

A Program Director's Guide to Evaluating STEM Education Programs:

Lessons Learned from Local, State, and National Initiatives

April 2013



MASSACHUSETTS
Department of
Higher Education



Donahue Institute
OFFICE OF THE PRESIDENT

Introduction

This guide is designed for program directors/managers, educators, and others responsible for developing and implementing science, technology, engineering and mathematics (STEM) education programs. These individuals (hereafter referred to as “program leaders”) play a central role in managing and facilitating a program’s evaluation, and in sharing findings with staff, participants, funders, and other stakeholders. A major aspect of expanding STEM education programs is providing compelling evidence of their effect. A well-planned evaluation helps staff understand, assess, document, and communicate issues of program effectiveness related to both implementation (i.e., which program elements are successful) and outcomes (i.e., what impact did the program have on participants). These results are important for funders, program designers and/or implementers, and other stakeholders.

This guide is not intended to provide an exhaustive description of STEM education evaluation practice and strategies, but rather to offer an introductory overview of evaluation and the program leader’s role in managing and facilitating it. As such, the purpose of this guide is to:

- Provide evaluation guidelines and resources for program leaders who are implementing STEM programs in schools and community-based organizations.
- Reduce “evaluation anxiety” for individuals who are not professional evaluators by providing guidelines for good evaluation protocols.
- Offer suggestions on how to develop a comfortable and productive program-evaluation team relationship.
- Emphasize the role of program evaluation in helping inform STEM education practice and policy.

Drawing on the promising practices and lessons learned from national, state, and local programs, this guide provides resources for a variety of issues related to STEM education program evaluation. Formulation of this guide drew strongly on the expertise of staff associated with the following programs and organizations:

- The National Science Foundation’s Innovative Technology Experiences for Teachers and Students (ITEST) Learning Resource Center at EDC.
- The Massachusetts Department of Higher Education’s Pipeline Fund @Scale Program Initiative.
- The University of Massachusetts Donahue Institute.

More information about each of these programs/organizations, and the STEM projects they encompass, can be found in the References and Further Reading section.

The guide that follows is divided into four main sections, but it is not necessary to read through the document from beginning to end to benefit from the guidelines and resources that are included. Those who are new to evaluation may find the more concrete examples

included in planning and implementing the evaluation more useful for their work. More experienced individuals may draw more heavily on the framing concepts or principles outlined in the section on using evaluation findings. An outline of the guide is provided below, and the table of contents provides links to each section so the reader can access relevant sections as needed.

- I. **Framing Concepts** – reviews a variety of “big picture” approaches for framing an evaluation plan;
- II. **Planning the Evaluation of your STEM Program** – focuses on operational and evaluation planning;
- III. **Implementing the Evaluation** – reviews strategies for collecting data from participants and how to analyze results and findings;
- IV. **Using Evaluation Findings from your STEM Program** – suggests ways to use findings to inform programming as well as ways to disseminate findings to stakeholders.

While this guide is focused on STEM education evaluation, the principles are relevant to all types of evaluation. Consequently, the resources in this guide come from a broad array of disciplines in addition to STEM education. Throughout the entire guide, lessons from the ITEST and @Scale programs are integrated to provide real-world examples to illustrate and support major themes in each section. In addition, there are live web-links embedded throughout this document and highlighted in the figures labeled “Online Resources.” Related print resources are listed under the [References and Further Reading section](#). Finally, we view this release of the guide as a “version 1.0,” and plan to further refine and enhance the document based on additional feedback from program leaders and others in the STEM education field.

Joyce Malyn-Smith, Ed.D
Principal Investigator
ITEST Learning Resource Center
Education Development Center, Inc.

Alyssa Na'im
Research/Evaluation Specialist
ITEST Learning Resource Center
Education Development Center, Inc.

David Cedrone
Associate Commissioner for Economic &
Workforce Development
MA Department of Higher Education

Jean Supel
Research Manager
Donahue Institute
University of Massachusetts

Acknowledgements

We gratefully acknowledge our colleagues who provided thoughtful contributions to this guide:

Peter Christopher

Communications and Policy Associate, Massachusetts Afterschool Partnership

Enrique Colón-Bacó

Research Associate, Education Development Center, Inc.

Keith Connors

Program Manager, STEM Pipeline Fund

Leslie Goodyear, Ph.D.

Principal Research Scientist, Education Development Center, Inc.

Christopher M. Himes, Ph.D.

STEM Program Manager, Academic Affairs, Massachusetts College of Liberal Arts

Katherine Honey

STEM Coordinator, Southeastern Massachusetts STEM Network

Ellise M. LaMotte

Director of Academic Services, Olin College

Beverly Parsons, Ph.D.

Executive Director, InSites, Inc.

Phyllis Russell

President and Executive Director, Jr.Tech, Inc

Veronica Thomas, Ph.D.

Professor of Human Development, Howard University

Table of Contents

Framing Concepts.....	p. 1
Culturally and Contextually Responsive Evaluation.....	p. 1
Evaluating Scale-up Programs.....	p. 3
Using Systems Thinking in Evaluation.....	p. 5
Planning the Evaluation of Your STEM Program.....	p. 11
Operational Planning.....	p. 11
Criteria for strong STEM education evaluations.....	p. 11
Selecting an evaluator for your STEM program.....	p. 12
Budgeting for the evaluation of your STEM program.....	p. 13
Meeting evaluation needs of multiple STEM education stakeholders.....	p. 14
Evaluation Planning.....	p. 18
Understanding the STEM program.....	p. 18
Creating the logic model.....	p. 19
Articulating STEM education outcomes.....	p. 20
Developing evaluation questions for your STEM program.....	p. 22
Determining the evaluation design and methods for your STEM program.....	p. 23
Developing data collection procedures and measures for your STEM program.....	p. 24
Implementing the Evaluation.....	p. 28
Collecting data: coordinating with program participants and stakeholders.....	p. 28
Analyzing and interpreting data and findings.....	p. 29
Using Evaluation Findings from Your STEM program.....	p. 31
Using findings to inform STEM programming and practice.....	p. 31
Communicating findings to STEM education stakeholders.....	p. 31
Glossary.....	p. 34
References and Further Reading.....	p. 36
About ITEST and @Scale.....	p. 36
Selected Online Resources.....	p. 36
Selected Print Resources.....	p. 37

Framing Concepts

In order to build the nation's capacity to compete in the global economy, the current and future workforce must be equipped with the technical and practical skills required to advance innovation-driven industries. Careers in, and skills associated with, science, technology, engineering, and mathematics (STEM) disciplines are, and will continue to be, in high demand. Furthermore, the majority of STEM-related careers tend to have higher wages and better benefits associated with them than many non-STEM careers.

STEM education programs are deliberately designed to interest students in, and academically prepare them for, STEM career pathways. The current composition of the STEM workforce heavily influences the design and delivery of these programs. Very often STEM education programs target specific populations underrepresented in the STEM workforce or specific careers projected to have employment shortages. Through a variety of learning experiences, in-school and out-of-school STEM educators expose participants to innovative experiences that can enhance students' interest in, and preparation for, STEM careers.

Because STEM education programs are embedded in several overlapping contexts (educational, social, developmental, economic, etc.), program leaders and evaluators should think broadly about the approaches and tools used to assess their effects. Below we discuss three areas that have particular significance for evaluating STEM education programs. Recommendations for additional information associated with each area can be found in the Online Resources sidebar and the References and Further Reading section.

Culturally and Contextually Responsive Evaluation

Preparation for a career begins before an individual ever sets foot in a classroom and continues throughout a lifetime through both in- and out-of-school experiences. Research indicates that it is important for individuals to participate in a variety of career development experiences throughout their "learning life" for several reasons:

- 1) Individuals relate to "careers" in very different ways at different ages – a concept that is further affected by culture, gender, and personal experience. A girl in the 1st grade will relate to the concept of "being an engineer" very differently from a girl in the 7th or 11th grade, even more so than a boy. As a result, learning about careers needs to be strongly embedded in age- and socially/culturally-appropriate contexts that account for the (a) developmental stage, (b) academic background, (c) demographic characteristics and (d) social environment of participants.
- 2) Individuals "learn how to learn" about something through multiple exposures. A boy visiting a zoo for the 1st time will have a very different experience than one visiting for the 3rd time, even if they are alike in every other way. As a result, learning about careers needs to be "built" through stages through multiple interventions so that both interest and knowledge can grow over time.

- 3) Different individuals learn about, and become interested in, things differently. Where one student might become interested in engineering through watching a NASA video embedded in a classroom experience, another might do so through playing with parts of a jet engine at the local science museum. Having a variety of experiences available helps increase the likelihood of at least one of them being the “spark” that engages any one individual.

It is important to keep these issues in mind when developing, implementing, and interpreting an evaluation so that methods of observation and documentation, and understanding the results of the educational experience, are appropriate. Some examples:

- 1) While it may be budget-friendly to borrow a survey used to document STEM interest among middle school students for a group of 2nd graders, it may not be developmentally appropriate. Even if it is written simply enough to be read by the younger students, the questions asked may assume a level of psychological maturity that the younger students have not yet attained.
- 2) Just as there is “head room” in the real estate industry in terms of the potential for improvement of a property, so too is there “head room” in education in terms of the potential for increases in interest and preparation. A survey that has been used primarily to show changes among students who experience or learn something for the 1st time may not be able to show changes among students who are experiencing the activity or learning the concept for the 3rd time.
- 3) A pre- and post-survey of students who engaged an activity showed a slight increase in their interest in science. However, when the results were broken out, it could be seen that while interest increased a great deal for a few students, it remained flat for most. This does not necessarily mean that the program was “bad” or poorly implemented. Clearly it did have a positive effect on some. It could just mean that it wasn’t the right activity, at the right time, for many of the students. An evaluation that tells us not only what works, but also for whom and under what conditions could lead us to a better understanding of why.

There is well-established evidence that points to the value of using culturally and contextually responsive evaluation practices in STEM education programs. For STEM education programs that specifically target populations underrepresented in the STEM workforce – racial/ethnic minorities, females, and individuals with disabilities – being responsive to various cultures and cultural experiences requires the program leader and evaluator understand the context of the program under consideration and use appropriate methods and tools in the evaluation. For example, using an assessment that references the cultural ideals of one group may alienate or be misinterpreted by other groups. Similarly, out-of-school time programs have a certain culture (i.e., related to informal learning) where standardized assessments may be inappropriate because they are too formal or resemble the high-stakes testing setting that occurs during the school day.

The take-away message here is that the quality of an evaluation is directly related to the program leader’s and evaluator’s depth of understanding of the nature of the program and

its participants, which is totally separate from the idea of the evaluation itself being simple or complex. For example, it is possible for a small, simple evaluation to be well-designed, program- and participant-appropriate, and return culturally and contextually valid and useful information. On the other hand, it is equally possible for a large, complex evaluation to be poorly designed, program- and participant-inappropriate, and yield culturally inappropriate information, thus limiting the usefulness of the results information. Keep this in mind when you talk to a potential evaluator or review an evaluation proposal.

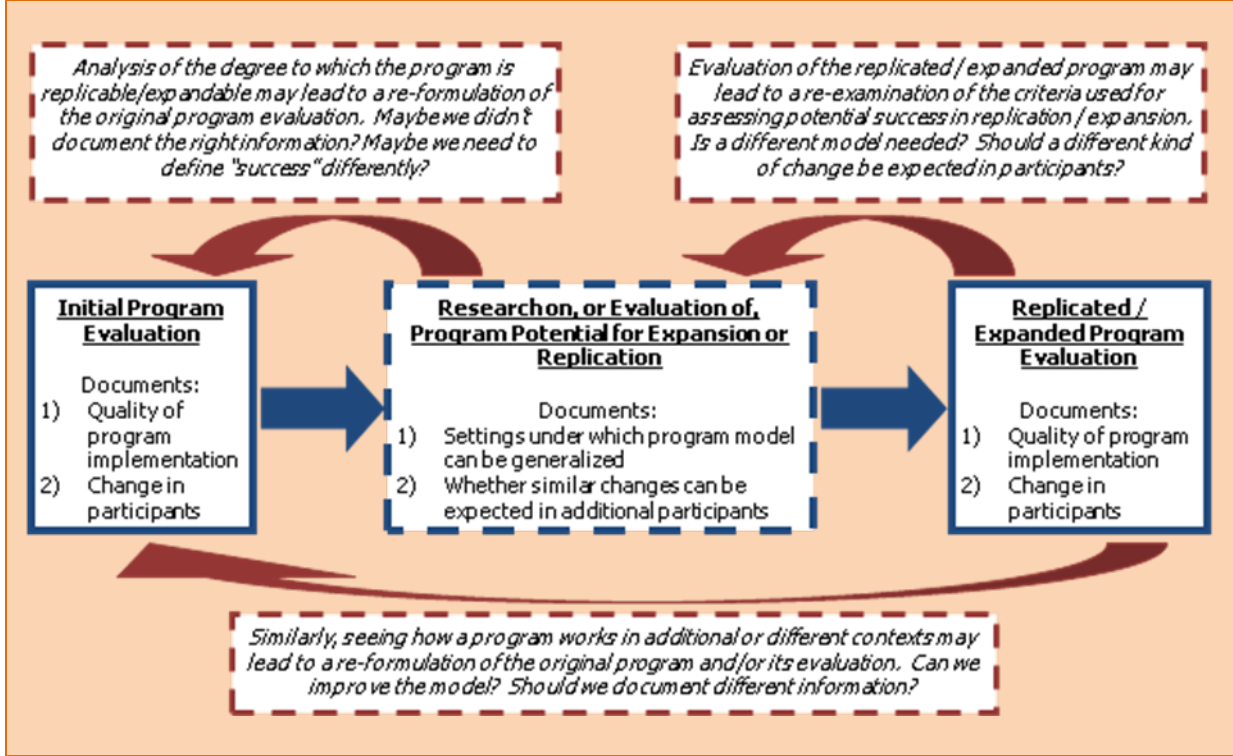
Evaluating Scale-up Programs

There is an increasing expectation of STEM programs to provide evidence of replication or scalability. Evaluations of these types of programs are initially focused on assessing the impact of the implementation of a single intervention or series of interventions, then move on to establishing evidence of successful replication in different settings under different circumstances. In both phases, there are intermediate and reinforcing steps that help refine the model. The evaluation may examine which types of activities are strongly correlated with participants' outcomes. For example, whether the "dosage" of key activities (duration and frequency) affects students' retention of STEM concepts. Or, if teachers who participate in online training significantly differ from those trained in-person in using hands-on activities with their students. The program leader can use such findings from the evaluation to further refine, then replicate, the STEM intervention.

The program leader and evaluator must acknowledge challenges that accompany generalizing local level STEM education interventions that occur with certain populations and under particular circumstances. From a programmatic perspective, the uniqueness of the local needs and contexts are not easily replicated or scaled up, and, in some cases, it may not be advantageous to do so. Acknowledging these difficulties and tailoring the evaluation to specifically explore these issues is a serious consideration when evaluating STEM education programs for these characteristics.

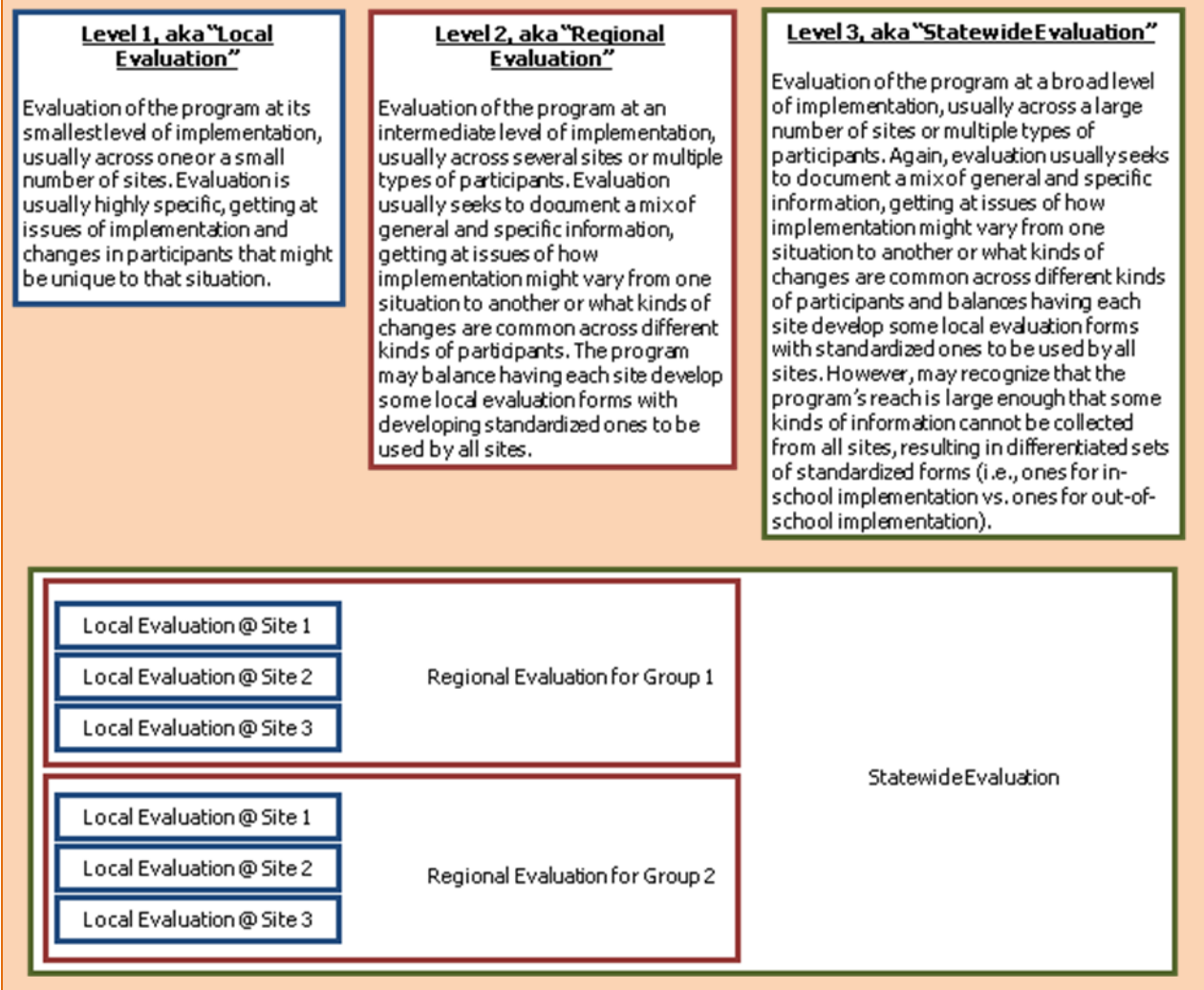
The take-away message here is that documentation of a program's success does not necessarily equate to documentation of potential successful expansion or replication. These are two different, although overlapping, questions. Figure 1 below provides a visualization of how the two types of evaluation interrelate:

Figure 1. Documenting potential expansion or replication



Furthermore, as programs expand or replicate, it is often necessary to implement multiple kinds of evaluations simultaneously in order to maintain integrity at both the original and replicated or scaled-up levels. Most common are only two levels, but it is possible for there to be even three or four. An example is shown in Figure 2 below.

Figure 2. Model of scaled-up program and evaluation



So, if you have a goal of eventually expanding or replicating your program, keep in mind the degree to which your evaluation plan is expandable and replicable. Alternatively, if you are looking to implement a program at your site that is part of an expansion/replication, keep in mind how you will balance collecting information necessary for evaluating your activities and participants specifically such that you do not overload yourself, your staff, or participants.

Using Systems Thinking in Evaluation

A system is "an interconnected set of elements that is coherently organized in a way that achieves something" (Meadows, 2008, p. XX). A systems thinking approach seeks to examine and understand both the parts and the relationships among the parts rather than

the parts on their own. STEM education evaluation benefits from systems thinking for a number of reasons:

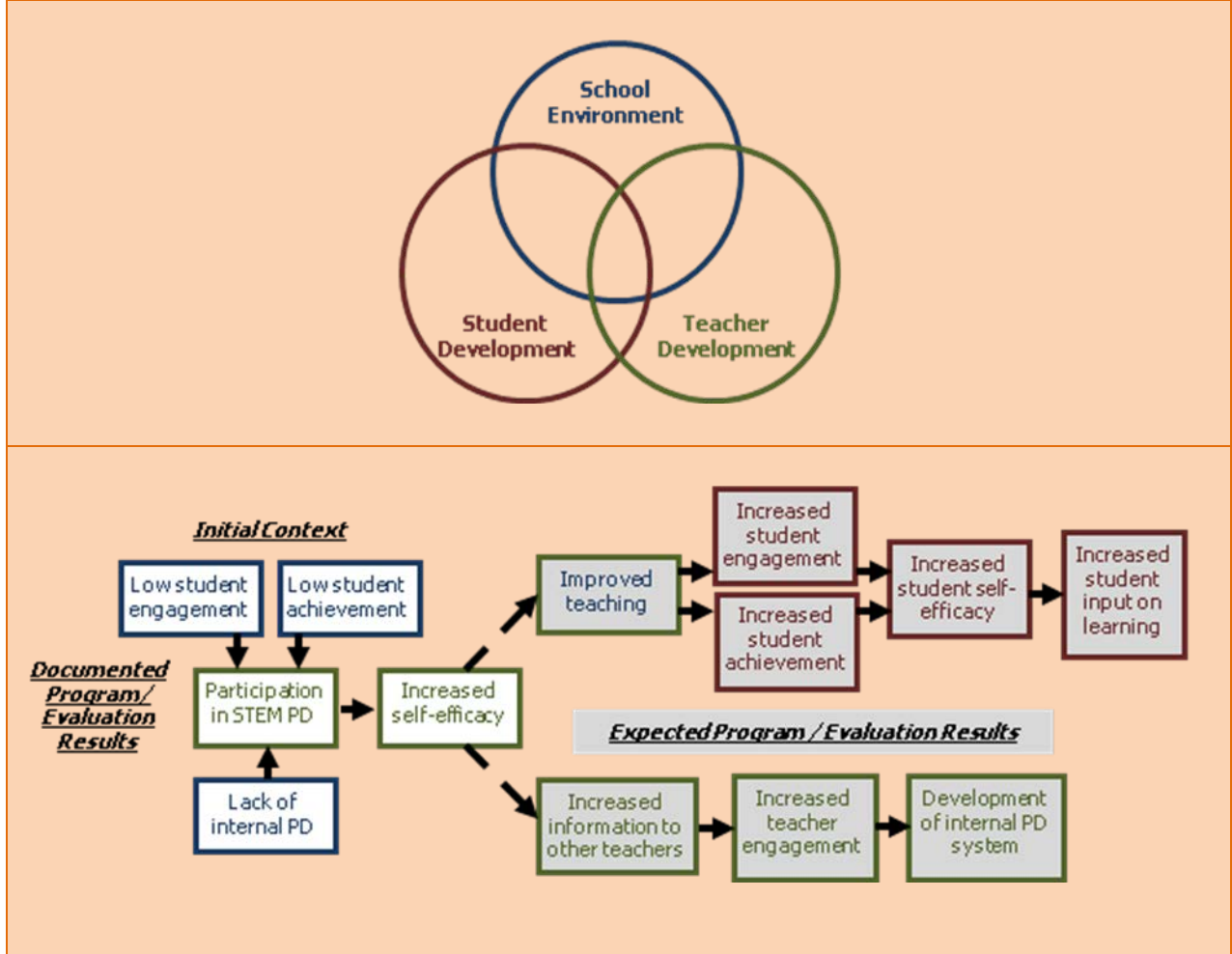
- STEM education programs exist within a system or network of both education and workforce development enterprises.
- The workforce development emphasis of STEM education programs relates to national, state and local economic priorities.
- STEM activities are often tasked with both increasing interest and increasing preparation.
- There is often a great deal of time between STEM education interventions (i.e., ones that occur during primary and secondary school) and the goal of these interventions, which is an increased number of new STEM employees (i.e., post-college employment).
- STEM education programs often bring together stakeholders from multiple sectors, or with varied priorities, such as business, PK-12 education, higher education, and non-profit organizations.

Systems thinking relates to program evaluation in two primary ways:

- 1) During the development of the program/evaluation logic model (during program planning):

Program funding is often limited to only a short period of time and the evaluation is limited to only current participants. This does not mean, however, that you should not consider the bigger picture of relationships and outcomes that contribute to, and are expected to result from, your program when you develop it and its evaluation. Even if your program and evaluation must be very limited in scope due to budget restrictions or funding requirements, you can still show the wider system of effects and feedback you expect to see over the long term through the development of a simple systems thinking diagram and a well-crafted logic model. For example, consider a one-time, two-week professional development seminar for teachers. The top portion of Figure 3 shows the primary systems that are expected to interrelate for that program. The bottom portion of Figure 3 provides a more elaborate diagram (a logic model) to show what the effects of the program would be beyond the program's scheduled evaluation period. Interestingly, the development of a logic model that links program development and evaluation itself is an exercise in systems thinking as insights gained through the representation of one aspect of the model may lead to a reconsideration of other aspects.

Figure 3. Using systems thinking to develop the logic model



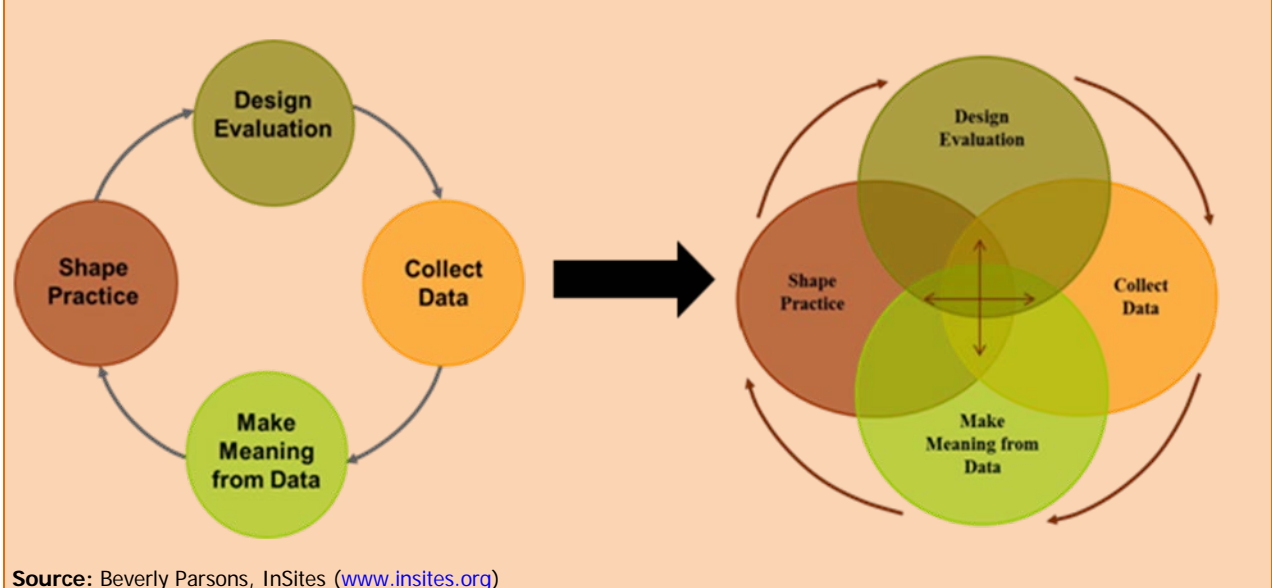
2) During the design and implementation of the evaluation:

One example of systems thinking within evaluation design is recognizing that the different stages of the evaluation process interrelate and affect one another, as depicted in Figure 4. Some examples of how these interrelationships manifest in practice include:

- What means are available for collecting data? If the phases of evaluation are viewed as more interconnected, the activities in one phase also can serve purposes within another phase. For example, focus groups structured with this in mind, can serve to not only collect data but also help make meaning from the data for participating stakeholders.
- Who is the intended audience for the evaluation? Systems thinking helps the evaluator see that data intended primarily for one audience can also be useful to other audiences whose work is connected to the project but they are not directly involved.

- What is the intended bigger change? Knowing this is central to defining what kind of data to collect and how meaning will be derived from the evaluation.

Figure 4. Traditional evaluation framework vs. systems-oriented evaluation



In this example, a systems-oriented evaluation draws on systems theory and modifies traditional approaches to evaluation. While the traditional evaluation has distinct phases and activities, the systems-oriented evaluation blurs the lines between these phases (Figure 4 above). Each phase including “shape practice,” which is typically seen as the last phase of the traditional evaluation framework, reinforces and informs the others. A systems oriented evaluation also brings attention to different evaluation questions and ways of interpreting data.

The take-away message here is that systems thinking leads program leaders, evaluators, and other STEM education stakeholders to consider how various components interrelate. By increasing their understanding of various interrelationships, it is possible for program leaders and evaluators to better understand the basis for, and expected outcomes of, their STEM education program.

The Online Resources below offer additional reports and tools about these framing concepts.



Online Resources

Culturally and Contextually Responsive Evaluation

- ❑ [The 2010 User-Friendly Handbook for Project Evaluation](#) – Chapter 7: A Guide to Conducting Culturally Responsive Evaluations (pp. 75-96)
- ❑ [American Evaluation Association \(AEA\) Statement on Cultural Competence in Evaluation](#)
- ❑ [Positionality Matters: Understanding Culture and Context from the Perspective of Key Stakeholders - Supplemental Resources](#) (ITEST Webinar slides and supplemental resources)

Evaluating Scale-up Programs

- ❑ [Chris Dede's Dimensions of Scale](#)
- ❑ [Harvard Family Research Project, Evaluation Exchange – Scaling Impact](#)
- ❑ [Examining the Depth Dimension of Scale for the GEAR-Tech-21 Project](#) (ITEST Project)
- ❑ [Scale-Up of ITEST Projects: Models, Challenges & Strategies](#) (ITEST Webinar)

Using Systems Thinking in Evaluation

- ❑ [Designing Initiative Evaluation: A Systems-orientated Framework for Evaluating Social Change Efforts](#)
- ❑ [Using Complexity Science Concepts When Designing System Interventions and Evaluations](#)
- ❑ [Questions that Matter: A Tool for Working in Complex Situations](#)
- ❑ [ZIPPER: A Mnemonic for Systems-Based Evaluation](#)
- ❑ [W.K. Kellogg Foundation's Designing Initiative Evaluation: A Systems-Oriented Framework for Evaluating Social Change Efforts](#)

Proceed to the next section, *Planning the Evaluation of your STEM Program*, to continue reading or click on one of the sections below:

Framing Concepts.....	p. 1
Culturally and Contextually Responsive Evaluation.....	p. 1
Evaluating Scale-up Programs.....	p. 3
Using Systems Thinking in Evaluation.....	p. 5
Planning the Evaluation of your STEM Program.....	p. 11
Operational Planning.....	p. 11
Criteria for strong STEM education evaluations.....	p. 11
Selecting an evaluator for your STEM program.....	p. 12
Budgeting for the evaluation of your STEM program.....	p. 13
Meeting evaluation needs of multiple STEM education stakeholders.....	p. 14
Evaluation Planning.....	p. 18
Understanding the STEM program.....	p. 18
Creating the logic model.....	p. 19
Articulating STEM education outcomes.....	p. 20
Developing evaluation questions for your STEM program.....	p. 22
Determining the evaluation design and methods for your STEM program.....	p. 23
Developing data collection procedures and measures for your STEM program.....	p. 24
Implementing the Evaluation.....	p. 28
Collecting data: coordinating with program participants and stakeholders.....	p. 28
Analyzing and interpreting data and findings.....	p. 29
Using Evaluation Findings from your STEM program.....	p. 31
Using findings to inform STEM programming and practice.....	p. 31
Communicating findings to STEM education stakeholders.....	p. 31
Glossary.....	p. 34
References and Further Reading.....	p. 36
About ITEST and @Scale.....	p. 36
Selected Online Resources.....	p. 36
Selected Print Resources.....	p. 37



Planning the Evaluation of your STEM Program

Planning the evaluation of a STEM education program involves both *operational planning* on the part of the STEM program leaders as well as *evaluation planning* on the part of both the evaluator and program leaders. This section describes these respective responsibilities and related tasks.

Operational Planning

Operational Planning refers to the primary evaluation-related duties for the program leader or program team, including determining criteria for strong STEM education evaluations¹, selecting an evaluator for your STEM program, budgeting for evaluation, and meeting evaluation needs of multiple stakeholders.

Criteria for strong STEM education evaluations

A program's criteria for a strong STEM education evaluation should draw on existing frameworks and standards established in the field of program evaluation. Below are a few major sources of best practices related to evaluation standards and quality:

- According to The Program Evaluation Standards (Yarbrough, D. B., Shulha, L. M., Hopson, R. K., and Caruthers, F. A., 2010), the core attributes of a quality evaluation are: utility, feasibility, propriety, accuracy, and accountability.² These standards and related statements of quality (30 in all) offer a list of criteria and considerations in support of strong program evaluation.
- The **W.K. Kellogg Foundation Evaluation Handbook** (W.K. Kellogg Foundation, 2004), states that a strong program evaluation should address three components: context, implementation, and outcome. According to the handbook, a good project evaluation should:
 - “examine how the project functions within the economic, social, and political environment of its community and project setting (context evaluation);
 - help with the planning, setting up, and carrying out of a project, as well as the documentation of the evolution of a project (implementation evaluation); and
 - assess the short- and long-term results of the project (outcome evaluation).” (p. 20)
- The **U.S. Government Accountability Office** (2012) lists four criteria for a good evaluation design:
 - 1) Be appropriate for the evaluation questions and context.
 - 2) Adequately address the evaluation question(s).
 - 3) Fit available time and resources.
 - 4) Rely on sufficient, credible data.

¹ The emphasis on “strong” relates to high integrity, not high cost. Part of setting up a high quality evaluation is recognizing what can be done well, even if it is very simple, with the resources available.

² Please refer to the *Glossary* at the end of this document for definitions to these and other underlined key terms.

- **The Guiding Principles for Evaluators** offers guidelines for sound, ethical practice for professional program evaluators. They also inform evaluation clients (and the general public) about the principles they can expect to be upheld by professional evaluators. The principles include:
 - **Systematic Inquiry:** Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.
 - **Competence:** Evaluators provide competent performance to stakeholders.
 - **Integrity/Honesty:** Evaluators ensure the honesty and integrity of the entire evaluation process.
 - **Respect for People:** Evaluators respect the security, dignity and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact.
 - **Responsibilities for General and Public Welfare:** Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

An interconnecting theme across these various sources is that the evaluation should be conducted in service to the client and stakeholders. Moreover, the evaluation must rely on sound and accurate data, and data sources, in order to inform appropriate decision making.

Selecting an evaluator for your STEM program

Choosing the evaluator for your STEM education program involves a range of considerations related to organizational and staff needs. The program staff must consider the internal resources and capacities to support a program evaluation, including:

- internal staff availability and expertise;
- site and communications logistics;
- availability and expertise of program volunteers;
- the timeline for implementing the evaluation and reporting results; and
- site, school, state and/or funder regulations.

While some organizations have internal departments that can conduct evaluation activities, *having* the capability doesn't mean you *can* or *should* use this capability. You should first weigh the value of hiring an external evaluator against using an in-house evaluator. External evaluators are seen as providing impartial judgments, thus lending credibility to the work, a perception that could be lost using an internal evaluator. Sometimes

What we have learned from ITEST

In addition to technical competence you should consider these questions in your selection of an evaluator:

- ◆ Does the evaluator specialize in the type of evaluation you want done (e.g., qualitative, quantitative, summative, formative)?
- ◆ Does the evaluator have experience evaluating STEM education programs?
- ◆ Is the evaluator knowledgeable about STEM education?
- ◆ Does the evaluator understand your STEM education program?

evaluator choices don't exist because program funding stipulations outline specific guidelines for the evaluation, such as requiring an external evaluator to assess the impact of the program.

The task of finding an external evaluator takes some effort, but the time and attention towards careful consideration is worth it if it benefits your program. Often, finding the right evaluator requires a competitive bidding process from a pool of qualified candidates. Personal recommendations from colleagues can also be valuable. However, be sure you vet recommended candidates, just as you would any candidate, with the general guidelines you identified for what you want from your evaluation (see criteria above) and the evaluator. An initial place to start your search is by asking colleagues who have hired evaluators to assess the impacts of their STEM education programs. They will likely be able to draw on their own experiences and suggest individuals who might fit your evaluation needs. Beyond that, there are a handful of online databases of evaluators, which are listed in the resources below. (See [Evaluator Directories](#) in the Online Resources below.)

Budgeting for the evaluation of your STEM program

Typically the budget for the STEM program evaluation is determined as a percentage of the program's total costs. Common budgets range from 5 percent (at the low end) to as much as 20 percent of the total funds allocated for the program. It is important to budget appropriately for the evaluator based on tasks and responsibilities, keeping in mind that an adequately-paid evaluator will garner numerous benefits for the program. A major aspect of determining the budget is identifying and estimating the various components of the evaluation. Program leaders are typically aware of these elements from their own experiences budgeting for operating expenses. Evaluation budgets should typically include estimates for:

- Staff, benefits, and consultant costs associated with the time spent on various evaluation-related activities (including things like planning, designing forms, conducting interviews or observations, administering surveys, transferring information from paper to computer, analyzing results, writing reports, making presentations, etc.)
- Travel expenses incurred by the evaluator
- Communications, printing, and mailing costs
- Supplies and equipment
- Subscriptions for online survey sites such as SurveyMonkey, or to purchase rights to a specialized survey (if necessary)
- Costs for meetings (e.g., space rental, travel for attendees) and/or incentives for participants
- The cost of obtaining approval for an evaluation that officially qualifies as research and for which an authorization is needed in order to survey and/or interview participants (often referred to as getting Institutional Review Board – or IRB – approval)
- General indirect or overhead costs associated with the company by which the evaluator is employed

In addition, the budget estimate can be increased or decreased by taking the following into account:

- Internal capacities/resources for handling evaluation-related tasks. Be very cautious in retaining any element of an evaluation because of the risk of quality control issues. While it's true you may save money by assigning an internal staff person to handle something as simple as transferring answers from paper surveys into a spreadsheet, it's also a possibility you could end up paying out more money to the evaluator for a mistake made by your employee requiring the evaluator (a) to review all of the data for potential other mistakes, and (b) to re-write any reports they based on your incorrect data.
- Whether the program has plans to seek grant-funding from new or additional sources whose evaluation requirements might be higher or different from the ones you are looking to fulfill in the near-term. If the answer is "Yes," then it would likely be a worthwhile investment to combine gathering information relevant to those purposes while simultaneously fulfilling your immediate evaluation needs.
- Issues of geographic proximity and/or use of technology. While finding a local evaluator may be desirable in many respects, those benefits might be outweighed by a misalignment of their skills to the job. In such situations, paying extra for travel expenses to obtain the right evaluator may be cost-effective over the long run. Depending on the nature of the evaluation work and the level of tech-suaveness of the program staff, communication tools (such as Skype and Adobe Connect) should be considered as possible alternatives for minimizing travel expenses.

Meeting evaluation needs of multiple STEM education stakeholders

While some may view the role of the program evaluator as that of a remote observer, collecting data and reporting findings, in reality an effective evaluator is deeply collaborative and engaged with the key stakeholders. Understanding their perspectives and interests gives breadth and depth to the evaluation, whether the evaluator is internal or external to the evaluation or program, or whether the stakeholders are engaged in the evaluation activities.

Stakeholders include any and all individuals or groups who are invested in the program and/or are affected by its outcomes. It is useful to think about stakeholders as belonging to three main groups: decision makers (e.g., funders; policy-makers; department, institution, or organization leaders), implementers (e.g., program staff, teachers), and recipients (e.g., students, parents, community) (Cronbach et al., 1980). It is also helpful to consider the "culture" of the stakeholder groups. This definition of culture includes not only typical references to beliefs, social norms, and practices of racial, ethnic, religious, and/or social groups, but also refers to values, goals, and practices of an institution or organization as well as those of a particular field or discipline. This definition provides a macro-level understanding, reflecting all stakeholders involved in a program and its evaluation.

Responding to stakeholders and involving them in the evaluation requires the program leaders to balance multiple goals:

- Stakeholders may have different evaluation needs. Funders are typically more interested in summative results showing impact of the program, while implementers are often concerned with formative information to guide program development.
- Evaluation design and data collection methods should accommodate not just language, age, and developmental requirements, but also situational contexts.
- Communication and reporting depends on the needs of the stakeholder. While all stakeholders should receive some information regarding program outcomes, what they receive, when they receive it, and the degree of detail that they receive depends on the goals of the program.

The topics in the second and third bullets above are addressed in greater detail in the remainder of this toolkit as we shift the focus to planning and implementing the evaluation, and later to using the evaluation findings.

What we have learned from @Scale

Just as too many cooks may spoil the soup, so too can too many stakeholders spoil an evaluation. It is entirely possible for an evaluation to run amuck in trying to please all stakeholders and, thereby overwhelming a program and its participants. Some potential consequences of “too much evaluation” include:

- ◆ **Evaluation Fatigue:** participants, whether children or adults, tire from too many survey and data requests resulting in numerous data collection problems (questions or forms not filled out, “gaming” surveys, etc.).
- ◆ **Data Overload:** excessive amounts of information without sufficient time for the evaluator or program leaders to process results in frustration and delay for them and the individuals/groups who requested it. This condition can also have longer term consequences that interrelate with “evaluation fatigue.” Participants become frustrated when they don’t see the information they gave being used, so they are less responsive to future requests.
- ◆ **Failure to Launch:** continuously analyzing surface-level data, though “interesting,” prevents the evaluator and program staff from delving more deeply into an area that is central to the program.



Online Resources

Evaluation Criteria and Standards

- [WK Kellogg Foundation Evaluation Handbook](#) (p. 20)
- [Designing Evaluations](#) (U.S. Government Accountability Office, 2012)
- [American Evaluation Association's Guiding Principles for Evaluators](#)

Guidelines for Selecting an Evaluator

- [The 2010 User-Friendly Handbook for Project Evaluation](#) – Appendix A
- [The U.S. Department of Health and Human Services \(2010\) Program Manager's Guide to Evaluation](#) – Chapter 3
- [WK Kellogg Foundation Evaluation Handbook](#) (p. 57-68)
- [When and How to Use External Evaluators](#) (Rutnik & Campbell, 2002)

Evaluator Directories

- [American Evaluation Association](#) (searchable by keyword and geography)
- [Evaluator Directory for EvaluATE](#) - Evaluation Resource Center for the NSF-funded Advanced Technological Education Program
- [Informal Science Evaluator Directory](#)

Budgeting for an Evaluator

- [A Checklist for Developing and Evaluating Evaluation Budgets](#) (Horn, 2001)
- [WK Kellogg Foundation Evaluation Handbook](#) (p. 56-57)

Engaging with and Meeting the Needs of Stakeholders

- [Positionality Matters: Understanding Culture and Context from the Perspective of Key Stakeholders](#)

Proceed to the next section to continue reading or click on one of the sections below:

Framing Concepts.....p. 1

Culturally and Contextually Responsive Evaluation.....p. 1

Evaluating Scale-up Programs.....p. 3

Using Systems Thinking in Evaluation.....p. 5

Planning the Evaluation of your STEM Program.....p. 11

Operational Planning.....p. 11

Criteria for strong STEM education evaluations.....p. 11

Selecting an evaluator for your STEM program.....p. 12

Budgeting for the evaluation of your STEM program.....p. 13

Meeting evaluation needs of multiple STEM education stakeholders.....p. 14

Evaluation Planning.....p. 18

Understanding the STEM program.....p. 18

Creating the logic model.....p. 19

Articulating STEM education outcomes.....p. 20

Developing evaluation questions for your STEM program.....p. 22

Determining the evaluation design and methods for your STEM program.....p. 23

Developing data collection procedures and measures for your STEM program.....p. 24

Implementing the Evaluation.....p. 28

Collecting data: coordinating with program participants and stakeholders.....p. 28

Analyzing and interpreting data and findings.....p. 29

Using Evaluation Findings from your STEM program.....p. 31

Using findings to inform STEM programming and practice.....p. 31

Communicating findings to STEM education stakeholders.....p. 31

Glossary.....p. 34

References and Further Reading.....p. 36

About ITEST and @Scale.....p. 36

Selected Online Resources.....p. 36

Selected Print Resources.....p. 37



Evaluation Planning

Planning the STEM evaluation requires close coordination and collaboration between the evaluator and the program leaders. Although the evaluator offers technical skills and expertise, he/she will need critical information about the program to make decisions about the appropriate scale and scope of the evaluation. This planning between evaluator and program leader involves acquiring a mutual understanding of the program, creating a logic model, articulating outcomes, developing evaluation questions, determining evaluation design and methods, and developing collection plans and measures. To make best use of your evaluator, you must provide them with a clear and deep understanding of your program and the essential questions you want answered.

Understanding the STEM program

The first step in the evaluation process involves understanding the program – gaining a clear picture of how, why, and under what conditions the program works. There are numerous reasons why the program leader should help the evaluator understand the program including to help the evaluator:

- Develop an accurate logic model.
- Articulate appropriate questions for the program’s goals.
- Understand where and when information can be collected.
- Decide on a timeline and methods for collecting and communicating information.

While it is important to take on the work of “understanding the STEM program” at the start of the evaluation, these elements can and should be revisited at all stages of the evaluation and programming. At a minimum, the evaluator and program team should revisit these components at key points in the program timeline in an effort to confirm anticipated results or identify the need to change direction (i.e., after a program begins, but before it reaches the half-way point; after the end of one cohort, but before the beginning of another; after the collection of data, but before reporting; etc.). It

is also important to engage key stakeholders in this process. Typically this work involves the evaluator and program staff who designed the program, but it may include those who deliver or implement the program (e.g., teachers), individuals or groups who have invested in the program (e.g., board members), and, in some cases, may even include the

What we have learned from ITEST

There are a variety of ways the program leader and evaluator can strengthen their collaboration. Successful strategies employed by ITEST program leaders and evaluators have included:

- ◆ Encouraging the evaluator to attend staff, advisory, and other program-related meetings and activities (in-person or electronically) to learn about the program, and build a relationship with the key stakeholders.
- ◆ Invite the program leader and other central program staff to sit in on evaluation meetings. These individuals can help inform discussions of instrument design, content, and format (i.e., what questions to ask and how to ask them).
- ◆ Regular meetings between the program leader and evaluator promote communication and collaboration. These check-ins can be used to report on progress, changes of direction, and program and evaluation management issues.

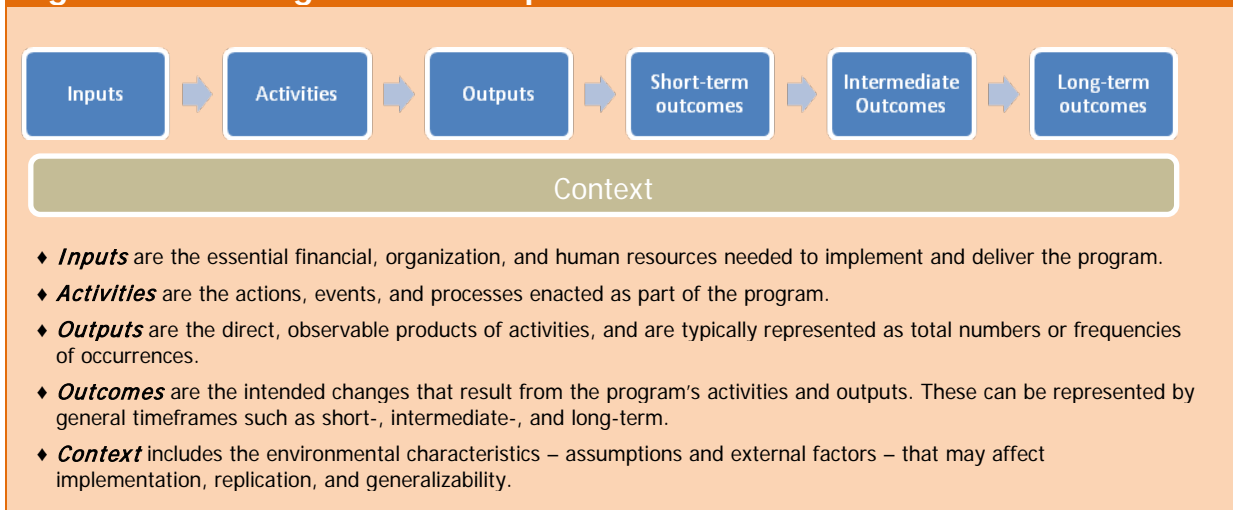
participants (e.g., students). Usually the evaluator takes the lead on these tasks and consults with stakeholders as appropriate.

Creating the logic model

Developing a logic model for the program, and articulating evaluation questions and intended outcomes, are initial iterative processes of the evaluation that inform one another and the direction of the evaluation itself. It may be that as the program logic model is developed, new evaluation questions and a better understanding of intended outcomes are generated. Similarly, as outcomes are clarified, essential components of the program model can be refined. Investing time at this stage will help lay a strong foundation for the evaluation as well as yield an in-depth understanding of the program for both the program leader and the evaluator.

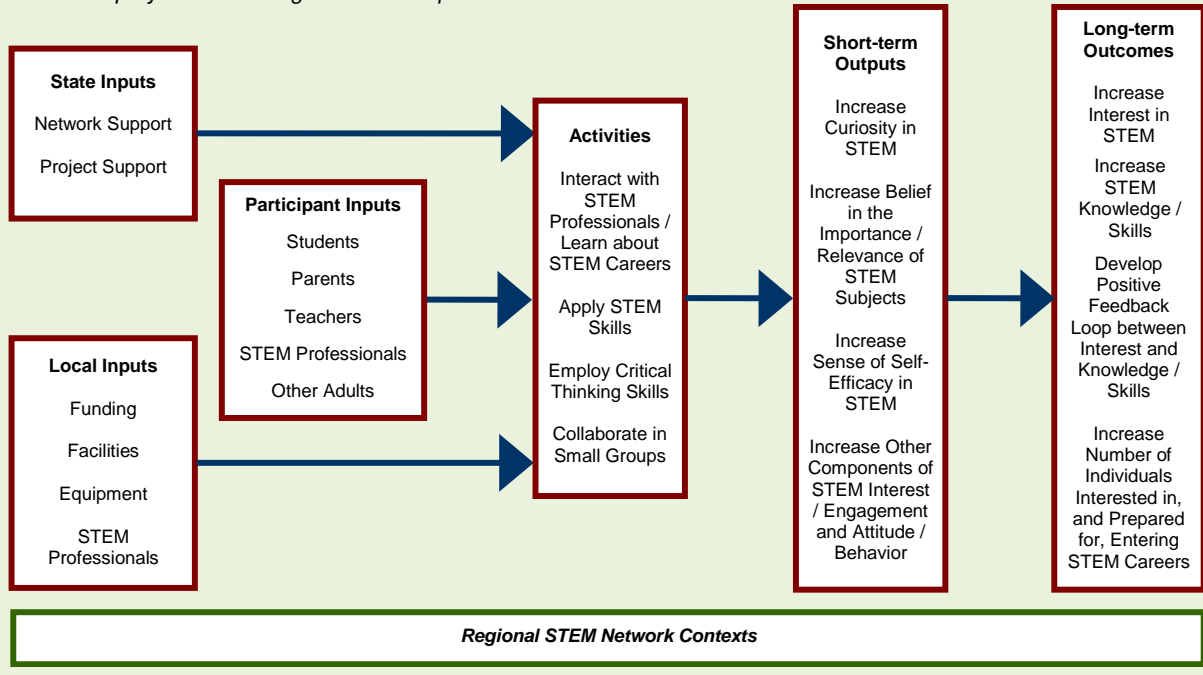
A logic model provides a visual diagram (often with text) that outlines how a program works along with its intended objectives and outcomes. The elements of a logic model identify the linkages among program resources or inputs, activities, outputs, and outcomes (see Figure 5). Evaluators can be a useful resource in identifying the type of logic model you choose and can work with you to draft one.

Figure 5. Basic logic model components



An example of a Logic Model from @Scale

Below is what might be seen as a “meta logic model” – that is, a logic model that was developed for an overarching statewide evaluation of multiple projects. This logic model was designed to inform standardized aspects of the statewide evaluation that would span all of the projects involved, while simultaneously aiding each individual project in defining its local or specific evaluation.



As the logic model above shows, STEM education programs often have similar overall short or long term outcomes. Despite these common overall outcomes, how a program achieves these outcomes may vary drastically. Often the process of developing the logic model helps program leaders identify the essential inputs and linkages for their STEM education program. This process will also help clarify underlying assumptions about how and why program works.

Articulating STEM education outcomes

Beyond identifying general intended STEM education outcomes during the process of developing the logic model, special attention needs to be focused on articulating **measurable** outcomes for the STEM education program and its evaluation. Defining the program’s outcomes helps the program staff, evaluators, and other stakeholders know what to look for and give clues on how to measure them. In general, outcomes are represented by positive changes in knowledge, skills, or behavior. In STEM education programs intended outcomes might include:

- Increased content knowledge
- Positive changes in pedagogical practices
- Increased knowledge and awareness of STEM careers
- Entry into STEM majors or careers.

It is also useful to consider any negative changes that might result from the program. When we introduce changes into a system, not all results are ones we desire.

What we have learned from ITEST

As the ITEST program has matured, projects have consistently identified and measured impacts in five main outcome areas:

- ◆ **Content Knowledge and Skills:** STEM subject matter knowledge
- ◆ **21st Century and other Skills:** collaboration, critical thinking, creativity, leadership
- ◆ **Affective Domains:** attitudes, confidence, interest, motivation persistence, self-efficacy
- ◆ **Career Development:** career decision-making, career interest, career knowledge
- ◆ **Teaching Practice:** teaching effectiveness, pedagogical knowledge, pedagogical practice

In the process of articulating the outcomes for their individual STEM education program, program leaders should also consider the STEM education policy climate of their state. If there are state goals regarding STEM education, how does their program contribute to them? In Massachusetts, the Governor's STEM Advisory Council has identified a set of core goals to advance the state's plan to address STEM education and workforce development. Figure 6 lists the six goals from the state's STEM plan,

Figure 6. Massachusetts STEM goals

1. Increase student interest in STEM.
2. Increase STEM achievement of PreK-12 students.
3. Increase the percentage of students who demonstrate readiness for college-level study in STEM fields.
4. Increase the number of students who graduate from a post-secondary institution with a degree in a STEM field.
5. Increase the number/percentage of STEM classes lead by effective educators, from PreK-16.
6. Align STEM education programs with the workforce needs of key economic sectors.

which can be applied to many STEM education programs. The full report, [A Foundation for the Future: Massachusetts' Plan for Excellence in STEM Education - Science, Technology, Engineering, and Mathematics](#), outlines the state's plan in greater detail.

Below are several online resources to help you in thinking about developing the logic model and articulating the outcomes for your STEM education program.

Online Resources

Logic Models and Outcomes

- ❑ [2010 User-Friendly Handbook for Project Evaluation \(pp. 15-23\)](#)
- ❑ [Evaluation Resource Center for Advanced Technological Education \(EvaluATE\) Webinar: Evaluation Tools](#)
- ❑ [University of Wisconsin-Extension's Logic Model Resources](#)
- ❑ [W.K. Kellogg Foundation's Logic Model Development Guide](#)
- ❑ [Center for Disease Control \(CDC\) Evaluation Workbook \(pp. 13-15\)](#)
- ❑ [Everything You Ever Wanted to Know about Logic Models But Were Afraid to Ask](#)
- ❑ [Matt Keene on Fuzzy Logic Models – Embracing and Navigating Complexity](#)

Developing evaluation questions for your STEM program

There are numerous approaches to developing evaluation questions. Often, initial questions are specified before a program is even evaluated. The process of developing a program's logic model and outlining its intended outcomes will likely shed light on particular questions that the program staff wants to have answered. Initial questions will be revised and refined accordingly.

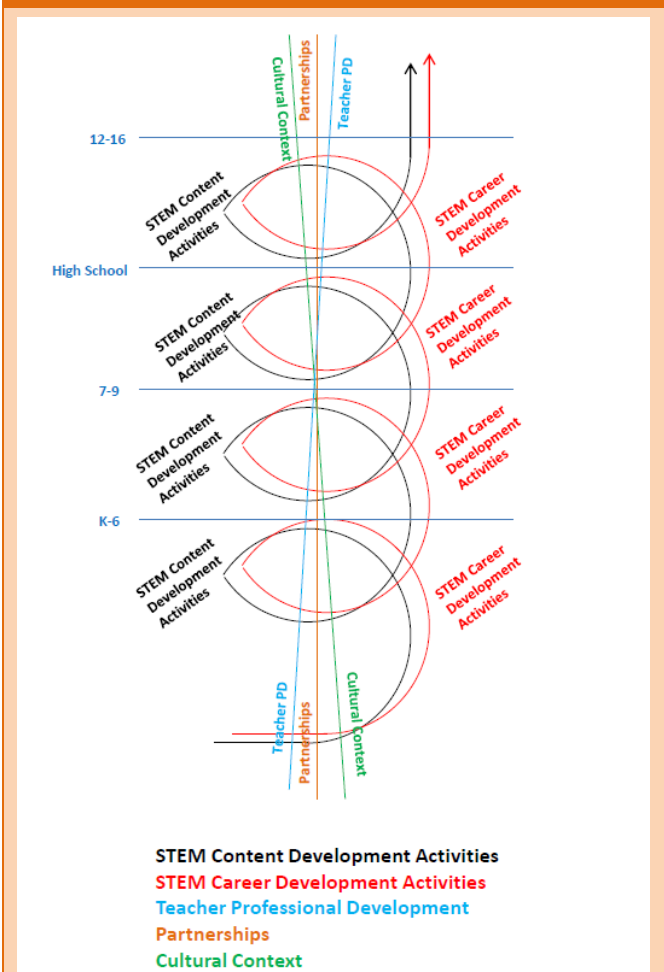
Evaluation questions will vary in scope, specificity, and number based on the type of program as well as the type of evaluation. Some evaluations will focus on *process or implementation* (How does a program work?), others will investigate *outcomes* (What was achieved?), and still others will examine a combination of *implementation and outcomes* (How did the program work, what was achieved, and why?). An evaluation that focuses not just on what happened, but also why, will reveal both program successes and challenges, and will help program staff and other stakeholders (like funders and policymakers) determine which elements of a program lead to better or poorer outcomes. In addition, funders increasingly ask programs to answer important questions beyond the individual project to advance the STEM education field.

What we have learned from ITEST

Over the last year the ITEST Learning Resource Center (LRC) and ITEST community formed a STEM Workforce Education Model Development Group to determine how best to learn about ITEST's role in STEM career development and preparing the next generation of STEM innovators. Although we found that many projects worked towards similar outcomes, the absence of a common framework and measures made it difficult to aggregate and learn from the efforts of the ITEST community at large.

In an effort to deepen our analysis and provide a common framework for the synthesis of lessons learned in the ITEST experience, the group explored career development literature and the ITEST portfolio of projects including projects' intended outcomes, evaluation questions and data collected. The ITEST STEM Workforce Education Helix that emerged from our work helped us visualize how different aspects of the ITEST experience come together to bring about the goal of youth's preparation to enter STEM careers. Those aspects visualized here and include: STEM Content Development Activities, STEM Career Development Activities, Partnerships, Educator Professional Development, and Cultural Context.

Figure 7. ITEST STEM Workforce Education Helix



What we have learned from ITEST (cont'd)

Our discussion of the above diagram led us to propose a Framework of Outcomes Related to Successful Pursuit of STEM Careers (click to link to the end of this section).

Based on this framework the following broad evaluation questions should be asked of projects preparing youth for STEM careers: In what ways does participation lead to...

1. Positive dispositions toward STEM Content and STEM Careers?
2. Increased STEM knowledge and STEM career knowledge?
3. Increased STEM Skills and STEM Career Skills?
4. Actions that increase likelihood that individuals will pursue STEM careers and STEM-enabled careers?

Determining the evaluation design and methods for your STEM program

The evaluation design depends in large part on the evaluation questions and what they aim to answer about the program. [Formative evaluation](#) provides information to help with decision-making and continuous improvement. [Summative evaluation](#) offers results at the end of a program or at a concluding period (e.g., at the end of one year or a multi-year period) and reports on the effectiveness of the program (which often informs decisions related to additional funding, expansion or replication). As noted above, evaluation questions fall into three main categories seeking to answer:

- How does the STEM education program work? (Implementation – for formative purposes)
- What did the STEM program achieve? (Outcomes – for summative purposes)
- How or why did the STEM program achieve its results? (Implementation and outcomes – for both)


Prioritizing which questions the evaluation will investigate, and to what extent it will explore them, will help determine the evaluation design. If the program is only interested in documenting what activities happen and how, a [descriptive analysis](#) of materials, curricula, activities, etc. will be sufficient. If, on the other hand, there is interest in describing the program as well as what changes occur as a result of implementation, a multi-faceted evaluation designed to document activities, assess outcomes, and potentially attribute [causation](#) would be required. This is an area where an evaluator has particular expertise, but is important for program staff to communicate their needs and offer guidance on the appropriateness of certain designs. Another consideration in the design of the evaluation is the need for formative or summative results for the program staff or other stakeholders. This is an area where communication between the evaluation and program teams is crucial to ensure common understanding on the purpose of the evaluation.

In addition to determining the appropriate design, the program leader and evaluator need to assess which methods to use for collecting data or other information. Generally, there are two methods for collecting information:

- 1) [Quantitative methods](#) - such as tests or surveys.
- 2) [Qualitative methods](#) - such as focus groups, interviews, or document reviews.

Qualitative methods are usually more time consuming than quantitative ones, but may yield richer descriptions and detect nuances in outcomes that quantitative methods cannot. However, quantitative methods may be more easily linked to other sources of evaluation (e.g., evaluation results from a similar program, or data from large, standardized surveys). Using a “[mixed methods approach](#)” – that is, utilizing both quantitative and qualitative methods within your evaluation – allows the program leader and evaluator to select the most appropriate tools for the various kinds of information that needs to be collected. Furthermore this approach allows the program to potentially connect its results to wider trends and patterns while simultaneously uncovering subtleties within the program.

Below are a few online resources that provide more technical detail on evaluation design and methods. Ultimately, the decision regarding which evaluation method to implement are the evaluator’s to make based on the jointly developed logic model, the evaluation questions the project leader seeks to have answered, and through an understanding of the kind of evidence needed to make claims about the program.

 **Online Resources**

Evaluation Design and Methods

- ❑ [The 2010 User-Friendly Handbook for Project Evaluation](#)
 - Data Collection Methods: Some Tips and Comparisons pp. 52-57
 - Review and Comparison of Selected Techniques pp. 58-74
- ❑ [Designing Evaluations](#) (U.S. Government Accountability Office, 2012)
- ❑ [Selecting an Appropriate Evaluation Design](#)
- ❑ [Fitting the Design to Context – ITEST Examples](#)

Developing data collection procedures and measures for your STEM program

Although evaluation questions and intended program outcomes help inform what types of data to collect, the types of measures used to assess program outcomes depend, in large part, on the evaluation design and methods. Another factor to consider is how change will be measured. In order to demonstrate an effect, it is useful to have both a “before” and an “after” program comparison. If the design allows for a comparison group, additional distinctions can be made against this group. Once the issue of outcomes, design, and methods has been resolved, the next step is to determine how to measure the program’s effects.

For programs seeking to collect quantitative data, surveys or tests are relatively easy to administer and analyze. Qualitative data can be collected via interviews, focus groups, or observations. As a starting point, program leaders should work with their evaluators to identify and locate existing tools and measures. These can be found through literature or internet searches. In addition, STEM programs that operate within a school, or those that are closely aligned with state standards, can coordinate with local or state education agencies to obtain tools or comparative data.

What we have learned from ITEST

In addition to traditional tools and measures, non-traditional sources of data through embedded assessments such as journals entries, scavenger hunts, or concept mapping activities can be used to enhance and illustrate examples of program effects.

A number of ITEST program leaders and evaluators have successfully used [embedded assessments](#) in their evaluations. During the Fourth Annual ITEST Principal Investigator Meeting, they shared their experiences with using embedded assessments:

- ◆ [A Multi-Pronged Approach to Embedded Assessment: Research and Evaluation in the GRADUATE Project](#)
- ◆ [Evidence of Success: Embedded Assessments in Photonics Leaders II \(PL2\)](#)
- ◆ [Using Embedded Assessment for Increasing Student Motivation and Teacher Engagement](#)
- ◆ [Using Student Performance Assessment as an Evaluation Tool](#)

The online resources listed below include a few different online databases where STEM education instruments can be found.



Online Resources

Online Instrument Databases and Examples

- ❑ [STEM Education Instrument Database](#)
- ❑ [Assessment Tools in Informal Science \(ATIS\)](#)
- ❑ [The ITEST Learning Resource Center's Online Evaluation Database: Examples from the collection](#)
- ❑ [Online Evaluation Resource Library \(OERL\)](#)
- ❑ [AAAS Project 2061 Science Assessment Website](#)



Framework of Outcomes Related to Successful Pursuit of STEM Careers

Desired Outcomes:

- *Employment in STEM Career*
- *Employment in STEM-enabled Career*

Examples of indicators	Categories			
	<p>Dispositions – Participation in ITEST projects leads to positive dispositions to STEM Content and STEM Careers. (STEM self-efficacy, interest, engagement, motivation, intentionality, agency, identity, resiliency, aspirations, ...)</p>	<p>Knowledge - Participation in ITEST projects leads to increased STEM Knowledge and STEM Career Knowledge. (Science, Technology, Engineering, and Math)</p>	<p>Skills - Participation in ITEST projects leads to increased STEM Skills and STEM Career Skills. (Application of Science, Technology, Engineering, and Math content and related tools)</p>	<p>Actions – Participation in ITEST projects, through the accumulation of positive dispositions and increased knowledge and skills, leads to actions that increase likelihood that students will pursue STEM careers and STEM-enabled careers. (Pursuit of in and out of school time STEM activities, Persistence in STEM Learning)</p>
STEM Content Outcomes	<p>Examples: Despite the lack of role models who look like me, I believe I'm good at science (Increased STEM Content Self-efficacy) I am interested in analyzing population data. (Increased STEM content interest)</p>	<p>Examples: Understands the relationship between DNA structure and function. (Science) Knows the quadratic equation (Math) Knows the principle if, then, else statement. (Technology)</p>	<p>Examples: Use probe ware to collect temperature differences in the soil (Science and Technology) Applies the iterative process of engineering design. (Engineering)</p>	<p>Examples: Students enroll in higher level STEM electives (Persistence in STEM learning) Students go to science museum with families (Pursuit of out of school time STEM activities) Participate in math competitions. (Pursuit of out of school time STEM activities)</p>
STEM Career Outcomes	<p>Examples: I believe I can be a geneticist (Increased STEM Career Self-efficacy) I want to be a statistician more than I did when the project began (Increased STEM Career Interest) There are career opportunities for me as a woman in computer science. (Identity)</p>	<p>Examples: Seismologists study earthquake data to determine the shifting pattern of tectonic plates. (Science, Technology and Math) Aware of the required courses and skills for a career in cyber security. (Technology and Engineering)</p>	<p>Examples: Propose different uses for a vacant lot sites in their neighborhood after collecting site data using urban planning tools (Technology tools used by scientists to solve problems) Develop an app to teach mathematics to children (Mathematics, Technology and Engineering)</p>	<p>Examples: Seeks volunteer, internship, and /or work-based STEM career experiences. Prepares a STEM career portfolio.</p>

(STEM career pathways, STEM careers vs STEM-enabled careers, types of problems commonly solved by STEM professionals, skills/knowledge needed by technicians, technologists and other STEM professionals for workplace success,...)

Outcomes:

- *Increased Persistence in STEM Learning*

Proceed to the next section, *Implementing the Evaluation*, to continue reading or click on one of the sections below:

Framing Concepts.....	p. 1
Culturally and Contextually Responsive Evaluation.....	p. 1
Evaluating Scale-up Programs.....	p. 3
Using Systems Thinking in Evaluation.....	p. 5
Planning the Evaluation of your STEM Program.....	p. 11
Operational Planning.....	p. 11
Criteria for strong STEM education evaluations.....	p. 11
Selecting an evaluator for your STEM program.....	p. 12
Budgeting for the evaluation of your STEM program.....	p. 13
Meeting evaluation needs of multiple STEM education stakeholders.....	p. 14
Evaluation Planning.....	p. 18
Understanding the STEM program.....	p. 18
Creating the logic model.....	p. 19
Articulating STEM education outcomes.....	p. 20
Developing evaluation questions for your STEM program.....	p. 22
Determining the evaluation design and methods for your STEM program.....	p. 23
Developing data collection procedures and measures for your STEM program.....	p. 24
Implementing the Evaluation.....	p. 28
Collecting data: coordinating with program participants and stakeholders.....	p. 28
Analyzing and interpreting data and findings.....	p. 29
Using Evaluation Findings from your STEM program.....	p. 31
Using findings to inform STEM programming and practice.....	p. 31
Communicating findings to STEM education stakeholders.....	p. 31
Glossary.....	p. 34
References and Further Reading.....	p. 36
About ITEST and @Scale.....	p. 36
Selected Online Resources.....	p. 36
Selected Print Resources.....	p. 37



Implementing the Evaluation

Once the major work of planning the evaluation is complete, the implementation phase occurs. This involves collecting data from participants and other stakeholders, then analyzing and interpreting data and findings.

Collecting data: coordinating with program participants and stakeholders

During the planning phase, thought must be given to how the data will be collected. It is during the implementation phase, however, that the evaluator and program leader or staff must manage the practical implications of these choices. Before any data are collected, the program leader and evaluator must determine if permission needs to be obtained from participants (or their guardians) and clearance granted by the organizations or schools involved. Once this has been determined, the evaluator and program leader should work together to identify specific processes and procedures to collect data.

The process of collecting data involves working closely with program participants and stakeholders to ensure that adequate data are collected (in terms of response rates) and that the data collected are accurate (in terms of quality). The program leaders provide a direct connection to program participants and should assist with scheduling access and other communications issues. The program leaders should also be mindful of overloading participants with too many requests for data, which can lead to [“evaluation fatigue.”](#) The evaluation and program teams should

What we have learned from ITEST

Over the last few years through the Annual Evaluation Peer Exchange – a two week online dialogue where ITEST evaluators, Principal Investigators, and staff share their best practices and lessons learned regarding evaluation – several re-emerging themes about data collection have surfaced. Below are several highlights from their recommendations:

- ◆ Carefully plan the timing of data collection to coincide with program activities and avoid interceding events (such as school vacations, holidays, and testing periods). Provide advance notice and repeated follow up with participants or those in charge of data collection.
- ◆ When possible, look for activities in other programs that might support data needed for your evaluation. Work together to tailor tools to serve both teams' needs.
- ◆ Provide positive feedback to participants. For teachers, share information with their principals, while maintaining respondents' confidentiality.
- ◆ If possible, provide incentives such as stipends, prizes, or awards for completed evaluation tasks.
- ◆ Make sure that activities involving teachers have a direct connection to their curriculum so the evaluation is not perceived as extra work.
- ◆ Incorporate face-to-face interactions when possible. These meetings help provide far richer data than often can be collected through a survey. If meeting in person, feed your participants!
- ◆ Share preliminary as well as final findings so participants (teachers and students, when appropriate) can see the value of the evaluation and the relevance to their effort.

work together to develop a plan to collect data that meets their respective needs with a balance towards not overburdening program participants and administrators.

Analyzing and interpreting data and findings

Typically the evaluator is responsible for analyzing data and findings from the STEM program, but the program leader plays an important role in contextualizing the results. Overall, the analysis should be framed by the evaluation questions and outcomes that are identified in the early steps of planning the evaluation. However, there should be room to look for and explore unintended outcomes or consequences of the STEM program and relationships to other parts of the system. The program leader can play an important role in adding narrative to numbers or details behind evaluation findings.

The types of analyses will depend upon the design and methods used in the program evaluation, but generally should identify patterns or commonalities, make comparisons or contrasts, and, in some cases, examine causation or [attribution](#). The analyses also relate back to the evaluation questions; the evaluator should use these questions as a core reference point. Analyzing quantitative data can range from simple descriptions (i.e., “the average score on the post-test was 20% higher than the average score on the pre-test”) to advanced statistical methods that assess the degree to which different participant characteristics related to the change they experienced. Qualitative data analysis can vary from narrative descriptions (for small samples or datasets) to [thematic coding](#) or [content analysis](#) (for larger samples or datasets) and generally takes more time and effort to analyze. Both types of analyses should be organized so that they begin by exploring individual changes or specific findings and then broaden to examine and interpret the results as they relate to larger issues. For example, “What do the results mean for the program and its participants? Are they significant or meaningful to the larger field of STEM education? If so, how?” As a whole, this will tell the story of the program and its impact; the findings may also help situate the program relative to broader educational or workforce development systems and priorities.

Proceed to the next section, *Using Evaluation Findings from Your STEM Program*, to continue reading or click on one of the sections below:

Framing Concepts.....p. 1

Culturally and Contextually Responsive Evaluation.....p. 1

Evaluating Scale-up Programs.....p. 3

Using Systems Thinking in Evaluation.....p. 5

Planning the Evaluation of your STEM Program.....p. 11

Operational Planning.....p. 11

Criteria for strong STEM education evaluations.....p. 11

Selecting an evaluator for your STEM program.....p. 12

Budgeting for the evaluation of your STEM program.....p. 13

Meeting evaluation needs of multiple STEM education stakeholders.....p. 14

Evaluation Planning.....p. 18

Understanding the STEM program.....p. 18

Creating the logic model.....p. 19

Articulating STEM education outcomes.....p. 20

Developing evaluation questions for your STEM program.....p. 22

Determining the evaluation design and methods for your STEM program.....p. 23

Developing data collection procedures and measures for your STEM program.....p. 24

Implementing the Evaluation.....p. 28

Collecting data: coordinating with program participants and stakeholders.....p. 28

Analyzing and interpreting data and findings.....p. 29

Using Evaluation Findings from your STEM program.....p. 31

Using findings to inform STEM programming and practice.....p. 31

Communicating findings to STEM education stakeholders.....p. 31

Glossary.....p. 34

References and Further Reading.....p. 36

About ITEST and @Scale.....p. 36

Selected Online Resources.....p. 36

Selected Print Resources.....p. 37



Using Evaluation Findings from your STEM program

An important feedback loop from the evaluation is the application of findings to inform and improve programming. A key aspect of this work involves communicating findings to STEM education stakeholders.

Using findings to inform STEM programming and practice

A well-planned and implemented evaluation will position the program leaders to gather data they can use to communicate the program's successes and lessons learned. Beyond serving the needs of the individual program, evaluation results can potentially contribute to broader learning and help advance knowledge about particular interventions or approaches to delivering STEM education programs.

One of the primary uses of evaluation findings is to help the program leader determine the effectiveness of his or her program. For example, for a STEM education program, one can learn the effect the program has on participants' STEM knowledge, skills, and/or attitudes. This information might be used to highlight successful approaches while also identifying areas for improvement. In this way, the program leader and staff can conclude what works and what might need to be modified or eliminated entirely. For STEM education programs charged with replication or scale-up interventions, the evaluation will help determine key components of the model to ensure fidelity in expansion.

An additional use of evaluation findings is to situate the individual program within the larger field of STEM education and workforce development. While most programs and stakeholders are chiefly concerned with the local implementation, there is value in sharing findings with other STEM education practitioners and policymakers. For example, it may be useful to determine how the program measures up to state or national goals regarding STEM education or compares with other programs in the state.

Communicating findings to STEM education stakeholders

Communicating evaluation findings to STEM education stakeholders can take a variety of forms. The evaluator will likely provide interim, annual, and/or final reports to the program leader and staff. These reports can, in turn, be re-distributed to other stakeholders. Evaluation results can also be conveyed through any standard communication tools the program might use on a regular basis, such as websites, newsletters, listservs, etc. Working alone, or with the help of the evaluator, the program leader should consider the various stakeholder groups – decision makers, implementers, and recipients – and determine exactly who should receive evaluation findings and in what format.

Emphasizing the use of evaluation findings and communicating them to various stakeholders helps establish and build buy-in for the evaluation process and shows transparency. Furthermore, communicating evaluation results completes the relationship link among the program leader with the participants and funders while validating their engagement as stakeholders. Program staff and others most immediately affected by the evaluation will see the value in collecting data from participants and those associated with them (such as parents and peers for youth participations or peers and administrators for educators or other adult participants). At an even higher level, within the STEM education field in general, publicly disseminating evaluation findings will help build the knowledge base for other practitioners, evaluators, and policy makers.



Online Resources

- [Utilization-Focused Evaluation \(U-FE\) Checklist](#)
- [Evaluation Resource Center for Advanced Technological Education \(EvaluATE\) Webinar: Maximizing Evaluation Impact](#)
- [The Basics of Evaluation Reporting](#)

Do you want to return to an earlier section of this document? Click on one listed below.

Framing Concepts.....	p. 1
Culturally and Contextually Responsive Evaluation.....	p. 1
Evaluating Scale-up Programs.....	p. 3
Using Systems Thinking in Evaluation.....	p. 5
Planning the Evaluation of your STEM Program.....	p. 11
Operational Planning.....	p. 11
Criteria for strong STEM education evaluations.....	p. 11
Selecting an evaluator for your STEM program.....	p. 12
Budgeting for the evaluation of your STEM program.....	p. 13
Meeting evaluation needs of multiple STEM education stakeholders.....	p. 14
Evaluation Planning.....	p. 18
Understanding the STEM program.....	p. 18
Creating the logic model.....	p. 19
Articulating STEM education outcomes.....	p. 20
Developing evaluation questions for your STEM program.....	p. 22
Determining the evaluation design and methods for your STEM program.....	p. 23
Developing data collection procedures and measures for your STEM program.....	p. 24
Implementing the Evaluation.....	p. 28
Collecting data: coordinating with program participants and stakeholders.....	p. 28
Analyzing and interpreting data and findings.....	p. 29
Using Evaluation Findings from your STEM program.....	p. 31
Using findings to inform STEM programming and practice.....	p. 31
Communicating findings to STEM education stakeholders.....	p. 31
Glossary.....	p. 34
References and Further Reading.....	p. 36
About ITEST and @Scale.....	p. 36
Selected Online Resources.....	p. 36
Selected Print Resources.....	p. 37



Glossary

Accountability Standards: Refers to program evaluation standards (from the Joint Committee on Standards for Educational Evaluation, JCSEE) that encourage adequate documentation of evaluations and a metaevaluative perspective focused on improvement and accountability for evaluation processes and products.

Accuracy Standards: Refers to program evaluation standards (JCSEE) that are intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality.

Activities: The actions, events, and processes enacted as part of the program.

Attribution: Assigning some quality or character to a person or thing.

Causation: Causation is the relation of cause to effect. In evaluation it may be useful to determine which elements contribute strongest to intended outcomes by establishing causal links in the development of the theoretical underpinnings or logic model as well as through statistical analyses of results from the program.

Content Analysis: Analytic approach used to determine the presence of certain words, concepts, themes, phrases, characters, or sentences within texts or sets of texts and to quantify this presence in an objective manner.

Context: The environmental characteristics – assumptions and external factors – that may affect implementation, replication, and generalizability.

Descriptive Analysis: Information and findings derived from text unlike statistical or quantitative findings that are derived from numbers.

Embedded Assessment: Alternative methods for assessing performance and progress that are integrated into instructional materials and are virtually indistinguishable from day-to-day activities.

Evaluation Fatigue: As a result of participants and other stakeholders being overburdened by evaluation components, especially excessive data collection, they become exhausted by the evaluation, which can lead to poor or incomplete data.

Feasibility Standards: Refers to program evaluation standards (JCSEE) that are intended to increase evaluation effectiveness and efficiency.

Formative Evaluation: Evaluation that provides information as the program is being developed or refined to help with continuous improvement.

Inputs: The essential financial, organization, and human resources needed to implement and deliver the program.

Logic Model: A visual diagram (often with text) that outlines how a program works along with its intended objectives and outcomes.

Mixed Methods Evaluation: An evaluation where the design includes the use of both quantitative and qualitative methods for data collection and analysis.

Outcomes: The intended changes that result from the program's activities and outputs. These can be represented by general timeframes such as short-, intermediate-, and long-term.

Outputs: The direct, observable products of activities, and are typically represented as total numbers or frequencies of occurrences.

Propriety Standards: Refers to program evaluation standards (JCSEE) that support what is proper, fair, legal, right and just in evaluations.

Qualitative: Using approaches in an evaluation that primarily focuses on descriptive and interpretive tools and methods.

Quantitative: Using techniques that involve the use of numerical measurement and data analysis based on statistical tools and methods.

Summative Evaluation: Evaluation that provides results at the end of a program or at a concluding period (e.g., at end the end of one year or multi-year period) and reports on the effectiveness of the program, which is often used to assist with decisions such as additional funding, replication, or even scale-up.

Thematic Coding: Most often used with qualitative data analysis. Thematic coding involves examining and recording patterns (or "themes") within data.

Utility Standards: Refers to program evaluation standards (JCSEE) that are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.

References and Further Reading

About ITEST and @Scale

The ITEST program was established by the National Science Foundation (NSF) in direct response to current concerns and projections about the growing demand for and shortages of STEM (science, technology, engineering, and mathematics) professionals in the United States; the program seeks solutions to help ensure the breadth and depth of the STEM workforce. Since 2003, ITEST has funded more than 195 projects focusing on engineering; computer, biological, and environmental sciences; and mathematics, impacting more than 225,000 students, 8,000 educators, and 3,000 parents and caregivers. The program also funds the [ITEST Learning Resource Center](#) to facilitate the ITEST Community of Practice, through individual and targeted technical support, and oversees the national dissemination of program models, materials, and best practices from the program.

Since 2003, the Massachusetts Department of Higher Education has administered the STEM Pipeline Fund. The STEM Pipeline Fund was established in association with Economic Stimulus Legislation to: (1) increase the number of Massachusetts students who participate in programs that support careers in fields related to STEM; (2) increase the number of qualified STEM teachers; and. (3) improve the STEM educational offerings available in public and private schools. Since 2003, the STEM Pipeline Fund has sponsored over 20 different projects that served over 10,000 students from over 150 school districts across the state. The @Scale initiative, advanced in 2012 by the Massachusetts Governor's STEM Advisory Council, is a strategic effort to focus public – STEM Pipeline Fund – and private resources in support of an integrated portfolio of education and workforce development projects aligned with the goals of the Commonwealth's [STEM Plan](#). Projects within this portfolio address STEM from Pre-K to college/workforce, and contribute to increasing the number and diversity of STEM-career interested, and prepared, individuals.

Selected Online Resources

Below is a list of additional, evaluation-related websites that contain a wealth of resources. While some of these have been referenced in earlier, Online Resources sidebars, they are included again here for your convenience:

[A Foundation for the Future: Massachusetts' Plan for Excellence in STEM Education](#)

[Bureau of Justice Assistance – Guide to Program Evaluation](#)

[Center for Advancement of Informal Science Education - Principal Investigator's Guide: Managing Evaluation in Informal STEM Education Projects](#)

[Centers for Disease Control and Prevention – Capacity Building Evaluation Guide](#)

[Centers for Disease Control and Prevention – Developing an Evaluation Plan](#)

[Centers for Disease Control and Prevention - Introduction to Program Evaluation for Public Health Programs: A Self-Study Guide](#)

[Evaluation Toolkit for Magnet School Programs](#)

[InSites – Communities of Learning, Inquiry, and Practice \(CLIPs\)](#)

[MassInsight Education – Evaluation Toolkit](#)

[McCormick Foundation – Program Evaluation Guide](#)

[National Staff Development Council – Teacher Professional Development Evaluation Guide](#)

[OECD Best Practice Guidelines for Evaluation](#)

[Pell Institute – Evaluation Toolkit](#)

[Research Councils UK – Evaluation: Practical Guidelines](#)

[The Educators’ Guide to Service-Learning Program Evaluation](#)

[The Evaluation Center at Western Michigan University](#)

[United Nations Development Program – Outcome Level Evaluation](#)

[US Department of Health and Human Services – Program Manager’s Guide to Evaluation](#)

Selected Print Resources

Below is a list of selected, recent articles related to issues that have been discussed in this document. Additional information on the issues, as well as suggestions for further reading, can be found in the journals from which these have been drawn, and in the references within each article.

Appleton-Dyer, Sarah, Janet Clinton, Peter Carswell, and Rob McNeill. *Understanding Evaluation Influence Within Public Sector Partnerships: A Conceptual Model*. American Journal of Evaluation. December 2012: 532-546.

Azzam, Tarek. *Evaluator Characteristics and Methodological Choice*. American Journal of Evaluation. September 2011: 376-391.

Botcheva, Luba, Johanna Shih, and Lynne C. Huffman. *Emphasizing Cultural Competence in Evaluation: A Process-Oriented Approach*. American Journal of Evaluation. June 2009: 176-188

Brousselle, Astrid, and François Champagne. *Program Theory Evaluation: Logic Analysis*. Evaluation and Program Planning. February 2011, Vol. 34, Is. 1: 69-78.

Bryson, John M., Michael Quinn Patton, and Ruth A. Bowman. *Working with Evaluation Stakeholders: A Rationale, Step-wise Approach and Toolkit*. Evaluation and Program Planning. February 2011, Vol. 34, Is. 1: 1-12.

Cartland, Jenifer, Holly S. Ruch-Ross, Maryann Mason, and William Donohue. *Role Sharing Between Evaluators and Stakeholders in Practice*. American Journal of Evaluation. December 2008: 460-477.

Cooksy, Leslie J., and Melvin M. Mark. *Influences on Evaluation Quality*. American Journal of Evaluation. March 2012: 79-84.

Guerra-López, Ingrid, and Sacip Toker. *An Application of the Impact Evaluation Process for Designing a Performance Measurement and Evaluation Framework in K-12 Environments*. Evaluation and Program Planning. May 2012, Vol. 35, Is. 2: 222-235.

Hall, Jori N., Jeehae Ahn, and Jennifer C. Greene. *Values Engagement in Evaluation: Ideas, Illustrations, and Implications*. American Journal of Evaluation. June 2012: 195-207.

Helitzer, Deborah, Christine Hollis, Brisa Urquieta de Hernandez, Margaret Sanders, Suzanne Roybal, Ian Van Deusen. *Evaluation for Community-based Programs: The Integration of Logic Models and Factor Analysis*. Evaluation and Program Planning. August 2010, Vol. 33, Is. 3: 223-233.

Kundin, Delia M. *A Conceptual Framework for How Evaluators Make Everyday Practice Decisions*. American Journal of Evaluation. September 2010: 347-362.

Newton, Xiaoxia A., Rebecca C. Poon, Nicole L. Nunes, Elisa M. Stone. *Research on Teacher Education Programs: Logic Model Approach*. Evaluation and Program Planning. February 2013, Vol. 36, Is. 1: 88-96.

O'Sullivan, Rita G. *Collaborative Evaluation within a Framework of Stakeholder-Oriented Evaluation Approaches*. Evaluation and Program Planning. November 2012, Vol. 35, Is. 4: 518-522.

Penuel, William R., and Barbara Means. *Using Large-Scale Databases in Evaluation: Advances, Opportunities, and Challenges*. American Journal of Evaluation. March 2011: 118-133.

Samuels, Maurice, and Katherine Ryan. *Grounding Evaluations in Culture*. American Journal of Evaluation. June 2011: 183-198.

Scheirer, Mary Ann, Melvin M. Mark, Ariana Brooks, George F. Grob, Thomas J. Chapel, Mary Geisz, Molly McKaughan, and Laura Leviton. *Planning Evaluation Through the Program Life Cycle / Expanding Evaluative Thinking: Evaluation Through the Program Life Cycle / Program Life Cycle Stage as a Guide to Evaluation Decision Making: Benefits, Limits, Alternatives, and Future Directions / Commentary on "Expanding Evaluative Thinking: Evaluation Through the Program Life Cycle" / A Lifer's Perspective on Life Cycle Evaluations / Evaluation Purpose and Use: The Core of the CDC Program Evaluation Framework / Lessons on the Program/Evaluation Life Cycle*. American Journal of Evaluation. June 2012: 263-294.

Wasserman, Deborah L. *Using a Systems Orientation and Foundational Theory to Enhance Theory-driven Human Service Program Evaluations*. Evaluation and Program Planning. May 2010, Vol. 33, Is. 2: 67-80.

Whitehall, Anna K., Laura G. Hill, and Christian R. Koehler. *A Comparison of Participant and Practitioner Beliefs About Evaluation*. American Journal of Evaluation. June 2012: 208-220.