EVALUATION FLASH CARDS

Embedding Evaluative Thinking in Organizational Culture

Developed by Michael Quinn Patton
Utilization-Focused Evaluation

St. Paul, Minnesota
Updated March 2017

OTTO BREMER TRUST™

CONTENTS

INTRODUCING THE EVALUATION FLASH CARDS

As part of our ongoing work to strengthen our support for communities, the trustees and staff of the Otto Bremer Trust engaged in a series of learning seminars on evaluation. In order to make the core concepts easily accessible and retrievable, we asked Michael Quinn Patton, who led the seminars, to create a set of basic reference cards. These became the Evaluation Flash Cards presented here, with the idea that a core concept can be revisited "in a flash." Illustrations of the concepts are drawn from Otto Bremer Trust grants. We hope this resource is useful to other organizations committed to understanding and improving the results of the programs they support.

These cards are not intended to be definitive, universally applicable, or exhaustive of possibilities.

ABOUT THE AUTHOR

Michael Quinn Patton is an independent evaluation consultant with 40 years of experience conducting evaluations, training evaluators, and writing about ways to make evaluation useful. He is former president of the American Evaluation Association and recipient of both the Alva and Gunnar Myrdal Award for outstanding contributions to evaluation use and practice and the Paul F. Lazarsfeld Award for lifetime contributions to evaluation theory, both from the American Evaluation Association. The Society for Applied Sociology honored him with the Lester F. Ward Award for outstanding contributions to applied sociology. He is the author of six books on evaluation, including *Essentials of Utilization-Focused Evaluation* (2012).

PERMISSION AND CITATION

The Otto Bremer Trust permits use of these Evaluation Flash Cards for non-commercial purposes, subject to full attribution (see the suggested citation reference below). For permission to use this material for commercial purposes, please contact the Trust at 651-227-8036 or communications@ottobremer.org.

Citation reference: Patton, Michael Quinn (2014).
Evaluation Flash Cards: Embedding Evaluative Thinking in Organizational Culture.
St. Paul, MN: Otto Bremer Trust, ottobremer.org.

1 EVALUATIVE THINKING

DISTINGUISH EVALUATIVE THINKING FROM EVALUATION.

Evaluation is activity. Evaluative thinking is a way of doing business.

Evaluative thinking is systematic *results-oriented thinking* about:

- What results are expected,
- How results can be achieved,
- What evidence is needed to inform future actions and judgments, and
- How results can be improved in the future.

Evaluative thinking becomes most meaningful when it is embedded in an organization's culture. This means that people in the organization expect to engage with each other in clarifying key concepts, differentiating means and ends, thinking in terms of outcomes, examining the quality of evidence available about effectiveness, and supporting their opinions and judgments with evidence.

Evaluative thinking is what characterizes learning organizations. Keeping up with research and evaluation findings becomes part of everyone's job. Inquiring into the empirical basis for assertions about what works and doesn't work becomes standard operating procedure as people in the organization engage with each other and interact with partners and others outside the organization. Critical thinking and reflection are valued and reinforced.

Infusing evaluative thinking into organizational culture involves looking at how decision makers and staff incorporate evaluative inquiry into everything they do as part of ongoing attention to mission fulfillment and continuous improvement. Integrating evaluation into organizational culture means "mainstreaming evaluation"—that is, making it central to the work rather than merely an add-on, end-of-project paperwork mandate.

INDICATORS THAT EVALUATIVE THINKING IS EMBEDDED IN AN ORGANIZATION'S CULTURE

- Evaluative thinking permeates the work, with conscious and constant reflection on project, program, regional, and organizational experience and the intention to implement improvements based on what is learned.
- Evaluative thinking is demonstrated in the implementation of well-focused programs and in the use of high-quality evaluations that feed into program and organizational decision making.
- Time and resources are allocated for reflection on evaluation findings and using those findings.

The antithesis of evaluative thinking is treating evaluation as a check-it-off compliance activity.

EVALUATIVE THINKING EMBEDDED AND VALUED AS A WAY OF DOING BUSINESS	EVALUATION AS A COMPLIANCE ACTIVITY
Thinking about what kinds of information are most needed for learning and improvement.	Focusing on evaluation contract requirements and procedures.
Reflecting together on evaluation findings, learning lessons, and applying them in future decisions.	Checking off that evaluation reports have been submitted and filed.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Pay attention to how, and how much, evaluative thinking is manifest, embedded, and valued.

BOTTOM LINE
Practice evaluative thinking. Like any important skill, evaluative thinking improves with practice and reinforcement.

2 EVALUATION QUESTIONS

BEGIN WITH BASIC DESCRIPTION.

Evaluation supports reality testing—finding out what is actually going on in a program. This can then be compared to what was intended and hoped for. But the first step is basically descriptive.

I keep six honest serving-men
(They taught me all I knew);
Their names are What and Why and When
And How and Where and Who.

— Rudyard Kipling (1865–1936), *The Elephant's Child*

For professionals as diverse as journalists, police detectives, lawyers, and evaluators, *Kipling's five Ws and one H* is the formula for full understanding and a complete report. These are descriptive, factual, and open-ended questions. None can be answered "yes" or "no." You have to find out what happened. When first entering a program situation (for example, on a site visit), it can be helpful to begin with some basic facts to get the lay of the land. Keep it simple: Who's proposing to do what? Where? When? How? Why?

EXAMPLE: A JOB-TRAINING PROGRAM

PROGRAM DESCRIPTION	PARALLEL EVALUATION QUESTIONS
Who: The target population is chronically unemployed people of color. The staff consists of "coaches" and trainers selected for their capacity to work with this population.	Who does the program actually serve? How does the actual population served compare to the targeted population?
What: Train participants in both "soft skills" and "hard skills" to get living-wage jobs with benefits in companies the program has cultivated.	What training do participants actually receive? How does the training received compare to the proposed training? What do companies report about the skills of participants hired?
Where: The main program operates in two local offices.	How does the location of the program affect its operation? Strengths and weaknesses of location?
How: The program uses an "empowerment curriculum" that engages participants in being accountable, responsible, and successful. Building on empowerment, the program offers skill training matched to the needs and interests of participants and job needs of companies.	How does the curriculum work in practice? What are participants' reactions? What is evidence of "empowerment," of acquisition of "soft" and "hard" skills, and of alignment between companies' needs and program participants' skills?
Why: Evaluation of successful employment programs shows that the combination of positive attitudes, appropriate behaviors for the workplace, and training in skills needed by companies leads to successful outcomes.	To what extent does the program reproduce the results documented in previous evaluations? How do the results of this program compare to other models?
When: Participants are generally in the program for 18 months to 2 years. The intended outcome is retention of a living-wage job with benefits for at least one year.	To what extent is the intended outcome actually attained?

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Use the full set of descriptive questions to get a comprehensive picture of what's being proposed.

BOTTOM LINE
Ground evaluation in basic descriptive questions.

3 LOGIC MODELS

MODELS CAN BE DISPLAYED AS A SERIES OF LOGICAL AND SEQUENTIAL CONNECTIONS. EACH STEP IN A LOGIC MODEL CAN BE EVALUATED.

A logic model is a way of depicting the program intervention by specifying inputs, activities, outputs, outcomes, and impacts in a sequential series.

EXPLANATIONS OF SOME OF THE TERMS USED IN LOGIC MODELS

- *Inputs* are resources like funding, qualified staff, participants ready to engage in the program, a place to hold the program, and basic materials to conduct the program. These inputs, at an adequate level, are necessary precursors to the program's activities.
- Participating in program *activities* and processes logically precedes outputs, like completing the program or getting a certificate of achievement.
- *Outputs* lead to short-term participant outcomes, like a better job or improved health.
- Short-term *outcomes* lead to longer-term *impacts*, like a more prosperous or healthy community.

Logic models are one way of answering the *It* question in evaluation. The logic model depicts what is being evaluated.

The primary criteria for judging a logic model are whether the linkages are logical and reasonable.

1. Are the inputs (resources) sufficient to deliver the proposed activities?
2. Will the proposed activities lead to the expected outputs?
3. Do the outputs lead logically and reasonably to the outcomes?
4. Will successful outcomes lead to desired impacts?

NOT LOGICAL AND REASONABLE	LOGICAL AND REASONABLE
Attending an after-school drop-in center will increase school achievement.	Participating in an after-school drop-in center will help keep kids out of trouble after school.
A safe house for victims of domestic abuse will lead to jobs.	A safe house for domestic abuse victims will provide support and stability to enable participants to figure out next steps and get referrals for longer-term help.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Does the proposal include a logic model? If so, is it reasonable and logical? Do the steps make sense?

BOTTOM LINE
Is the proposed logic model sequence from inputs to impacts logical and reasonable?

4 THEORY OF CHANGE

TESTING A THEORY OF CHANGE CAN BE AN IMPORTANT CONTRIBUTION OF EVALUATION.

A theory of change explains how to produce desired outcomes. It is explanatory. A logic model just has to be sequential (inputs before activities, activities before outcomes), logical, and reasonable. In contrast, a theory of change must explain why the activities produce the outcomes.

EXAMPLE

A program to help homeless youth move from the streets to permanent housing proposes to:

1. Build trusting relationships with the homeless youth;
2. Work to help them feel that they can take control of their lives, instill hope, and help them plan their own futures; and
3. Help them complete school, both for their economic well-being and to help them achieve a sense of accomplishment.

This approach is based on *resilience* research and theory. Resilience research and theory posits that successful youth: (1) have at least one adult they trust and can interact with, (2) have a sense of hope for the future, (3) have something they feel good about that they have accomplished, and (4) have at least some sense of control over their lives.

The issue that arises in examining a proposal based on a theory of change is **whether the proposed program activities constitute a practical and reasonable implementation of the theory**. Does the program provide specific and concrete experiences that reflect the theory of change? The key conceptual and real-world challenge is translating a theory of change into an actual implemented program with real outcomes.

Evaluation of a program with an explicit theory of change is sometimes called "theory-driven evaluation" because the evaluation can be a test of the theory. If the program fails to produce the predicted outcomes, the critical interpretative and explanatory issue becomes: Did the program fail because the theory was inadequately implemented, or because the theory itself was inadequate? This is the difference between *implementation failure* versus *theory failure*, a longstanding and important distinction in evaluation.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

How explicit and articulate is the program's theory of change?

BOTTOM LINE

Can a program identify a theory of change based on research and, if so, can it demonstrate how it will translate the theory into an actual program?

5 EVALUATION VS. RESEARCH

EVALUATION AND RESEARCH HAVE DIFFERENT PRIMARY PURPOSES, BUT THE STATE OF RESEARCH KNOWLEDGE AFFECTS WHAT EVALUATION CAN CONTRIBUTE.

Evaluation generates improvements, judgments, and actionable learning about programs. Research generates knowledge about how the world works and why it works as it does.

Scientific research is undertaken to discover knowledge, test theories, and generalize across time and space. Program evaluation is undertaken to inform decisions, clarify options, identify improvements, and provide information about programs and policies within contextual boundaries of time, place, values, and politics. Research informs science. Useful evaluation supports action.

Research informs evaluation in that the more knowledge that exists about a problem, the more an evaluation can draw on that knowledge. For example, research shows that children immunized against polio do not get polio. Therefore, evaluation of an immunization program can stop at determining that children have been immunized and confidently calculate how many cases of polio have been prevented based on epidemiological research. The evaluation design does not have to include follow-up to determine whether immunized children get polio. That question has been settled by research.

A program aimed at getting senior citizens to exercise to improve their health does not have to prove that exercise improves health and contributes to a longer, higher quality life. Health research has demonstrated that. Evaluation of the exercise program, then, only has to demonstrate that it is effective in getting seniors to exercise at the levels shown by research to be effective.

In contrast, there is little research on homeless youth. The knowledge gap is huge. So evaluation has to be more developmental and exploratory because the research foundation is weak.

RESEARCH	EVALUATION
Purpose is testing theory and producing generalizable findings.	Purpose is to determine the effectiveness of a specific program or model.
Questions originate with scholars in a discipline.	Questions originate with key stakeholders and primary intended users of evaluation findings.
Quality and importance judged by peer review in a discipline.	Quality and importance judged by those who will use the findings to take action and make decisions.
Ultimate test of value is contribution to knowledge.	Ultimate test of value is usefulness to improve effectiveness.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Find out if research supports a program proposal. Have those submitting the proposal done their homework in finding out and taking into account what research shows?

BOTTOM LINE
Distinguish research from evaluation. Use research to inform both program and evaluation designs.

6 DOSAGE

DIFFERENT DEGREES OF INTERVENTION AND ENGAGEMENT PRODUCE DIFFERENT LEVELS OF OUTCOMES.

Dosage effects refer to the fact that different people engage in and experience a program with different degrees of intensity. A higher dose of engagement should be related to higher-level outcomes.

EXAMPLE

A youth community center reports serving 300 kids each quarter.

QUESTION

What are different degrees of dosage for those 300 kids?

DATA

High dosage/high outcomes:

Thirty kids come to the center after school every day. They have important, ongoing relationships with staff. They benefit greatly from the staff's mentoring, homework help, personal support, and individualized problem solving.

Medium dosage/medium outcomes:

Fifty kids come to the center about once a week for a specific program, like a volunteer program that helps them improve reading; they get some modest help on a specific outcome (reading).

Low dosage/minimal outcomes:

Another 220 kids come once a quarter for pizza night, or a Friday night dance. This is a source of recruiting and connection to the community, but it is really outreach rather than "serving" those kids.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Explore how aware the program is of variations in dosage and the implications of those variations.

BOTTOM LINE
Watch for and understand dosage effects. All programs have them.

7 DISAGGREGATION

WHAT WORKS FOR WHOM IN WHAT WAYS WITH WHAT RESULTS?

Subgroups in programs have different experiences and different outcomes. Disaggregation refers to distinguishing the experiences and outcomes of different subgroups.

EXAMPLE

A program aims to prevent teenage pregnancies. The program typically reports aggregate results for all teens served (ages 13–19). The reported success rate is 60 percent, which means that 60 percent of the teens do not get pregnant during the year they are engaged in the program.

DISAGGREGATED DATA

- Success rate for teens aged 16–19: 80 percent
- Success rate for teens aged 13–15: 40 percent

LESSON

The overall 60 percent success rate for all teens disguises the fact that the program is highly effective with older teens and relatively ineffective with younger teens. Indeed, some outcomes are different. The program works to help older teens maintain safe and supported independence but attempts to get younger teens integrated into a family, either their own or a foster family. In reality, the two subgroups constitute different approaches with different results. The disaggregated data can help decision makers target improvements to the subgroups for whom the program is less effective—and learn from those that show higher levels of impact.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Explore the capacity of the program to disaggregate data for learning, management, and reporting.

BOTTOM LINE
When looking at overall results for a program, ask about the disaggregated results for important subgroups.

8 CHANGING DENOMINATORS, CHANGING RATES

DIFFERENT DENOMINATORS PRODUCE DIFFERENT RESULTS.

To understand and interpret data on rates and performance indicators, like the participation rate in a program, the drop-out rate, or the completion rate, pay special attention to the denominator.

EXAMPLE

A local job-training program reports a 40 percent drop-out rate. The denominator for this program's rate is based on the number who have completed the initial training and signed the program contract. Thus, the drop-out rate is NOT based on the number who initially enroll in the program but rather the number who enroll *and* complete the course *and* sign the contract. Half of the initial enrollees do not reach that stage.

ILLUSTRATIVE DATA

1. Number who enter the program from January to June: 200
2. Number who complete course and sign contract: 100
3. Contract signing rate: 50 percent (100/200 = 50 percent)
4. Number who drop out before job placement: 40
5. Drop-out rate for contract signers is 40 percent (40/100 = 40 percent)
6. Drop-out rate for ALL enrollees is 70 percent (140/200 = 70 percent)
7. Program completion (placed in a job): 60
8. Completion rate of contract signers: 60 percent (60/100 = 60 percent)
9. Job retention one year after placement: 30 participants
10. Job retention rate: 50 percent (30/60 = 50 percent)
11. Job retention percentage of all participants who enroll: 15 percent (30/200 = 15 percent)

LESSON

Different rates have different denominators. Different denominators yield different rates. Programs define and calculate drop-out and completion rates differently, which makes comparisons difficult.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Explore how the program computes key indicators like participation, completion, and drop-out rates.

BOTTOM LINE
Be clear about the denominator being used when rates are reported.

9 SMART GOALS

NOT ALL GOALS ARE CREATED EQUAL.

Traditionally, evaluation has been synonymous with measuring goal attainment. The most basic evaluation question is: To what extent is the program attaining its goals? To evaluate goal attainment, goals have to be clear enough to permit evaluation.

A CLEAR GOAL HAS FIVE DIMENSIONS, WHICH FORM THE ACRONYM **SMART**:

Specific

Measurable

Achievable

Relevant

Time bound

EXAMPLES

Weak goal:

Improve quality of life. This goal is vague and general (not specific). What is meant by quality of life? How would it be measured? What's the timeframe?

SMART goal:

Graduates will get a job paying a living wage with benefits and keep the job for at least a year.

- The outcome is **specific** *(get and keep a certain kind of job)*
- The goal is **measurable** *(living-wage job with benefits)*
- The goal is **achievable** *(the level of aspiration is reasonable)*
- The outcome is **relevant** *(the goal is aimed at the chronically unemployed; getting and keeping a living-wage job is relevant to both participants and society)*
- The goal is **time bound** *(keep the job at least one year)*

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

When reviewing goals, examine if they are SMART.

BOTTOM LINE
Goal statements vary tremendously. Not all are SMART.

10 DISTINGUISHING OUTCOMES FROM INDICATORS

EVALUATION DEPENDS ON IMPORTANT DISTINCTIONS. ONE SUCH DISTINCTION IS OUTCOMES VS. INDICATORS.

An outcome is a clear statement of the targeted change.
An indicator is a measurement of the outcome.

EXAMPLES OF TYPES OF OUTCOMES	ILLUSTRATIVE INDICATORS
Change in circumstances	Number of children in foster care who are safely reunited with their families of origin
Change in status	Number of unemployed who become employed
Change in behavior	Number of former truants who regularly attend school
Change in functioning	Measures of increased self-care among nursing home residents
Change in attitude	Score on an instrument that measures self-esteem
Change in knowledge	Score on an instrument that measures understanding of the needs and capabilities of children at different ages

An indicator is just that, an indicator. It's not the same as the desired outcome, but only an indicator of that outcome. A score on a reading test is an indicator of reading capability but should not be confused with a particular person's true capacity to read. Many kinds of things affect a test score on a given day. Thus, *indicators are inevitably approximations*. They are imperfect and vary in validity and reliability.

Figuring out how to measure a desired outcome is called *operationalizing* the outcome. The resources available for measurement will greatly affect the kinds of data that can be collected for indicators. For example, if the desired outcome for abused children is no subsequent abuse or neglect, regular in-home visits and observations, including interviews with the child, parent(s), and knowledgeable others, would be desirable, but such data collection is expensive. With constrained resources, one may have to rely on data collected routinely by government through mandated reporting—that is, official, substantiated reports of abuse and neglect over time. Moreover, when using such routine data, privacy and confidentiality restrictions may limit the indicator to aggregate results quarter by quarter rather than one that tracks specific families over time.

Another factor affecting indicator selection is the demands data collection will put on program staff and participants. Short-term interventions such as food shelves, recreational activities for people with developmental disabilities, drop-in centers, and one-time community events do not typically engage participants with a high enough dosage level to justify collection of sophisticated data. Many programs can barely collect data on end-of-program status, much less follow-up data six months after program participation. Programs may need to develop the capacity to measure outcomes.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Examine the clarity of proposed outcomes and the meaningfulness of indicators.

BOTTOM LINE
Outcomes are the desired results; indicators are how you know about outcomes. The key is to make sure that the indicator is a reasonable, useful, and meaningful measure of the intended participant outcome.

11 PERFORMANCE TARGETS

WHAT'S THE BULL'S-EYE?

A performance target specifies the level of outcome that is hoped for, expected, or intended.

What percentage of participants in employment training will have full-time jobs six months after graduation? 40 percent? 65 percent? 80 percent? What percentage of fathers failing to make child support payments will be meeting their full child support obligations within six months of intervention? 15 percent? 35 percent? 60 percent?

Setting performance targets should be based on data about what is possible. The best basis for establishing future performance targets is past performance. "Last year we had 65 percent success. Next year we aim for 70 percent." Lacking data on past performance, it may be advisable to wait until baseline data has been gathered before specifying a performance target. Arbitrarily setting performance targets without some empirical baseline may create artificial expectations that turn out unrealistically high or embarrassingly low. One way to avoid arbitrariness is to seek norms for reasonable levels of attainment from other, comparable programs, or review the evaluation literature for parallels. Just making up arbitrary or ambitious performance targets is not very useful.

SEPARATE GOALS FROM INDICATORS AND PERFORMANCE TARGETS.

- *Desired outcome:* All children will be immunized against polio.
- *Indicator:* Health records when children enter school show whether they have been vaccinated.
- *Performance target:* Children receive four doses of IPV: a dose at 2 months, at 4 months, and at 6–18 months; and a booster dose at 4–6 years.

As indicators are collected and examined over time, it becomes more meaningful and useful to set performance targets.

EXAMPLE

Consider this outcome statement: Student achievement test scores in reading will increase one grade level from the beginning of first grade to the beginning of second grade.

Such a statement mixes together and potentially confuses the (1) specification of a desired outcome (better reading) with (2) its measurement (achievement test scores) and (3) the desired performance target (one grade level improvement).

Specifying the desired outcome, selecting indicators, and setting targets are separate decisions. They are related, of course, but each should be examined on its own merits. For example, there are ways other than standardized tests for measuring achievement, like student portfolios or competency-based tests. The desired outcome should not be confused with its indicator.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Examine the appropriateness and basis of performance indicators.

BOTTOM LINE
The challenge is to make performance targets realistic, meaningful, and useful.

12 QUALITATIVE EVALUATION

QUALITATIVE DATA COMES FROM OPEN-ENDED INTERVIEWS, ON-SITE OBSERVATIONS, FIELDWORK, SITE VISITS, AND DOCUMENT ANALYSIS.

Qualitative evaluation uses case studies, systematically collected stories, and in-depth descriptions of processes and outcomes to generate insights into what program participants experience and what difference those experiences make.

Suppose you want to evaluate learning to read. If you want to know how well children can read, give them a reading test (quantitative data). If you want to know what reading means to them, you have to talk with them (qualitative data). Qualitative questions aim at getting an in-depth, individualized, and contextually sensitive understanding of reading for each child interviewed. Of course, the actual questions asked are adapted for the child's age, language skills, school and family situation, and purpose of the evaluation. But regardless of the precise wording and sequence of questions, the purpose is to hear children talk about reading in their own words; find out about their reading behaviors, attitudes, and experiences; and get them to tell stories that illuminate what reading means to them. You might talk to groups of kids about reading as a basis for developing more in-depth, personalized questions for individual interviews. While doing field work (actually visiting schools and classrooms), you would observe children reading and the interactions between teachers and children around reading. You would also observe what books and reading materials are in a classroom and observe how they are arranged, handled, and used. In a comprehensive inquiry, you would also interview teachers and parents to get their perspective on the meaning and practice of reading, both for children and for themselves, as models children are likely to emulate.

EXAMPLES OF QUALITATIVE EVALUATION	QUALITATIVE DATA COLLECTED, SYNTHESIZED, AND REPORTED
Evaluate the principles that guide work with homeless youth, both to improve effective use of principles and find out the impacts on youth.	Case studies of diverse homeless youth using shelters and youth programs; in-depth interviews with youth, street workers, and shelter or program staff; review of files; focus groups with youth to understand their perspectives and experiences.
Evaluate the role of community colleges in rural communities.	Interview community college teachers, students, and administrators about their experiences and perspectives; interview key community people and leaders; do case studies of successful students compared to drop-outs.
Evaluate a community leadership program.	Interviews with program participants about the leadership training, then follow-up community case studies to find out what they do with the training.
Evaluate a drop-in center for inner-city Native American youth.	Work with Native American leaders to develop culturally appropriate questions. Observe. Interview. Report patterns.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Develop skills in open-ended interviewing and systematic site visit observations—emphasis on being skilled and systematic. Document what you see and hear. Analyze and synthesize *qualitatively*.

BOTTOM LINE

Qualitative evaluation captures and communicates the perspectives, experiences, and stories of people in programs to understand program processes and outcomes from their viewpoint.

13 TRIANGULATION THROUGH MIXED METHODS

ANY SINGLE SOURCE OF DATA, LIKE INTERVIEWS, FOCUS GROUPS, OR SURVEYS, HAS BOTH STRENGTHS AND WEAKNESSES.

Using multiple methods increases confidence in overlapping patterns and findings. Checking for consistency across different data sources is called triangulation.

The term triangulation is taken from land surveying. Knowing a single landmark only locates you somewhere along a line in a direction from the landmark, whereas with two landmarks you can take bearings in two directions and locate yourself at their intersection. The notion of triangulating also works metaphorically to call to mind the world's strongest geometric shape—the triangle.

The logic of triangulation is based on the premise that no single method ever adequately solves the problem of interpreting how much the weakness of any particular method may give a false or inadequate result. Because different kinds of data reveal different aspects of a program, multiple methods of data collection and analysis provide more grist for the interpretation mill. Combinations of interviewing, observation, surveys, performance indicators, program records, and document analysis can strengthen evaluation. Studies that use only one method are more vulnerable to errors.

COMBINING QUANTITATIVE AND QUALITATIVE DATA

Statistics tell us about the size or scope of an issue, like the number of homeless youth, how many rural people lack access to quality dental care, or whether the number of children in poverty is increasing or decreasing.

Qualitative data tells us what the numbers mean through the perceptions of program participants and staff. Open-ended interviews with program participants, case studies, and site visits provide insights into how to interpret and make sense of the numbers. Stories also put faces on the numbers and humanize statistics so that we never forget that behind the numbers are real people living their lives.

Strong evaluations include both quantitative and qualitative data. Triangulating across statistics and stories make each data source more valuable, meaningful, and credible.

EXAMPLE

A site visit to a housing development turned up statistics on residents' characteristics, diversity, and income levels as well as the needs people expressed and stories about living in the housing development. Staff learned that to live in this development "you need to work, be in school, or have formal volunteering occurring." An evaluation going forward might inquire how this policy works in practice. Statistics would reveal patterns of work, school attendance, volunteering, and resident turnover. Open-ended interviews would find out how residents and staff experience these policies—the attitudes, knowledge, behaviors, and feelings that affect the desired outcome of building a vibrant residential community.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

When reviewing a proposal or conducting a site visit, look for both numbers and stories, and examine the consistency or conflicts between these different data sources.

BOTTOM LINE

***The evaluation ideal is:** No numbers without stories; no stories without numbers. Learn what each kind of data reveals and teaches, and how to use them together: triangulating.*

14 IMPORTANT AND RIGOROUS CLAIMS OF EFFECTIVENESS

NOT ALL FINDINGS ARE THE SAME. WHAT'S WORTH PAYING ATTENTION TO? WHAT MATTERS MOST?

The most powerful, useful, and credible claims are those that are of major importance and have strong empirical support. Claims can be important or unimportant, and the evidence for the claims can be strong or weak. The ideal is strong evidence supporting claims of major importance.

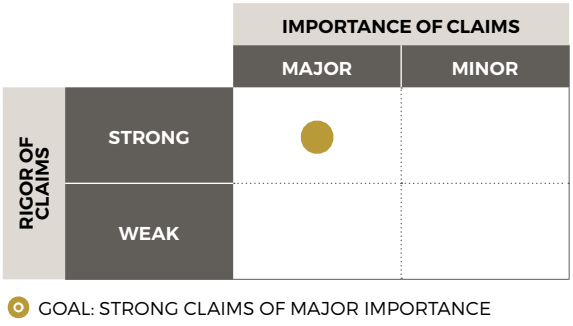
Example of an effectiveness claim: Programs serving homeless youth are contributing significantly to reducing youth homelessness in the Twin Cities.

CHARACTERISTICS OF CLAIMS OF MAJOR IMPORTANCE

- Involve making a difference, having an impact, or achieving desirable outcomes
- Deal with a problem of great societal concern
- Affect a relatively large number of people
- Provide a sustainable solution (something that lasts over time)
- Save money and/or time—that is, accomplish something with less money and in less time than is usually the case (an efficiency claim)
- Enhance quality
- Claim to be new or innovative
- Show that something can actually be done about a problem—that is, claim the problem is malleable
- Involve a model or approach that could be used by others (meaning the model or approach is clearly specified and adaptable to other situations)

CHARACTERISTICS OF STRONG CLAIMS

- Provide valid, believable evidence in support
- Include data from multiple points in time (longer periods of follow-up provide stronger evidence of sustained change)
- Are about a clear intervention (model or approach) with solid implementation data
- Clearly specify outcomes and impacts (behavior outcomes are stronger than opinions, feelings, and knowledge)
- Include comparisons and/or replications in the evidence
- Are based on more than one kind of evidence or data (i.e., triangulation of data):
  - *Quantitative and qualitative data*
  - *Multiple sources (e.g., youth, parents, teachers, and staff corroborate results).*



PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

If the project is successful, what kinds of claims will be made?

BOTTOM LINE
Review claims, carefully examining the importance of the claim and the strength of the evidence.

15 ACCOUNTABILITY EVALUATION

DIFFERENT TYPES OF EVALUATION SERVE DIFFERENT PURPOSES.

Accountability-focused evaluation determines whether funds were expended appropriately to accomplish intended results.

ACCOUNTABILITY QUESTIONS

- Are funds being used for intended purposes?
- Are goals and targets being met?
- Are resources being efficiently allocated?
- Are problems being handled?
- Are staff qualified?
- Are only eligible participants being accepted into the program?
- Is implementation following the approved plan?
- Are quality control mechanisms in place and being used?

Answers to these questions determine whether resources are well managed, being used for approved purposes, and efficiently attaining desired results.

EXAMPLES	ACCOUNTABILITY EVALUATION QUESTIONS
A job-training program provides comprehensive job training to low-income adults.	— How many low-income adults received training? — What training was provided?
A dental services provider delivers affordable dental services to low-income children and expectant mothers in North Dakota.	— How many low-income children and expectant mothers are served? — What services are provided with what results? — Were funds used as approved?
A local community college seeks to build capacity and increase organizational effectiveness by purchasing a database to be shared by area community colleges.	— Was the database purchased? — Is it being used by area community colleges? — Has organizational effectiveness increased?

The utility of an accountability system depends on who is held accountable, by whom, for what. Accountability is most meaningful when those held accountable actually have the capacity to achieve the things for which they are held accountable, within the timeframes expected.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Find out how the program is currently thinking about and dealing with accountability issues.

BOTTOM LINE
Accountability requires clarity about who is accountable to whom for what.

16 FORMATIVE EVALUATION

DIFFERENT TYPES OF EVALUATION SERVE DIFFERENT PURPOSES.

Formative evaluation supports program improvement. The emphasis is on forming, shaping, and improving, thus the term formative.

FORMATIVE EVALUATION QUESTIONS

- What works and what doesn't?
- What are the program's strengths and weaknesses?
- What's the feedback from participants in the program about what should be improved?
- How do different subgroups respond—that is, what works for whom in what ways and under what conditions? (If one size doesn't fit all, how can the needs of different people be met?)
- How can outcomes and impacts be increased?
- How can costs be reduced?
- How can quality be enhanced?

The emphasis in these formative questions is on improvement.

EXAMPLES	FORMATIVE EVALUATION QUESTIONS
A local program provides job training for low-income adults.	— What are the program's strengths and weaknesses from the perspective of participants? — What can be improved?
A local program aims to help victims of domestic violence get jobs and improve their lives.	— How can links to partners for referral services be strengthened?
A local organization offers a program to empower those in poverty and limited by poor education.	— Across the variety of services offered, which ones are working well and which need improvement? — In the empowerment gatherings, what works for whom in what ways, with what outcomes? What can be learned from feedback to improve the empowerment gatherings?

The utility of formative evaluation depends on a willingness to distinguish strengths from weaknesses and acknowledge what needs improvement. Grantees often fear reporting weaknesses or problems to funders. Formative evaluation requires mutual trust and a shared commitment to learning, improving, and getting better.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Find out how the project or program will attend to learning and improvement.

BOTTOM LINE

Formative evaluation requires openness to learning and a commitment to ongoing improvement. Funders contribute to mutual trust by valuing learning and supporting honest communication about what's working and what needs improvement.

17 SUMMATIVE EVALUATION

DIFFERENT TYPES OF EVALUATION SERVE DIFFERENT PURPOSES.

Summative evaluation judges the overall merit, worth, and significance of a project. The term summative connotes a summit (important) or summing-up judgment.

The focus is on judging whether a model is effective. Summative evaluation is used to inform decisions about whether to expand a model, replicate it elsewhere, and/or "take it to scale" (make it a statewide, region-wide, or national model).

SUMMATIVE EVALUATION QUESTIONS

- Does the program meet participants' needs effectively and efficiently?
- Is the model well specified and standardized so that the resources needed, services delivered, and outcomes attained are clear?
- What are the key factors that support success? What key contextual factors affect outcomes?
- To what extent can outcomes be attributed to the intervention? Is the program theory clear?
- Does the evidence support the theory?
- How do outcomes and costs compare to other options? Is the model cost-effective?
- What unanticipated outcomes have been found? With what implications?
- Is the model *ready for prime time?* Is it ready for replication in other places?

The emphasis in these summative questions is on testing and validating a model. Summative evaluation serves to inform major decisions about the value of a model for future dissemination.

EXAMPLES	SUMMATIVE EVALUATION QUESTIONS
A local job-training program aspires to be a model program for moving chronically unemployed people of color into living-wage jobs with benefits in high-quality companies that offer career opportunities.	— What are the critical elements of the model? — What are the program's employment outcomes? Can these be attributed to the program? — What are the costs per participant? — Has the model been sufficiently implemented and evaluated to recommend it to others for replication?
A local micro-lending program provides $1,000 grants to meet emergency or critical needs of low-income participants in a variety of partner agencies. It wants to expand throughout the region.	— What exactly is this model? How does it work? With what outcomes? — What niche does the model fill? Does it meet an important enough need to merit expansion to other communities?

The utility of summative evaluation is the focus on informing major decisions about a model's effectiveness and, therefore, its relevance and dissemination to other communities.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

Find out if the project or program aspires to be a replicable model.

BOTTOM LINE

Summative evaluation requires rigorous evidence because the stakes are high. The evaluation data must be high quality and credible to external stakeholders interested in the model.

18 DEVELOPMENTAL EVALUATION

DIFFERENT TYPES OF EVALUATION SERVE DIFFERENT PURPOSES.

Developmental evaluation supports innovation and development, especially in complex and dynamic situations.

Accountability evaluation, formative evaluation, and summative evaluation all depend on there being a program or model to evaluate. Developmental evaluation is relevant when the program or model is still being created and those involved are figuring out what they want to do and how to do it. Thus, developmental evaluation helps people developing new initiatives to get ongoing, real-time feedback about what is emerging and its implications for making a difference.

EXAMPLE

A collaboration to support homeless youth involves several organizations, each with its own projects and evaluations. As individual agencies, they are engaged in accountability reporting and formative evaluation to increase effectiveness. But the overall collaborative initiative is just beginning to be created as the organizations work together. This is a new development. As they collaborate on both programming for homeless youth and overall initiative-level evaluation, they are creating new ways of working together and developing new possibilities for greater impact. Developmental evaluation supports that generative and collaborative process as it unfolds.

KEY DEVELOPMENTAL EVALUATION QUESTIONS

- What is being developed? With what implications?
- Given where we (the collaborative group members) are now, what are the next steps?
- What's happening in the larger community context that affects how we work together and what we do together? (This involves attention to the political, economic, social, and cultural context and implications of what is being developed.)

The emphasis in these evaluation questions is on what is being developed and understanding the context within which adaptation and development are occurring.

EXAMPLE	DEVELOPMENTAL EVALUATION QUESTIONS
A local program is developing a new empowerment and job-training program to serve the transition needs of men of color coming out of prison.	— What issues are emerging as the program is being developed? — How is the program adapting to the challenges of this population?

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

To what extent is the proposed project innovative and/or operating in a turbulent environment? How does the project prepare to use evaluation to support adaptation and development?

BOTTOM LINE
Developmental evaluation supports innovative development. It fills a specific evaluation niche.

BOTTOM LINE ON DIFFERENT TYPES OF EVALUATION
Evaluations can serve diverse purposes, including accountability, program improvement, making overall summative judgments and decisions, and supporting innovative development. Make sure the type of evaluation matches the situation and is appropriate to the nature of the program.

19 THE IT QUESTION

WHEN WE SAY "IT WORKS" OR "IT DOESN'T WORK," WHAT'S THE IT?

The It is the program model being implemented—and evaluated.

EXAMPLES

- A local job-training program has a structured curriculum that aims to create a positive attitude about undertaking employment training and taking personal responsibility for success (not being a victim).
- Habitat for Humanity has developed a model for how to engage volunteers and low-income people together in building a home affordable to and owned by a low-income family.
- ReStore is a model nonprofit retail outlet specializing in the resale of quality new and previously owned building materials, home furnishings, and appliances at 50 to 75 percent below retail prices. The funds from ReStore sales are used to support Habitat for Humanity.

Answering the It question has two parts:

1. What intervention is being undertaken to effect change?
2. What outcomes are expected from the intervention?

Examples of It hypothesis statements:

- The job-training program expects the empowerment curriculum to change attitudes and behaviors so that participants complete employment training and obtain and retain a living-wage job.
- Habitat for Humanity expects its model to produce houses that low-income families own and maintain.
- ReStore is a model of social entrepreneurship aimed at generating funds (operating at a "profit") to help support Habitat for Humanity programs.

Asking about the *It* means going beyond a program's name or label to find out what it means and actually does. Additional examples of models:

- A nonprofit supports a rating and improvement system for early learning and child care providers. The system is the *It*.
- Boy Scouts of America runs a program that mentors, works to strengthen, and develops relationships for at-risk youth. The program is the *It*.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

- What exactly is the model being proposed?
- What outcomes is the model expected to produce?
- What evidence will be generated about how the model works?

BOTTOM LINE
Be clear about the It (model) in a proposal. That's what will be evaluated.

20 FIDELITY OR ADAPTATION

DIFFERENT APPROACHES TO DISSEMINATING MODELS REQUIRE DIFFERENT EVALUATION APPROACHES.

Two opposing approaches to implementing a model have very different evaluation implications.

The two approaches follow.

1. ***Fidelity-focused programming and evaluation*** means a national model is being implemented in a local community and is supposed to be implemented exactly as prescribed in the national model. Fidelity-focused program models provide best practices and standard operating procedures that amount to a recipe for success. A McDonald's Big Mac is supposed to be the same anywhere in the world.

   Core evaluation questions:

   — Is the local model faithfully and rigorously implementing the standard model as specified?

   — Is the local model getting the results promised by the national model?

2. ***Adaptation-focused programming and evaluation*** means a national model offers principles and guidance, but local implementation will be adapted to fit the local context. The Pew Children's Dental Campaign is an example of a national approach to bridging the gap between coverage and care that provides an overarching framework for research and policy engagement, but it has to be adapted to a statewide context.

   Core evaluation questions:

   — How is the national framework being adapted locally?

   — What are the implications of these adaptations for outcomes?

   — Is the local adaptation getting the results promised by the national model?

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

When funds are requested to implement a model being done elsewhere, find out whether implementation is expected to be fidelity-focused or adaptive in nature.

High fidelity will typically require capacity development and technical support from those who have developed and implemented the model elsewhere. This usually includes already-developed evaluation instruments and tools.

Adaptation will typically require astute understanding of local context and capacity to articulate how the local situation will influence the adaptive process and local outcomes.

BOTTOM LINE

Distinguish fidelity-focused proposals from adaptive-focused proposals. The implications for programming and evaluation are substantial.

21 HIGH-QUALITY LESSONS LEARNED

LESSONS CAN BE GENERATED FROM ALL KINDS OF EXPERIENCES AND DATA. HIGH-QUALITY LESSONS ARE THOSE THAT ARE SUPPORTED BY DIVERSE TYPES OF EVIDENCE.

High-quality lessons are supported by multiple sources of information. Knowledge confirmed from multiple sources increases confidence that a lesson is valid and can be used to inform decisions and future actions.

A common problem when an idea becomes highly popular—in this case the search for lessons learned—is that the idea loses its substance and meaning. Anybody who wants to glorify his or her opinion can proclaim it a "lesson learned." High-quality lessons, in contrast, represent principles extrapolated from multiple sources and cross-validated that inform future action. In essence, high-quality lessons constitute validated, credible, trustworthy, and actionable knowledge.

Places to look for potential lessons

1. Evaluation findings—patterns across programs
2. Basic and applied research findings
3. Cross-validation from multiple and mixed methods, both quantitative and qualitative
4. Reflective practice wisdom based on the experiences and insights of practitioners
5. Insights reported by program participants
6. Expert opinion
7. Cross-disciplinary findings and patterns
8. Theory as an explanation for the lesson and its mechanism of impact

Assessment criteria for judging the quality of lessons

- Importance of the lesson learned
- Strength of the evidence connecting intervention lessons to outcomes attainment
- Consistency of findings across sources, methods, and types of evidence

The idea is that the greater the number of supporting sources for a "lesson learned," the more rigorous the supporting evidence; and the greater the cross-validation from supporting sources, the more confidence one has in the significance and meaningfulness of a lesson. Lessons learned with only one type of supporting evidence would be considered a "lessons learned hypothesis." Nested within and cross-referenced to lessons learned should be the actual cases from which practice wisdom and evaluation findings have been drawn. A critical principle here is to maintain the contextual frame for lessons learned—that is, to keep lessons grounded in their context. For ongoing learning, the trick is to follow future applications of lessons learned in new settings to test their wisdom and relevance over time—and adapt accordingly.

EXAMPLE

The importance of intervening in preschool years for healthy child development and later school success is supported by numerous evaluations, basic research on child development, expert knowledge, practitioner wisdom, and child development theory. In contrast, lessons about how to work effectively with troubled teenagers are weak in evidence, theory, research, and number of evaluations.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

What lessons are program proposals based on? What are the sources of evidence supporting supposed lessons? To what extent do similar lessons show up in different sites, proposals, and reports?

BOTTOM LINE

Distinguish opinions and single-source lessons from high-quality, cross-validated lessons. The former are hypotheses. The latter constitute actionable knowledge.

22 EVALUATION QUALITY STANDARDS

EVALUATION CAN AND SHOULD BE EVALUATED. SO WHAT'S A GOOD EVALUATION?

The evaluation profession has adopted standards that are criteria for what constitutes a good evaluation.

A high-quality evaluation is:

- Useful
- Practical
- Ethical
- Accurate
- Accountable

EXAMPLE

A foundation commissions an evaluation of focus work on youth homelessness. The first phase of the evaluation documents that:

- the targeted number of new beds and services were added to shelters; and
- the grantees collaborated to design an evaluation of the critical factors that lead to permanent housing and stability for homeless youth.

The grantees and foundation staff use the Phase 1 evaluation findings to develop a proposal for Phase 2.

The foundation's trustees use the evaluation findings and proposal based on the findings to inform (1) their decision about whether to fund the next stage of the youth homelessness work and (2) how to shape future work.

The findings are useful—and actually used—because they are *practical* (concrete conclusions are reported that can be applied to improve programs), *ethical* (data was gathered in a way that showed respect for youth and program staff serving youth), and *accurate* (the data is meaningful and the findings are credible). The evaluation was worth what it cost because it was used to improve the work and inform future decision-making *(accountability)*.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

When a grantee submits evaluation data as part of a proposal, a foundation program officer asks:

1. How do you use the evaluation findings? What changes, improvements, or decisions have you made based on evaluation findings?
2. What is the process for collecting evaluation data? To what extent is the process practical, manageable, and sustainable?
3. How do staff and program participants experience the evaluation process? Do they find it meaningful and respectful?
4. How is accuracy ensured in data collection? What steps are taken to ensure that the evaluation findings are credible?
5. Based on your evaluation approach (and the answers to the preceding questions), what do you consider to be the strengths and weaknesses of your evaluation process and findings?

Note: These questions are asked only when grantees have made significant evaluation claims as part of the proposal.

BOTTOM LINE

Focus on evaluation use. Don't let evaluation become just compliance reporting.

23 COMPLETE EVALUATION REPORTING

THE ELEMENTS OF A COMPREHENSIVE EVALUATION REPORT

What? *What are the findings? What does the data say?*
So what? *What do the findings mean? Making interpretations and judgments.*
Now what? *Action implications and recommendations.*

Four distinct processes are involved in making sense of evaluation findings:

1. *Analysis* involves organizing raw data into an understandable form that reveals basic patterns and constitutes the evaluation's empirical findings, thereby answering the *what?* question.
2. *Interpretation* involves determining the significance of and explanations for the findings. This is Part One of answering the *so what?* question.
3. *Judgment* brings values to bear to determine merit, worth, and significance, including the extent to which the results are positive or negative, good or bad. This is Part Two of answering the *so what?* question.
4. *Recommendations* involve determining the action implications of the findings. This means answering the *now what?* question.

The graphic below depicts the inter-relationships among these four dimensions of evaluation sense making. The three fundamental questions—*What? So what? Now what?*—are connected to the four evaluation processes of (1) analyzing basic findings, (2) making interpretations, (3) rendering judgments, and (4) generating recommendations.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

To what extent does the proposal reflect evaluative thinking? If the program has past reports, do they reflect the distinctions between what, so what, and now what?

BOTTOM LINE
When reviewing an evaluation report, watch for distinctions between basic findings, interpretations, judgments, and recommendations—and the logical alignment and consistency among these elements.

24 UTILIZATION-FOCUSED EVALUATION

MAKE ATTENTION TO USE THE DRIVING FORCE BEHIND EVERY DECISION IN AN EVALUATION.

Utilization-focused evaluation begins with the premise that evaluations should be judged by their utility and actual use. Therefore, evaluators should facilitate the evaluation process and design an evaluation with careful consideration of how everything that is done, from beginning to end, will affect use.

Use concerns how real people in the real world apply evaluation findings and experience the evaluation process. Therefore, the focus in utilization-focused evaluation is on *intended use by intended users*.

— Who is the evaluation for?

— How is it intended to be used?

EXAMPLES OF DIFFERENT INTENDED USERS WITH LIKELY DIFFERENT INFORMATION NEEDS	EXAMPLES OF DIFFERENT USES OF EVALUATION
Program staff	Evaluation feedback to improve a program (formative evaluation).
Program director	Summative evaluation findings to decide whether to expand a model to new sites.
Government policymakers	Accountability evaluation to determine if funds were spent appropriately as intended, or to determine whether to invest in the program more broadly.

Utilization-focused evaluation does not advocate any particular evaluation content, model, method, theory, or even use. Rather, it is a process for helping primary intended users select the most appropriate content, model, methods, theory, and uses for their particular situation. Situational responsiveness guides the interactive process between evaluator and primary intended users. This means that the interactions between the evaluator and the primary intended users focus on fitting the evaluation to the particular situation with special sensitivity to context. A utilization-focused evaluation can include any evaluative purpose (formative, summative, developmental), any kind of data (quantitative, qualitative, mixed), any kind of design (e.g., naturalistic, experimental), and any kind of focus (processes, outcomes, impacts, costs, and cost-benefit, among many possibilities). *Utilization-focused evaluation is a process for making decisions about these issues in collaboration with an identified group of primary users, focusing on their intended uses of evaluation.*

A psychology of use undergirds and informs utilization-focused evaluation. Intended users are more likely to use evaluation if they understand and feel ownership of the evaluation process and findings. They are more likely to understand and feel ownership if they have been actively involved. By actively involving primary intended users, the evaluator is training users in use, preparing the groundwork for use, and reinforcing the intended utility of the evaluation every step along the way.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

How utilization focused is the evaluation portion of a proposal? Is evaluation just compliance and reporting oriented, or does there appear to be a commitment to making evaluation truly useful?

BOTTOM LINE

When reviewing an evaluation proposal or report, is it clear who is intended to use the evaluation and for what purposes?

25 DISTINGUISH DIFFERENT KINDS OF EVIDENCE

EVIDENCE-BASED PROGRAMS

The label "evidence based" is widely used. The question is: What does it mean, and what's the evidence?

Evidence about program effectiveness involves systematically gathering and carefully analyzing data about the extent to which observed outcomes can be attributed to a program's interventions.

Evaluators distinguish three types of evidence-based programs:

1. *Single-Summative*. Rigorous and credible summative evaluation of a single program.
2. *Meta-Analysis*. Systematic "meta-analysis" (statistical aggregation) of the results of a group of programs all implementing the same model in a high-fidelity, standardized, and replicable manner to determine best practices.
3. *Principles-Based*. Synthesis of the results of a group of diverse programs all adhering to the same principles but each adapting those principles to its own particular target population within its own context.

TYPE OF EVIDENCE-BASED PROGRAM	EXAMPLES	EVALUATION FOCUS AND FINDINGS
Single program summative	A local job-training program.	Evidence of the model's effectiveness for one particular site: Extensive, systematic, multi-year monitoring and evaluation data, including external summative evaluation on job placement and retention outcomes, will yield evidence-based conclusions about this particular program.
Meta-analysis	Results of implementing a standardized quality improvement and rating system for childcare providers in multiple sites.	Evidence of effectiveness across multiple sites: The quality-rating program is being implemented as a standardized, prescribed model, applying the same criteria and tool to all childcare providers. Systematic aggregate statistical analysis of standardized processes and outcomes will yield evidence-based best practices.
Principles-based synthesis	Youth homelessness work engaging programs operated by six organizations that share common principles and values but operate independently.	Evidence of effective principles: Each program is unique and provides different services but all work from a common set of principles of engagement, even though the implementation techniques built from those principles might vary from program to program. For example, "harm reduction" is a guiding principle. A synthesis of findings from case studies of their processes and outcomes will yield evidence-based effective principles.

PROPOSAL REVIEW AND SITE VISIT IMPLICATIONS

When a program claims to be evidence based, inquire into the nature of the evidence and the type of evidence-based program it aspires or claims to be.

BOTTOM LINE

Evidence-based programs must have evidence, but different kinds of evidence-based programs make different claims. Beware simple opinions masquerading as evidence. Beliefs are beliefs. Beliefs about program effectiveness must be evaluated to become an evidence-based program or model.

30 E. 7th St. Ste. 2900, St. Paul, MN 55101–2988
Main 651 227 8036 ***Toll-free*** 888 291 1123

OTTOBREMER.ORG