

## Glossary of Key Terms

---

**Assessment:** A type of data collection which may be designed to examine the country or sector context to inform project design, or an informal review of projects. (Source: USAID new evaluation policy).

**Baseline study:** An analysis describing the situation prior to a development intervention, against which progress can be assessed or comparisons made.

**Before-and-after design:** Design that is done by comparing key measures after the intervention began with measures taken before the intervention began. Pre-tests and post-tests are common before-and-after measures (the “before” measure often is called the baseline). The collection of baseline data is also known as Baseline study (see above). Note: Before-and-after design is a weak quasi-experimental design and it is insufficient by itself to demonstrate that the intervention alone caused the change observed.

**Bias:** The extent to which a measurement, sampling, or analytic method systematically underestimates or overestimates the true value of a variable or attribute.

**Causal tracing strategies:** A type of nonexperimental design that consists of arguments for causal relationships based on theory of change models and logically ruling out alternative or rival explanations.

**Cause and effect questions:** Questions that seek to determine what difference the intervention makes. In effect, Cause and Effect Questions ask if the desired results have been achieved and whether it is the intervention that has caused the results. Cause and effect questions often imply before & after or with & without comparisons.

**Case study design:** A nonexperimental design that provides an in-depth comprehensive description and understanding of an intervention as a whole and in its context.

**Cluster Evaluation:** Evaluation that looks at groups of similar or related interventions. A cluster evaluation is not intended to determine efficacy or accountability of an intervention. The intent is to determine what happened across the clusters and to ascertain common themes and lessons learned. Information is reported only in aggregate and stakeholder participation is key.

**Control group:** A randomly selected group that does not receive the services, products or activities of the program being evaluated.

**Control variables:** Extraneous factor that is kept constant so as to minimize any possible effect it might have on an intervention's outcome. Control variables allow the evaluator to rule out some alternative explanations even when random assignment is not possible.

**Context of an Evaluation (Evaluation Context):** The extent to which aspects of a program's setting are considered in the evaluation. Qualitative inquiry is contextual, recognizing that a unique array of contexts influences the program and its outcomes. It may refer to the local and national context within which the evaluation will be implemented. Examples include the economic context, political context, policy, legal and administrative contexts of an evaluation.

**Counterfactual:** A hypothetical statement of what would have happened (or not) had the program not been implemented.

**Descriptive Questions:** "snapshot" questions that ask of what is. They seek to understand or describe a program or process or attitudes towards it. Descriptive Questions are straight forward questions (who? what? where? when? how? how much/many?). They solicit the opinion of beneficiaries.

**Developmental Evaluation:** an evaluation approach developed by Michael Quinn Patton for programs and situations that are highly emergent (difficult to plan and predict), highly dynamic and rapidly changing, and where relationships are interdependent and non-linear rather than simple and linear (cause-effect).

**Evaluation approach:** A way of looking at or conceptualizing an evaluation in a general way. An evaluation approach often incorporates a philosophy and set of values. Typical evaluation approaches might include "utilization-focused evaluation," "goal-based evaluation," "multi-site evaluation" or "meta-evaluation." The choice of evaluation approach depends largely on the context. Approaches are not necessarily mutually exclusive, and evaluations may combine elements of two or more approaches. The same planning steps must be taken regardless of approach including, defining evaluation questions, identifying measures, collecting and analyzing data and reporting and using findings. See Road to Results, Chapter 5, for more examples of evaluation approaches.

**Evaluability assessment:** A study conducted to determine a) whether the program is at a stage at which progress towards objectives is likely to be observable; b) whether and how an evaluation would be useful to program managers and/or policy makers; and, c) the feasibility of conducting an evaluation.

**Evaluation design:** The methodology selected for collecting and analyzing data in order to reach defensible conclusions about program or project efficiency and effectiveness.

**Evaluation plan:** A detailed description of how the evaluation will be implemented that includes the resources available for implementing the plan, what data will be gathered, the research methods to be used to gather the data, a description of the roles and responsibilities of sponsors and evaluators , and a timeline for accomplishing tasks.

**Evaluation planning matrix:** Matrix that organizes evaluation questions and plans for collecting information to answer them. The matrix links descriptive, normative and cause-and-effect questions to the design and methodologies.

**Evaluation Scope of Work:** A written description of the objectives, tasks, methods, deliverables and schedules for an evaluation. An evaluation Scope of Work is often used for contracting evaluations with external evaluation service providers. An evaluation SOW may cover the full evaluation or specific components within an evaluation.

**Evaluation stakeholder Analysis:** A process to identify and assess the importance of key people, groups of people, or institutions that may significantly influence the success of an evaluation.

**Evaluation standards:** A set of criteria against which the completeness and quality of evaluation work can be assessed. The standards measure the utility, feasibility, propriety and accuracy of the evaluation. Evaluation standards must be established in consultation with stakeholders prior to the evaluation.\*

**Evaluator competencies:** Core competencies for evaluation practitioners usually include level of professional practice, ability to conduct systematic inquiry, ability to carry out situational analysis, project management skills, as well as reflective practice (personal evaluation and pursuit of professional development) and interpersonal competence. These essential competencies are not endorsed across the board by associations in the evaluation field although many organizations are looking for ways to standardize these competency designations. USAID has developed its own specialized evaluator competencies for Evaluation Specialists and Program Managers.

**Experimental design:** A methodology in which research subjects are randomly assigned to either a treatment or control group, data is collected both before and after the intervention, and results for the treatment group are benchmarked against a counterfactual established by results from the control group.

**Ex-post:** (after the fact) refers to the time period after a given intervention has been administered. Usually used in outcome and impact evaluations to determine the effect of a given intervention on the outcomes measured.

**External Validity:** The degree to which findings, conclusions, and recommendations produced by an evaluation are applicable to other settings and contexts.

**Formative Evaluation:** An evaluation intended to help improve the implementation and outcomes of the ongoing project or program being evaluated. By contrast, a *summative evaluation* is intended to assess the impacts of effects of a completed program but not usually to improve its implementation.

**Gender:** a social construct that refers to relations between and among the sexes, based on their relative roles. It encompasses the economic, political, and socio-cultural attributes, constraints, and opportunities associated with being male or female. As a social construct, gender varies across cultures, is dynamic and open to change over time. Because of the variation in gender across cultures and over time, gender roles should not be assumed but investigated. Note that “gender” is not interchangeable with “women” or “sex.”

**Gender analysis:** Examines access to and control over resources by men and women. It also refers to a systematic way of determining men’s and women’s often differing development needs and preferences and the different impacts of development on women and men. Gender analysis frameworks can provide an important basis (or baseline) for future evaluations

**Generalizability:** To generalize findings to a defined target population (such as all low-income households). Usually this is done through application of quantitative methods and construction of representative sample. Also see external validity.

**History Effect:** When there is a possibility that events that occurred during the course of the intervention or between repeated measures, that are not part of the intervention, may have influenced the outcomes. For a group of individuals a historical threat to internal validity must identify an event that simultaneously affects most of the individuals enough to appreciably change the measured trait.

**Indicators:** Quantitative or qualitative variable that provides reliable means to measure a particular phenomenon or attribute.

**Indicator targets:** Specific and measurable levels of project performance ideally set during the project planning stage. For USAID, targets are set for indicators at the assistance objective (AO), Intermediate Result (IR), and output levels. Targets are generally used for performance management .

**Instrumentation Effect:** This occurs if the reliability of the instrument changes. Changes can be a result of measurement scale or standards changes or recalibration of instruments of measurement (e.g. recalibrating a health clinic’s weight scale during a nutrition intervention). This causes a lack of reliability of the instrument used to measure them.

**Internal validity:** The degree to which conclusions about causal linkages are appropriately supported by the evidence collected. The seven (7) main threats to internal validity of evaluations include the follow effects: 1) history, 2) maturation, 3) repeated testing, 4) selection, 5) mortality, 6) regression to the mean, and 7) instrumentation. Each are described elsewhere in the Glossary.

**Interrupted time series design:** An evaluation design that collects observations at multiple time points before and after an intervention (interruption). The design attempts to detect whether the intervention has had an effect significantly greater than the underlying trend. Also called a time series design.

**Impact evaluation:** measures the change in a development outcome that is attributable to a defined intervention; impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change. Impact evaluations in which comparisons are made between beneficiaries that are randomly assigned to either a —treatment or a —control group provide the strongest evidence of a relationship between the intervention under study and the outcome measured.

**Logical framework:** A management tool used to improve the design and evaluation of interventions that is widely used by development agencies. It is a type of logic model that identifies strategic project elements (inputs, outputs, outcomes, impact) and their causal relationships, indicators, and the assumptions or risks that may influence success and failure. Related term: Results Framework.

**Longitudinal design:** A type of interrupted time series design in which repeated measures of the same variable are taken from the same subjects. When used for descriptive questions, a longitudinal design may be used to find out, for example, whether children attending an enrichment program maintain learning gains over time.

**Managing for Results :** is the systematic process of monitoring the achievements of program activities; collecting and analyzing performance information to track progress toward planned results; using performance information and evaluations to influence decision-making and resource allocation; and communicating results to advance organizational learning and communicate results to stakeholders.\*

**Maturation Effect:** This occurs when results are caused by aging or development. Changes that naturally occur as a result of the passage of time include growing older, getting smarter and gaining experience. Maturation may be conceived as occurring in two forms: short or long term. **Short-term maturation** is demonstrated by fatigue and learning. **Long-term maturation** deals with psychophysical development, cultural changes, and environmental changes that can affect psychological constructs.