final decision data entry into his/her **Access database**.
6. The **Co-reviewer** should send the completed reliability spreadsheet back to Vincent (Vincent@bpacal.com). A specific due date will be provided.
 - a. Please name the saved reliability spreadsheet in this way: *“Full state name your initials Reliability Database.xlsx”*. E.g., *“Alabama NR Reliability Database.xlsx”*.
 - b. **Note:** only the **Co-reviewer** needs to send the Excel file back to Vincent. We only need one completed Reliability Database per state.

Next: Complete Coding for Each State:

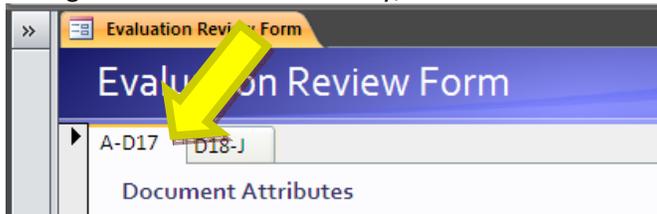
1. The reviewers should code the rest of the state and sub-grantee documents, splitting up the work between them. Projected time allotted for coding: for Lead reviewers = 1.5 days per state; for co-reviewers = 3 days per state.
2. Reviewers should submit their completed Access databases to Vincent (Vincent@bpacal.com). A specific due date will be provided.
3. Use the Google Docs spreadsheet to record your coding progress. After you finish a document (or set of documents), note it in the tracker. Please do this right away, as this provides the team with information about how the task is progressing.
4. **Let Task Leader know ASAP if you will have difficulty completing your coding assignments** by the due dates. We can make rearrangements according to people’s availability; but only if you let us know you need assistance!

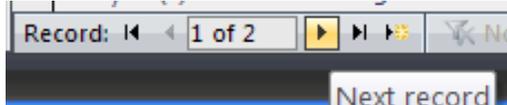
Instructions for using the Access database

1. Open the Evaluation Review Access database. Save it on to your hard drive and rename using this naming convention: *[full state name]_[your initials]* (e.g., Colorado_AB).
2. There will be a security warning towards the top of the screen. Click "Options" then click "Enable this content", and then "ok." **This step is VERY important.** The form will not work correctly if this step is skipped.



3. When starting to enter a new form entry, **ALWAYS** start with the first tab (A-D17).



4. **Start by answering question A8.** You will not be able to fill out other fields until this question is answered.
 - a. **NOTE:** Once an option is selected, you will not be able to modify unless you create a new record.
5. Fill in all relevant fields.
 - a. Fields that do not need to be filled out are grayed out and locked from editing.
 - b. **NOTE:** To reduce data entry errors, you can only enter a 6 digit number (all ID's are 6 digits) in the fields for 'A2. Document ID' and 'A2a. Re-enter Document ID' in the tab "A-D17". This field is automatically populated in the "D18-J" tab. However, please do check that the IDs in field A2 in both tabs have the correct document ID.
6. It is recommended that you finish coding a document in one sitting. However, if you need to stop and continue later, just re-open the database. The form will open on the first document ID in numerical order. Find the correct record (document ID number) by either:
 - a. Clicking on the arrow pointing towards the right until you see the Document ID you need to complete. (NOTE: not the arrow with the star)
 - b. Or, highlighting the Document ID field (A2) and pressing [CTRL] + [F] on your keyboard. Type the Document ID you are searching for in the "find what" section and press "find next." This will bring you to the correct record and you can continue entering data.
7. **NOTE:** There is no need to save in Access. It is automatically saving as you go. Be aware that it is very easy to overwrite data because of this feature.

Appendix G

Detailed Information on 21st CCLC Evaluation Measures Used by SEAs and Sub-Grantees

Measures		Sources of data										
Student Academic Outcome Measures	Number of states using this measure (48)	State Standardized Test	School/Classroom Grade	Teacher Assessment (e.g. Teacher Survey or Questionnaire)	Other Source (including Local and Formative Assessments)							
Math Achievement	38	21	16	10	8							
Reading/ELA Achievement	38	21	16	10	9							
Other Content Area Achievement*	8	2	3	1	3							
Other Measure*	15	1	2	10	6							
Student Non-Academic Outcome Measures	Number of states using this measure (48)	School Records	Teacher Assessment (e.g. Teacher Survey)	Standardized Instrument (e.g. Social Skills Rating Scale)	Student Assessment (e.g. Student Survey)	Parent Assessment (e.g. Parent Survey)	PPICs	Other				
Behavior	38	1	33	0	9	6	3	7				
Disciplinary Incidents	8	3	1	0	1	1	1	5				
School/Classroom Attendance	28	4	17	0	3	3	5	5				
Satisfaction/Attitude Toward School	15	0	12	0	7	3	0	3				
Homework Completion	24	1	19	0	2	3	0	1				
Other*	15	0	9	0	2	3	0	2				
Parent/Family Outcomes	Number of states using this measure (48)	Parent Assessment (e.g. Parent Survey)	Teacher Assessment (e.g. Teacher Survey)	School Records	PPICs	Other						
Satisfaction with Child's School	3	2	0	0	0	0						
Parent Involvement in School/Classroom	4	1	0	0	0	3						
Parent Employment	1	0	0	0	0	1						
Other Measure*	7	3	1	0	0	1						
Process Measures	Number of states using this measure (48)	Teacher assessment (e.g. teacher survey)	Parent assessment (e.g. parent survey)	Student assessment (e.g. student survey)	Staff Assessment (e.g. Survey)	School Records	Focus Groups	Interviews	Observations	Program Tracking/ Attendance Records/PPICs	Other	
Adult-to-Student Ratio	3	1	0	0	1	0	0	0	0	1	3	
Community Partnerships	29	2	0	0	3	0	0	1	0	12	12	
Core Academic Activities	35	3	1	2	2	3	2	1	3	17	11	
Recreational and Enrichment Activities	36	2	1	3	2	0	1	1	3	17	13	
Links to School Day	15	1	0	0	3	0	1	1	2	2	7	
Service Hours Provided	27	1	0	0	0	0	0	1	0	14	7	
Counseling and Mentoring to Students	13	0	0	0	0	0	0	1	0	7	3	
Program Attendance	40	4	0	1	1	0	2	1	0	22	7	
Services to Adults	23	3	3	0	1	0	0	1	0	11	8	
Communications with Parents	12	3	5	0	2	0	1	0	1	1	7	
Parent Involvement in 21st Century Program	21	3	9	0	4	0	0	1	1	4	9	
Parent Satisfaction with 21st Century Program	13	0	11	0	0	0	0	1	0	1	1	
Teacher/Administrator Satisfaction with 21st Century Program	2	1	0	0	0	0	1	0	0	0	1	
Student Satisfaction with 21st Century Program	11	1	0	10	0	0	1	0	1	0	1	
Program Implementation Issues	12	1	1	0	2	0	1	1	1	1	6	
Other Measure*	18	2	0	1	4	1	0	1	0	4	6	

* tabulated from data collected in Option Year 1 and Option Year 2, including updated documents from Base Year.

Measures		Sources of data									
Student Academic Outcome Measures	Percent of SGs using this measure (458)	State Standardized Test	School/ Classroom Grade	Teacher Assessment (e.g. Teacher Survey or Questionnaire)	Other Source (including Local and Formative Assessments)						
Math Achievement	62%	43%	34%		15%	25%					
Reading/ELA Achievement	62%	45%	35%		14%	25%					
Other Content Area Achievement*	9%	45%	45%		12%	21%					
Other Measure*	19%	2%	3%		22%	26%					
Student Non-Academic Outcome Measures	Percent of SGs using this measure (458)	School Records	Teacher Assessment (e.g. Teacher Survey)	Standardized Instrument (e.g. Social Skills Rating Scale)	Student Assessment (e.g. Student Survey)	Parent Assessment (e.g. Parent Survey)	PPICs	Other			
Behavior	52%	9%	50%	3%	14%	13%	2%	14%			
Disciplinary Incidents	21%	29%	6%	0%	1%	1%	1%	9%			
School/Classroom Attendance	47%	27%	31%	0%	2%	7%	3%	17%			
Satisfaction/Attitude Toward School	30%	0%	35%	0%	22%	13%	1%	20%			
Homework Completion	33%	3%	46%	0%	16%	12%	0%	4%			
Other*	19%	3%	49%	1%	30%	22%	0%	40%			
Parent/Family Outcomes	Percent of SGs using this measure (458)	Parent Assessment (e.g. Parent Survey)	Teacher Assessment (e.g. Teacher Survey)	School Records	PPICs	Other					
Satisfaction with Child's School	6%	22%	0%	0%	0%	4%					
Parent Involvement in School/Classroom	9%	28%	10%	8%	0%	33%					
Parent Employment	1%	0%	0%	0%	20%	20%					
Other Measure*	10%	58%	40%	2%	0%	64%					
Process Measures	Percent of SGs using this measure (458)	Teacher assessment (e.g. teacher survey)	Parent assessment (e.g. parent survey)	Student assessment (e.g. student survey)	Staff Assessment (e.g. Survey)	School Records	Focus Groups	Interviews	Observations	Program Tracking/ Attendance Records/PPICs	Other
Adult-to-Student Ratio	16%	1%	0%	0%	0%	1%	0%	1%	7%	32%	7%
Community Partnerships	38%	2%	2%	0%	2%	1%	0%	6%	5%	25%	25%
Core Academic Activities	45%	2%	3%	4%	1%	2%	0%	4%	10%	30%	10%
Recreational and Enrichment Activities	47%	0%	0%	6%	0%	0%	0%	5%	11%	34%	10%
Links to School Day	19%	5%	6%	7%	2%	0%	0%	13%	9%	13%	19%
Service Hours Provided	28%	0%	0%	0%	0%	0%	0%	0%	2%	38%	5%
Counseling and Mentoring to Students	11%	0%	0%	0%	0%	0%	0%	4%	2%	29%	10%
Program Attendance	54%	4%	2%	3%	0%	5%	0%	0%	0%	45%	11%
Services to Adults	29%	0%	4%	1%	0%	1%	1%	0%	2%	36%	6%
Communications with Parents	26%	13%	31%	1%	1%	0%	0%	5%	5%	16%	25%
Parent Involvement in 21st Century Program	33%	4%	25%	3%	1%	0%	0%	7%	7%	25%	11%
Parent Satisfaction with 21st Century Program	31%	11%	58%	1%	0%	0%	0%	1%	1%	10%	3%
Teacher/Administrator Satisfaction with 21st Century Program	13%	21%	0%	2%	2%	0%	0%	3%	0%	3%	2%
Student Satisfaction with 21st Century Program	30%	1%	1%	60%	0%	0%	0%	1%	2%	0%	2%
Program Implementation Issues	30%	12%	9%	0%	1%	0%	1%	9%	18%	14%	20%
Other Measure*	17%	5%	9%	9%	10%	1%	1%	11%	14%	18%	22%

* tabulated from data collected Option Year 1 and Option Year 2, and does not include Base Year sub-grantee data due to question inequities

Measure percentages are calculated by taking the number of sub-grantees that use that measure divided by the total sub-grantees who submitted evaluation reports.

Source percentages are calculated by taking the number of sub-grantees that indicated a particular source divided by the number of sub-grantees who indicated the measure.

Appendix H

Evaluation Framework for 21st CCLC Programs

21st Century Community Learning Centers Grant Monitoring Support

Contract No. ED-04-CO-0027

Task Order No. 0005

For the U.S. Department of Education,
Office of Elementary and Secondary Education

Evaluation Framework for 21st CCLC Programs

April 15, 2011

Submitted to:

Office of Elementary and Secondary Education
U.S. Department of Education
400 Maryland Avenue, SW
Washington, DC 20202

Submitted by:

Berkeley Policy Associates
440 Grand Avenue, Suite 500
Oakland, California 94610-5085
Telephone: 510-465-7884
www.berkeleypolicyassociates.com



Table of Contents

Introduction..... 1

1. Qualified Evaluator..... 1

2. Articulated Program Goals and Measurable Objectives..... 2

3. Design Appropriate for Measuring Program Quality and Effectiveness 3

4. Analysis and Reporting..... 6

5. Use of Evaluation Results..... 8

Evaluation Framework for 21st Century Community Learning Centers (21st CCLC) Programs

Introduction

This document has been developed to serve as a basic framework for the evaluation of 21st Century Community Learning Centers (CCLC) programs. This framework was developed by Berkeley Policy Associates (BPA) in collaboration with the U.S. Department of Education, and is based on current standards of practice, evaluation research¹ and the goals of the 21st CCLC program. The Department has contracted with BPA to support its 21st CCLC monitoring efforts, including providing technical assistance to State Educational Agencies (SEAs) for effective evaluations that can be used to support program improvement. This framework provides a basic structure for addressing both the state requirement to conduct a comprehensive statewide evaluation of the programs and activities provided with 21st CCLC funds, and the states' role in monitoring and supporting evaluation efforts at the local sub-grantee level, as described in the U.S. Department of Education Office of Elementary and Secondary Education's 21st CCLC Non-Regulatory Guidance²:

H-5: State evaluation requirements:

States must conduct a comprehensive evaluation (directly, or through a grant or contract) of the effectiveness of programs and activities provided with 21st CCLC funds. In their applications to the Department, States are required to describe the performance indicators and performance measures they will use to evaluate local programs. State must also monitor the periodic evaluations of local programs and must disseminate the results of these evaluations to the public.

H-6: Evaluation requirements for local grantees:

Each grantee must undergo a periodic evaluation to assess its progress toward achieving its goal of providing high-quality opportunities for academic enrichment. The evaluation must be based on the factors included in the *principles of effectiveness*.³ The results of the evaluation must be: (1) used to refine, improve, and strengthen the program and to refine the performance measures; and (2) made available to the public upon request. Local grantees, working with their SEAs, must evaluate the academic progress of children participating in the 21st CCLC program.

¹ For example: Joint Committee on Standards for Educational Evaluation, *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users*, 3rd Edition, 2010.

² U.S. Department of Education Office of Elementary and Secondary Education, *21st Century Community Learning Centers Non-Regulatory Guidance*, February 2003.

³ As described in Section 4205(b) of the Elementary and Secondary Education Act (ESEA) principles of effectiveness stipulate that programs: (A) be based upon an assessment of objective data regarding the need for before and after school programs (including during summer recess periods) and activities in the schools and communities; (B) be based upon an established set of performance measures aimed at ensuring the availability of high quality academic enrichment opportunities; and (C) if appropriate, be based upon scientifically based research that provides evidence that the program or activity will help students meet the State and local student academic achievement standards.

As the non-regulatory guidance suggests, at the state level, the focus of the statewide comprehensive evaluation of the 21st CCLC program is on evaluating the effectiveness of programs and activities provided with 21st CCLC funds, and SEAs are also responsible for monitoring local evaluation efforts. At the sub-grantee level, the focus is on assessing progress toward providing high quality services, using evaluation results to support program improvement. At both the SEA and sub-grantee levels, evaluation is to be guided by performance measures, and results are to be made available to the public.

This framework is intended for use by SEAs in support of their 21st CCLC grants. SEA coordinators and evaluators can use this framework to plan or assess the status of their comprehensive state-wide evaluations. SEAs can also use this framework to provide technical assistance and guidance to their sub-grantees in conducting local evaluations. This framework describes five key features of effective program evaluations, and gives examples of how these features are operationalized. It is recommended that 21st CCLC evaluations at both the state and local levels include the following five key features:

1. Qualified Evaluator
2. Articulated Program Goals and Measurable Objectives
3. Design Appropriate for Measuring Program Quality and Effectiveness
4. Analysis and Reporting
5. Use of Evaluation Results

1. Qualified Evaluator

To ensure both the quality and the credibility of the evaluation, it is important that evaluations be conducted by a qualified evaluator, either an individual or team of people with appropriate expertise and experience conducting evaluations of education or afterschool programs. This applies to any evaluation study, whether at the SEA or the sub-grantee level.

- Qualified evaluators have formal training in research and/or evaluation methods and have previous experience planning and conducting program evaluations.
 - *Examples of relevant training include:* A Master's degree or Ph.D. in education or a social science discipline, training in rigorous evaluation design and using relevant qualitative and quantitative methodologies such as conducting interviews and focus groups and/or analyzing survey and administrative datasets.
- Qualified evaluators have content knowledge of, and experience evaluating or studying, educational programs, school-based programs, and/or specifically after-school programs.
 - *Examples of relevant knowledge and experience include:* Experience evaluating other 21st CCLC programs or other school or community programs aimed at increasing student academic achievement, experience collecting and analyzing student outcome data (e.g. standardized test scores, grades) and implementation data (e.g., observing classrooms, surveys about program perception, collecting information about program quality).

- Qualified evaluators are independent of the 21st CCLC program thus avoiding any potential or perceived conflict of interest.

2. Articulated Program Goals and Measurable Objectives

It is recommended that evaluations explicitly articulate the goals of the program being evaluated and specify how program effectiveness and progress towards program goals are measured. At the SEA level, program goals align with the overall purposes of the 21st CCLC grant program. At the sub-grantee level, goals and activities are aligned with the state goals but may also reflect local priorities. According to federal statute, the purposes of the 21st CCLC are to:

- (1) Provide opportunities for academic enrichment, including providing tutorial services to help students, particularly students who attend low-performing schools, to meet state and local student academic achievement standards in core academic subjects, such as reading and mathematics;
 - (2) Offer students a broad array of additional services, programs, and activities, such as youth development activities, drug and violence prevention programs, counseling programs, art, music, and recreation programs, technology education programs, and character education programs, that are designed to reinforce and complement the regular academic program of participating students; and
 - (3) Offer families of students served by community learning centers opportunities for literacy and related educational development.⁴
- Program goals reflect a “theory of change⁵” or “logic model⁶” which defines the building blocks that are expected to contribute to the long term outcomes. The three broad purposes stated above embody the theory that providing opportunities for academic enrichment, additional youth development and enrichment services, and literacy services to families will result in better academic outcomes for students.
 - While the goals provide the overall theory or logic of the program, measuring success involves identifying measurable indicators for achieving program goals. Effective evaluations explicitly state and incorporate program goals and objectives in all phases of the process including planning, design, and reporting.
 - SEA and sub-grantee evaluations can address the same basic program goals and evaluation questions, or sub-grantees may supplement the state goals with additional goals that are specific to their local needs.
 - *Examples of state program goals:* Increase students reading skills.

⁴ Part B, Section 4201 (a), *Elementary and Secondary Education Act (ESEA) of 1965*, as amended.

⁵ Weiss, Carol, *New Approaches to Evaluating Comprehensive Community Initiatives*, Aspen Institute Roundtable on Community Change, 1995

⁶ Rogers, P.J. 'Logic models' in Sandra Mathison (ed) *Encyclopedia of Evaluation*. Beverly Hills, CA: Sage Publications, 2005.

- *Examples specific sub-grantee goals:* improve communication with teachers at host school in order to strengthen linkages between 21st CCLC activities and school day lessons. Increase middle schools students' academic performance in English language arts.
- Objectives are specific statements that include measurable indicators for reaching the goals.
 - *Example of state goal and measurable objective:*
Goal: To improve student achievement in math.
Objective: To increase the percentage of students participating in 21st CCLC achieving grade level proficiency in math by 10% on the state math assessment
 - *Example of sub-grantee goal and measurable objective:*
Goal: In a community where violence and behavior are particular challenges, a program goal may be to improve school safety.
Objective: Reduce student disciplinary incidents among students participating in the 21st CCLC program by 15%.

3. Design Appropriate for Measuring Program Quality and Effectiveness

It is recommended that evaluations use designs that are systematic, well-documented, and measure progress towards achieving program goals and objectives. Designs should be sufficiently rigorous to measure the quality of implementation and to support a reasonable hypothesis that the program is, or is not, contributing to achieving the desired outcomes.

Comprehensive and effective evaluation designs include the following components:

- **Evaluation Questions:** Evaluations explicitly articulate the purpose or questions that the evaluation is designed to address.
 - *Examples of evaluation questions include:* Is the statewide 21st CCLC program reaching the target population? How well are sub-grantee activities aligned with the goals and objectives of the state's 21st CCLC program? Is the 21st CCLC program contributing to an increase in reading scores for student participants?
- **Measures:** As specified in the non-regulatory guidance, SEAs are required to specify performance indicators and performance measures that are used to evaluate sub-grantee programs. In some cases SEAs may specify a uniform set of performance measures statewide. In other cases, SEAs may want to allow sub-grantees the flexibility to choose between specific performance measurement options, or supplement a core set of statewide measures with additional measures specific to the objectives of their local programs. Comprehensive evaluations include both process and outcome measures.
 - *Process measures include measures of implementation fidelity (was the program implemented as intended?), program quality, and program intensity or dosage. Examples of process measures include:* program attendance, types

- of academic or enrichment activities, frequency of these activities, or student/parent/staff satisfaction with the program.
- *Outcome measures are measures of behavior or performance (usually of students) that the program is designed to improve. Examples of outcome measures include:* standardized test scores, grades, school attendance records, rates of suspension and other disciplinary actions based on district data.
 - **Integrating Process and Outcome Measures:** Comprehensive evaluations combine process and outcome measures. Outcome measures identify “what” has been achieved. Process measures supplement outcome measures with information about “how” programs are implemented. Evaluations designed to combine these two types of measures can explore “why” programs may be more successful in some areas than others and what strategies might be effective in addressing program weaknesses. This approach results in an evaluation that is designed to support program improvement.
 - *Example of integrating process and outcome measures at the state level:* The state evaluator may find that some sub-grantees have shown greater student achievement gains than others. Review of sub-grantees’ Quality Improvement Process reports shows that several sub-grantees with lower student achievement gains have identified the need to increase attendance. Such findings could help the SEA identify a need for TA to sub-grantees on successful strategies for increasing and maintaining high student attendance.
 - *Example of integrating process and outcome measures at sub-grantee level:* an evaluator may find that reading scores have significantly increased for 21st CCLC participants but math scores have remained stable. Through focus groups, students may reveal that staff members have found ways to make reading groups fun and have created ways to keep student engaged. Such findings could help programs identify successful practices and apply those strategies to math activities, in order to increase student interest and engagement in math. Such information will be uncovered only by asking the right evaluation questions, and linking them to program goals and objectives.
 - **Rigorous Design:** Using the most rigorous evaluation design that is feasible will provide the best quality evaluation. Simply reporting achievement on performance measures without some analysis of how the program’s achievements compare to the results that would have been achieved in the absence of the program is not considered to be a rigorous design. Even comparing program outcomes from one year to the next is not considered a rigorous design, if the comparison does not either follow the same group of students over time or control for differences in the characteristics of students from one year to another. The following are examples of different types of rigorous evaluation designs:
 - ***Experimental (randomized control trial) design:*** The only way to truly determine causality (if the outcomes achieved are attributable to the program) is through an experimental study using random assignment. In such studies, students or schools would be recruited to (or express interest in) the program and then be randomly assigned to either a program or control group.

Experimental designs can be challenging to implement and costly, so they may not be feasible for many grantees.

- **Comparison group designs:** quasi-experimental designs compare outcomes between two groups but do not randomly assign individuals to the two groups. Some examples of comparison groups include:

Comparison with district or state averages. This is the simplest type of comparison, and while it does not take into account potential differences between participants and non-participants, it does use district or state averages as a kind of benchmark against which the program can gauge its relative success.

Comparison with a similar group or community. For example, outcomes for adolescents in a Boys and Girls Club in one neighborhood might be compared with outcomes for adolescents in another Boys and Girls Club in a similar neighborhood.

Comparison with matched individuals. For example, comparisons might be made between students involved in a program and students not involved in that program who are matched to program participants in terms of key variables such as their age, gender, race, grades, receipt of free/reduced lunches, absenteeism, and other characteristics.

Use of statistical methods to control for measured and unmeasured variables. For example, pre-test and post-test scores for participants can be compared with scores for a comparison group in that school or agency the year before the program opened, controlling for student characteristics.

Regression discontinuity design is the most rigorous quasi-experimental design, but it can be used only under very specific conditions. If students are admitted to a program based on exceeding a “cutoff” score on a consistent pre-program measure (such as income, test scores, or grade point average), and if an outcome measure is available for both admitted and non-admitted students (those above and below the “cutoff”), this design may be possible.

- **Single group pre-/post-test design:** This design is the least rigorous and while it does provide a measure of change for the individual student participants, it cannot be used to infer that the change is due to the program.

*[For more information on social science research designs used to evaluate educational programs, see: <http://www.socialresearchmethods.net/kb/design.php> or Chapter A6 (p. 201) of *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users, 3rd Edition (2010)*]*

- **Stakeholder Representation:** To produce results useful for program improvement, evaluations collect data from all relevant stakeholders, that is, representatives of all of the key parties who participate in or are directly affected by the program. These include students, teachers, parents, program staff and community partners.
 - *Examples of collecting data from key stakeholders:* Interviews, focus groups, or surveys of students, teachers, and parents. At the SEA level, an evaluator may

interview the state 21st CCLC staff for their input about the program. At the sub-grantee level, an evaluator might interview or survey relevant community partners.

- **Proper Documentation:** Evaluations document their designs, methods, sources of data and outcomes. Evaluators should describe the methodology used, data collection strategies and instruments used, analysis plan employed, and any assumptions made. Procedures and methods should be systematic and purposeful.
 - *Example:* For a sub-grantee evaluation, if 10 students were interviewed at a school about the program, the evaluation should describe how and why those students were selected to be interviewed. Those students should be described (without identifying the individual students); the reader should get a sense of whether those students are representative of other 21st CCLC participants or other students at the school.
- **Data Management:** Evaluations use information management and storage procedures that maintain the accuracy of data.
 - *Example:* Evaluators ensure that data files are backed up; evaluators can have research assistants double enter data for accuracy, all data elements and files are carefully and accurately labeled, all data and artifacts (interviews, documents collected, etc) are securely stored in the evaluator's office or other safe facilities. Quality control checks are in place to ensure that data are managed and analyzed carefully and accurately. Analysis procedures are documented and accessible to the program or a third party should they be needed for replicating the analysis at a later time.
- **Ethical Standards:** Evaluators maintain the confidentiality of participants and use methods and procedures that meet ethical standards.
 - *Example:* Experienced evaluators are familiar with ethical standards and evaluation participants' rights in their state and local context. Students are not individually identified in evaluation reports, and informed consent is obtained if students or parents will be interviewed. For more information, evaluators may visit U.S. Department of Health and Human Services, Office for Human Research Protections at <http://www.hhs.gov/ohrp/>.

4. Analysis and Reporting

Data collected is analyzed to answer the evaluation questions, and evaluation reports document both the evaluation methods and results so that findings and conclusions can be clearly articulated and shared with relevant stakeholders.

- Evaluation reports use data analysis procedures that can statistically determine if an effect is found for program participants.

- *Examples of statistical analyses:* regression, analysis of variance, or t-tests accompanied by significant testing to determine whether any differences found are real differences or are due to random error.
- Evaluation reports include an explanation of how the findings are linked to program goals and evaluation questions.
 - *Example of linking findings to state program goals and evaluation:* If one program goal is to improve reading skills among student participants, the evaluation report would include a question such as, “Did 21st CCLC program contribute to improvement in reading scores for participants?” The report would then describe how the necessary information was gathered and analyzed. The findings would interpret the analysis to state whether a program effect was indicated.
 - *Example of linking findings to sub-grantee goals and evaluation questions:* If one of the sub-grantee’s goals is to reduce disciplinary incidents, the evaluation report would include a question such as “How does the number of disciplinary incidents during the current year compare with the previous year?” Then, rather than simply presenting the number of disciplinary incidents, the findings would be presented in terms of whether the goal of reducing disciplinary incidents had been achieved.
- Evaluation reports describe the characteristics of the sample used to evaluate the program.
 - *Examples of descriptions of sample:* A statewide evaluation might provide information on how many students, centers, or sub-grantees are in the sample. It might also include information on the demographic characteristics of the students or the size or type of programs (e.g. faith-based organization, school district).
- Evaluation reports include a description of the data collection methods, including response rates, and sources of information.
 - *Example of description of methods:* An evaluation that includes a teacher survey would describe the survey instrument, to whom the survey was administered or given, and who completed the survey. It would also provide a response rate (how many surveys were returned and analyzed in comparison to the number of surveys distributed.)
 - Evaluation reports describe any limitations associated with their designs or methods, and their associated limitations in interpreting their findings.
- Evaluation reports provide recommendations linked to program goals based on findings from the data, including identified strengths and areas for improvement.

5. Use of Evaluation Results

As mentioned earlier, the non-regulatory guidance requires that sub-grantees use evaluation results to refine, improve, and strengthen their program and to refine the performance measures. Effective use of evaluation results includes:

- Creating and carrying out an improvement plan based on the findings from the evaluation.
 - *Examples of SEA uses of results:* identify technical assistance needs of sub-grantees (e.g. strategies for increasing attendance); set academic performance targets for the coming year.
 - *Examples of sub-grantee uses of results:* identify program needs (e.g., better recruitment of participants); prioritize which academic programs to emphasize to meet academic performance targets in the coming year.
- Engaging the evaluator in the program improvement process.
 - *Example of state level evaluator role in improvement process:* attend management team meetings to consult with the management team on the interpretation and use of evaluation results to identify sub-grantee TA needs and set performance targets for the coming year.
 - *Example of sub-grantee level evaluator role in improvement process:* facilitate meetings with program staff to engage them in the process of synthesizing evaluation findings and developing action steps.