



Harvard Family Research Project  
**Selected Evaluation Terms**

March 2002

By Priscilla Little, Harvard Family Research Project

Evaluation terminology can be confusing to even the most seasoned of evaluators. This resource provides commonly accepted definitions to evaluation terms frequently used in the out-of-school time field. For “real life” examples of how these terms are used, check our Out-of-School Time Evaluation Database, currently offering detailed evaluation profiles of over 20 out-of-school time programs nationwide. This resource is available at [www.gse.harvard.edu/hfrp/projects/afterschool/evaldatabase.html](http://www.gse.harvard.edu/hfrp/projects/afterschool/evaldatabase.html).

At the end of this list are answers to three common evaluation questions:

- What is the difference between performance measurement and program evaluation?
- What are my evaluation design options?
- What are the differences between quantitative and qualitative data?

**Accountability:** Accountability means that a public or private agency, such as a state education agency, that enters into a contractual agreement to perform a service, such as administer 21<sup>st</sup> Century Community Learning Center 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:*

- This is a review and analysis of existing program records and other information collected by the program. The information analyzed in a document review was 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:*

- Interviews and focus groups are conducted with evaluation and program/initiative stakeholders. These include, but are not limited to, staff, administrators, participants and their parents or families, funders, and community members.
- Interviews and focus groups can be conducted in person or over the phone. Questions posed in interviews and focus groups are generally open-ended and responses are documented in full, through detailed note-taking or transcription.
- 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:*

- Observation is 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:*

- These 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.
- Like the information analyzed in a document review, these data were not gathered with the purposes of the evaluation in mind; they are pre-existing data that inform the evaluation.

#### *Surveys/Questionnaires:*

- Surveys and questionnaires are also conducted with evaluation and program/initiative stakeholders. These are usually administered on paper, through the mail, in a highly structured interview process in which respondents are asked to choose answers from those predetermined on the survey, or more recently, through email and on the Web.
- The purpose of surveys/questionnaires is to gather specific information—often regarding opinions or levels of satisfaction, in addition to demographic information—from a large, representative sample.

#### *Tests/Assessments:*

- These 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. For example, the administration of achievement tests at certain intervals to gauge progress toward expected individual outcomes documented in the evaluation.

### **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.
- Non-experimental evaluation designs 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 of quasi-experimental designs include: comparison group pre-test/post-test design, time series and multiple time series designs, 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 out-of-school time 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):** “The ongoing monitoring and reporting of program accomplishments, particularly progress toward pre-established goals” (sometimes also called outcomes).<sup>1</sup> 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—in this case, the out-of-school time program or initiative's outcomes. 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.

---

<sup>1</sup> U.S. Government Accounting Office, April 1998.

## A Few Questions Explained

### 1. What is the difference between performance measurement and program evaluation?

	<b>Performance Measurement</b>	<b>Program Evaluation</b>
<b>Purpose</b>	Provides a broad, shallow snapshot of program functioning. Typically answers the question of whether a program has achieved its objectives, expressed as measurable performance standards.	Provides a narrower, deeper examination of program functioning. Typically answers questions of why a program worked, unintended benefits or consequences of a program, and how a program might be improved or changed.
<b>Components</b>	Identification of program goals or outcomes, indicators to measure progress, and regular collection and reporting of data.	Collection of broader range of information on program performance and its context. Information often includes both qualitative and quantitative data.
<b>Scope</b>	Usually involves data collection from all sites.	Usually involves data collection from only a subset of sites.
<b>Timeframe</b>	Annually, or at least at pre-determined intervals.	As needed.
<b>Uses</b>	To examine progress over time, to compare sites, to understand progress toward pre-established outcomes. Can serve as an early warning system to management and a tool for improving accountability to the public.	The more in-depth nature of program evaluation allows for an overall assessment of whether the program works and identification of adjustments that may improve its results. Program evaluation is also used to determine whether a program "caused" outcomes to be achieved.

## 2. What are the main features and trade-offs in design choice?

	Main Feature	Benefits/Trade-Offs
<b>Experimental Design</b>	Random assignment of individuals to either treatment (i.e., an out-of-school time program) or control groups (i.e., no out-of-school time program); groups are usually matched on general demographic characteristics and compared to each other to determine program effects.	The strongest design choice when interested in establishing a cause-effect relationship. Experimental designs prioritize the impartiality, accuracy, objectivity, and validity of the information generated. They allow for causal and generalizable statements to be made about a population or impact on a population by a program.
<b>Quasi-Experimental Design</b>	Features non-random assignment of individuals to treatment and comparison groups, as well as the use of controls to minimize threats to the validity of conclusions drawn. Often used in real-life situations when it is not possible to use random assignment.	Quasi-experimental designs prioritize the impartiality, accuracy, objectivity, and validity of the information generated. However, non-random assignment makes causal and generalizable statements harder to ascertain than when using an experimental design.
<b>Non-Experimental Design</b>	No use of control or comparison groups; typically relies on qualitative data sources such as interviews, observation, and focus groups.	Non-experimental designs are helpful in understanding participants' program experiences and in learning in detail about program implementation. No causal or generalizable conclusions can be drawn using a non-experimental design.

## 3. What is the difference between quantitative and qualitative data?

	Definition and Uses	Methods
<b>Quantitative Data</b>	<ul style="list-style-type: none"> <li>▪ Numeric information that is subject to statistical analyses</li> <li>▪ Can be used to compare outcomes associated with an intervention</li> </ul>	<ul style="list-style-type: none"> <li>▪ Tests/assessments</li> <li>▪ Secondary source/data review (i.e., pre-existing data sources)</li> <li>▪ Surveys/questionnaires</li> </ul>
<b>Qualitative Data</b>	<ul style="list-style-type: none"> <li>▪ Text-based information, collected systematically</li> <li>▪ Can be used to understand how a program operates and how participants experience the program</li> </ul>	<ul style="list-style-type: none"> <li>▪ Document review</li> <li>▪ Interviews</li> <li>▪ Focus groups</li> <li>▪ Observation</li> </ul>