Harvard Family Research Project
August 2002

Evaluating Municipal Out-of-School Time Initiatives—
Learning From the Field[1]

For many cities, out-of-school time (OST) programming is uncharted territory. Because of its newness, relatively little is known about OST best practices, program implementation, cost effectiveness, and impact. However, in these times of decreasing public resources and increasing and competing demands for public investments, it is necessary for funders, policymakers, and their constituents to know which investments are effective and how programs can be improved. This situation makes it imperative that those developing policies and implementing OST programs are able to learn, over time, whether OST investments are working, how they can be improved, and whether they should be expanded. In other words, cities need to grapple with the issue of *evaluation*.

To help inform municipal leaders as they craft evaluations of their OST initiatives, the Harvard Family Research Project (HFRP) reviewed and analyzed evaluation reports from 15 out-of-school time programs/initiatives actively engaged in evaluation. This brief provides a thumbnail sketch of the evaluation questions, methods, approaches, and indicators being used by cities across the country to expand our knowledge base about out-of-school time programs.[2] To accompany this overview, HFRP prepared a companion summary table of the 15 evaluation efforts described herein, available on the Internet at www.gse.harvard.edu/hfrp/projects/afterschool/resources/munimatrix.pdf (Acrobat file).

Overview of City-Level Programs/Initiatives
The 15 city-level OST programs/initiatives included in our review are:[3]

- ADEPT Drug and Alcohol Prevention Project, New Orleans
- After-School Achievement Program, Houston
- District of Columbia 21st Century Community Learning Center Summer Program, Washington, D.C.
- Extended-Day Tutoring Program, Memphis
- Fort Worth After School Program, Fort Worth
- Los Angeles' Better Educated Students for Tomorrow, Los Angeles
- Making the Most of Out-of-School Time, Boston, Chicago, Seattle
- New York City Beacons Initiative, New York City

---

[1] Prepared for The National League of Cities' *Your City's Families* Conference Pre-Conference Institute Municipal Leadership for Expanded Learning Opportunities, May 2002
[2] A longer brief, with recommendations for municipal leaders, will be available in 2003.
[3] For more information on each of these programs, visit the Harvard Family Research Project Out-of-School Time Evaluation Database at: www.gse.harvard.edu/hfrp/projects/afterschool/evaldatabase.html.

- Project for Neighborhood Aftercare School-Based After School Program, Nashville
- San Diego's "6 to 6" Extended School Day Program, San Diego
- San Francisco Beacons Initiative, San Francisco
- The After School Corporation, New York City
- Totally Cool Totally Art, Austin
- Virtual Y, New York
- YouthPlaces Initiative, Baltimore

Examining their size, scope, and program mission reveals that these 15 city-level initiatives present a diverse picture of municipal out-of-school time efforts. The initiatives range in size from quite small, 200 participants per year in Nashville's Project for Neighborhood Aftercare School-Based After School Program, to 76,000 youth and 33,000 adults served annually by the New York City Beacons Initiative. Some have been in operation for almost 15 years, such as Los Angeles' Better Educated Students for Tomorrow (LA's BEST), started in 1988; others are recent initiatives, such as Baltimore's YouthPlaces, started in 1999. Initiative missions range from providing instruction in visual arts (Totally Cool, Totally Art) to fostering improved literacy outcomes (Virtual Y) to building a system of quality out-of-school time care for children and youth (Making the Most of Out-of-School Time).

Despite their differences, this set of initiatives has some important commonalities:

- All of the multi-site initiatives are based in urban areas.
- Most receive core funding from local government.
- All of the initiatives chose to conduct evaluations of their programs in order to improve their programs and/or assess program impact.

What Do Cities Want to Know About Their OST Programs?
Across all evaluations, the questions that cities sought to answer predominantly fall into a few broad categories, listed below in order of the most common question to the least common question:

- What are we doing and how could we do it better?
- What is the impact of the program/initiative?
- What is the quality of the program/initiative?
- What are the characteristics/perceptions/experiences of participants?
- What are the costs of the program/initiative?
- What factors affect the impact of the program/initiative?

This list of evaluation questions reflects the dual purposes for which cities evaluate their OST initiatives—both to collect data for program improvement and to create a data-driven argument for sustainability based on proven results.

What Types of Evaluation Are Cities Conducting?
City initiatives are conducting both formative and summative evaluations in order to answer a broad range of evaluation questions.

Formative Information
Formative evaluations are conducted during program implementation in order to provide information that will strengthen or improve the program being studied—in this case, the out-of-school time program or initiative. Formative evaluation findings typically point to aspects of program implementation that can be improved for better results—how services are provided, how staff are trained, or how leadership and staff decisions are made.

All of the initiatives in this review conducted formative evaluations to better understand the initiatives themselves. Of these formative evaluations, most collected data to document: activity implementation; recruitment and participation; program context/infrastructure (including transportation); and staffing/training patterns, issues, and needs. Over half of the evaluations collected data on participant satisfaction and parent/community involvement. Fewer than half collected data on costs/revenues and systemic infrastructure (including partnerships).

Summative Information
Summative evaluations are conducted either during or at the end of a program's implementation. They determine whether a program's intended outcomes have been achieved—in this case, the out-of-school time program or initiative's goals. Summative evaluation findings typically judge the overall effectiveness or "worth" of a program based on its success in achieving its outcomes, and can be important in determining whether a program should be continued. Summative outcomes can be short-term or longer term, depending on the purpose of the evaluation.

Almost all of the initiatives in this review also conducted summative evaluations to examine the various impacts of the initiative on participants and the community. Of these summative evaluations, most collected data on academic and youth development outcomes. Fewer than one quarter of the evaluations collected data on family, community, prevention, systemic, and workforce development outcomes.

It is important to note that while many city initiatives are conducting summative evaluations, which by definition means they are collecting outcomes data, they are primarily doing so employing *non-experimental* evaluation designs. While this enables cities to make summary statements about participant outcomes, and demonstrate program "worth," the use of non-experimental design limits the ability of evaluators to determine if outcomes are actually a result of the OST initiative and, therefore, to make statements about the effectiveness of an overall program/initiative. Using comparison or control groups, as is done with experimentally and quasi-experimentally designed evaluations, does allow for this determination of causality, and ultimately, judgments about the effectiveness of the OST initiative. One-third of the city initiatives included in this review employed quasi-experimental designs to assess academic outcomes of the participants.

What Data Collection Methods Do Cities Use?
City OST initiatives are using many different methods to gather data about the functioning and impact of their programs. Data collection methods can be understood as the way in which evaluators approach answering evaluation questions. Most evaluated city initiatives use multiple data collection methods, including document review,

interviews/focus groups, observation, secondary sources/data review, surveys/questionnaires, and tests/assessments. Each city initiative studied here uses an average of four different data collection methods. The list below shows the number of evaluations that used each data collection method across the 15 city programs/initiatives.

- Interviews/focus groups (12 of 15)
- Observation (11 of 15)
- Surveys/questionnaires (11 of 15)
- Secondary sources/data review (10 of 15)
- Document review (9 of 15)
- Tests/assessments (4 of 15)

The most common method used by city OST initiatives is the interview/focus group. This is closely followed by observation and surveys and questionnaires. All of these methods allow evaluators to gather information from program participants and stakeholders about their experiences with the OST program and their perceptions of the OST program. Many programs are also using document and data review, in which case the evaluator uses existing documents and data to provide details about everything from program rules and regulations to academic performance to rates of absenteeism, to name a few. Fewer than one-third of the evaluated city initiatives included here used tests/assessments to assess program impact.

What Indicators Do Cities Use to Measure Results?

Findings reported across the evaluations provide a broad range of examples of the indicators that cities use to measure results. Our analysis classifies types of indicators used to measure two key outcome domains that are the most frequently measured and used to make claims about the effectiveness of OST programs—academic achievement and youth development. Table 1 lists the range of indicators used to measure academic achievement and youth development, and the data sources used to obtain information about the measure. Table 1 illustrates that there are many ways to define and measure academic achievement and youth development. Further, it reveals that most city-level OST evaluations rely on qualitative reporting by parents, program participants, principals, and school-day teachers to assess participant outcomes. Very few evaluations use standardized assessment measures of student achievement; even fewer use validated assessments of participant behavior.

Table 1: Measures and Data Sources for Outcome Areas[4]

Outcome Area	Data Source
Academic Achievement	
▪ Academic performance in general	▪ Parent report, principal report
▪ Attendance/absenteeism	▪ School records, parent report, principal report
▪ Attendance in school related to level of program participation	▪ School records
▪ Attendance in school related to achievement	▪ School records, standardized tests
▪ Attitude toward school	▪ Child report
▪ Behavior in school[5]	▪ Standardized behavior scales by teachers
▪ Child's ability to get along with others	▪ Parent report
▪ Child's liking school	▪ Parent report
▪ Child's communication skills	▪ Parent report
▪ Child's overall happiness	▪ Parent report
▪ Cooperation in school	▪ Child report
▪ Effectiveness of school overall	▪ Principal report
▪ Effort grades	▪ School records
▪ English language development	▪ Child report
▪ Expectations of achievement and success	▪ Child report, teacher report
▪ Family involvement in school events	▪ Principal report
▪ Grade point average	▪ School records
▪ Grades in content areas (math, reading, etc.)	▪ School records, parent report
▪ Homework performance	▪ Parent report, principal report
▪ Learning skills development	▪ Teacher report
▪ Liking school more	▪ Child report
▪ Motivation to learn	▪ Parent report, teacher report
▪ Reading	▪ Child report, principal report, test scores
▪ Safety—viewing school as a safe place	▪ Child report
▪ Scholastic achievement assessed by knowledge about specific subjects	▪ Parent report
▪ Standardized test scores	▪ SAT-9, state assessments (TCAP)
Youth Development	
▪ Adults in the OST program care about youth	▪ Child report
▪ Awareness of community resources	▪ Child report
▪ Behavior change toward new program component	▪ Parent and child report
▪ Child's self-confidence	▪ Parent report
▪ Exposure to new activities	▪ Principal report
▪ Facing issues outside of OST program	▪ Child report
▪ Interaction with other students in OST	▪ Child report
▪ Interest in non-academic subjects (art, music, etc.)	▪ Child report
▪ Leadership development/opportunities	▪ Child report
▪ Opportunities to volunteer	▪ Child report
▪ Productive use of leisure time	▪ Child report
▪ Sense of belonging	▪ Child report
▪ Sense of community	▪ Child report
▪ Sense of safety	▪ Child report
▪ Sources of support for youth	▪ Child report

---

[4] Compiled from a review of findings from 26 city-level evaluation reports; for brevity, "child" refers to youth of any age participating in the OST program.
[5] School behaviors included in the scales are: frustration tolerance, distraction, ignoring teasing, nervousness, sadness, aggression, acting out, shyness, and anxiety.

Conclusion

This overview of city-level out-of-school time evaluations illustrates the variety of approaches, methods, and indicators that cities across the country are using to collect data for program improvement and to demonstrate the effectiveness of OST initiatives. It also reveals that cities have many examples to draw from as they begin to craft their own evaluations. There is no formula for the evaluation of OST initiatives, but with good examples from other cities, the task of crafting an evaluation that best matches an initiative's goals is realistic.

Prepared by: Priscilla M. D. Little and Flora Traub, Harvard Family Research Project

For more information on the evaluation of municipal out-of-school time initiatives, please visit the Harvard Family Research Project OST website, at www.gse.harvard.edu/hfrp/projects/afterschool/about.html, or contact our organization at:

Harvard Family Research Project
Harvard University Graduate School of Education
3 Garden Street
Cambridge, MA 02138
Tel: 617-495-9108
Email: hfrp@gse.harvard.edu
Website: www.hfrp.org

Appendix A: Glossary of Selected Evaluation Terms

Accountability
A public or private agency, such as a state education agency, that enters into a contractual agreement to perform a service, such as administer 21st CCLC programs, will be held answerable for performing according to agreed-on terms, within a specified time period, and with a stipulated use of resources and performance standards.

Benchmark
(1) An intermediate target to measure progress in a given period using a certain indicator. (2) A reference point or standard against which to compare performance or achievements.

Data Collection Methods
Document Review: A review and analysis of existing program records and other information collected by the program. Information analyzed in a document review is not gathered for the purpose of the evaluation. Sources of information for document review include information on staff, budgets, rules and regulations, activities, schedules, attendance, meetings, recruitment, and annual reports.

Interviews/Focus Groups: Conducted with evaluation and program/initiative stakeholders, including: staff, administrators, participants and their parents or families, funders, and community members. Can be conducted in person or over the phone. Questions posed are generally open-ended. The purpose of interviews and focus groups is to gather detailed descriptions, from a purposeful sample of stakeholders, of the program processes and the stakeholders' opinions of those processes.

Observation: An unobtrusive method for gathering information about how the program/initiative operates. Observations can be highly structured, with protocols for recording specific behaviors at specific times, or unstructured, taking a more casual "look-and-see" approach to understanding the day-to-day operation of the program. Data from observations are used to supplement interviews and surveys in order to complete the description of the program/initiative and to verify information gathered through other methods.

Secondary Source/Data Review: Sources include data collected for other similar studies for comparison, large data sets such as the Longitudinal Study of American Youth, achievement data, court records, standardized test scores, and demographic data and trends. Data are not gathered with the purposes of the evaluation in mind; they are pre-existing data that inform the evaluation.

Surveys/Questionnaires: Conducted with evaluation and program/initiative stakeholders. Usually uses a highly structured interview process in which respondents are asked to choose answers from those predetermined on the survey and administered on paper, through the mail, or more recently, through email and on the Web. The purpose of surveys/questionnaires is to gather specific information from a large, representative sample.

Tests/Assessments: Data sources include standardized test scores, psychometric tests, and other assessments of the program and its participants. These data are collected with the purposes of the evaluation in mind.

Evaluation Design
Experimental Design: Experimental designs all share one distinctive element—random assignment to treatment and control groups. Experimental design is the strongest design choice when interested in establishing a cause-effect relationship. Experimental designs for evaluation prioritize the impartiality, accuracy, objectivity, and validity of the information generated. These studies look to make causal and generalizable statements about a population or impact on a population by a program or initiative.

Non-Experimental Design: Non-experimental studies use purposeful sampling techniques to get "information-rich" cases. Types include: case studies, data collection and reporting for accountability, participatory approaches, theory-based/grounded-theory approaches, ethnographic approaches, and mixed method studies.

Quasi-Experimental Design: Most quasi-experimental designs are similar to experimental designs except that the subjects are not randomly assigned to either the experimental or the control group, or the researcher cannot control which group will get the treatment. Like the experimental designs, quasi-experimental designs for evaluation prioritize the impartiality, accuracy, objectivity, and validity of the information generated. These studies look to make causal and generalizable statements about a population or impact on a population by a program or initiative. Types include: comparison group pre-test/post-test design, time series and multiple time series designs, non-equivalent control group, and counterbalanced designs.

Formative/Process Evaluation
Formative evaluations are conducted during program implementation in order to provide information that will strengthen or improve the program being studied—in this case, the after school program or initiative. Formative evaluation findings typically point to aspects of program implementation that can be improved for better results, like how services are provided, how staff are trained, or how leadership and staff decisions are made.

Indicator
An indicator provides evidence that a certain condition exists or certain results have or have not been achieved. Indicators enable decision makers to assess progress towards the achievement of intended outputs, outcomes, goals, and objectives.

Performance Measurement (also called Performance Monitoring)
According to the U.S. Government Accounting Office, it is "the ongoing monitoring and reporting of program accomplishments, particularly progress toward pre-established goals" (sometimes also called outcomes). Performance measurement is typically used as a tool for accountability. Data for performance measurement is often tied to state indicators and is part of a larger statewide accountability system.

Summative/Outcome Evaluation
Summative evaluations are conducted either during or at the end of a program's implementation. They determine whether a program's intended outcomes have been achieved. Summative evaluation findings typically judge the overall effectiveness or "worth" of a program based on its success in achieving its outcomes, and are particularly important in determining whether a program should be continued.