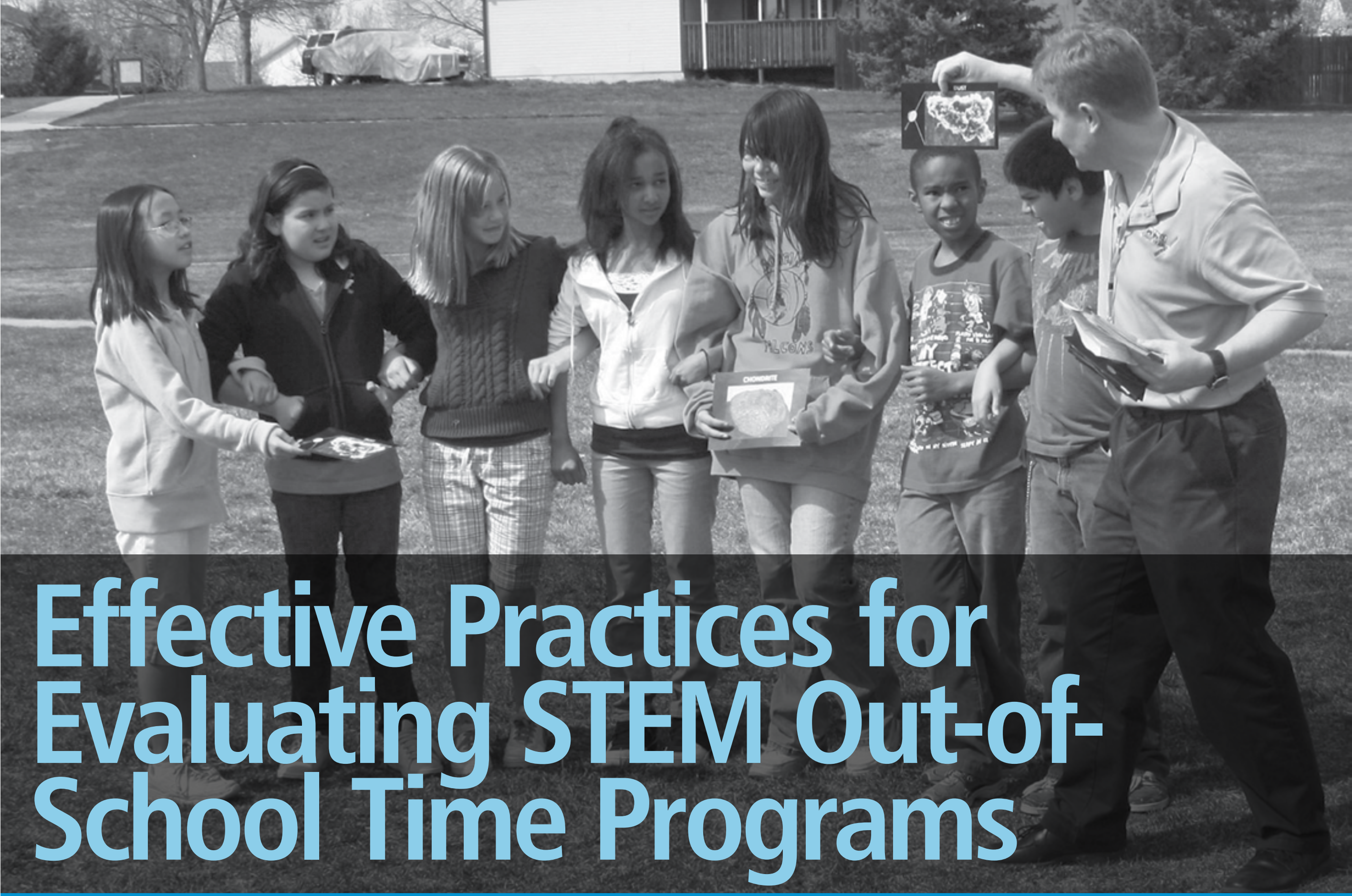


Effective Practices for Evaluating STEM Out-of-School Time Programs

by Stephanie B. Wilkerson and Carol M. Haden

Science, technology, engineering, and mathematics (STEM) programs in out-of-school time (OST) are designed to supplement school work, ignite student interest, and extend STEM learning. From interactive museum exhibits to summer-long science camps, opportunities for informal student engagement in STEM learning abound.

What difference do these programs make, and how can we improve them? These questions preoccupy educators and funders alike. OST program developers and providers can benefit from understanding why evaluation is critical to the success of STEM OST programs, what data collection methods are appropriate, and how to effectively communicate and report findings. In this article, we share lessons from our experience in each of these areas and provide examples of how effective practices play out.

Why Evaluate?

OST programs operate under funding constraints, with tight budgets and ever-increasing calls for accountability. In the past, the results of evaluations commissioned to satisfy the accountability requirements of funding agencies or supervisory organizations often went unread and unused. Now, program staff increasingly understand the value of incorporating evaluation into program design, from inception through delivery. Including evaluation in program planning in early stages allows for adaption and improvements along the way (Frechtling, 2010). As described below, "utilization-focused" evaluations (Patton, 2008) provide planners with valuable information

STEPHANIE B. WILKERSON, Ph.D., is president of Magnolia Consulting (www.magnoliaconsulting.org). For 15 years, she has led evaluations of STEM education programs funded by NASA, the U.S. Department of Education, National Geographic Learning, the Virginia Museum of Natural History, and the National Science Foundation. She specializes in integrating formative and summative evaluation methods into program development. She co-authored a meta-analysis of OST program effects on at-risk students in the *Review of Educational Research*.

CAROL M. HADEN, Ed.D., is senior evaluator at Magnolia Consulting. She has extensive experience evaluating formal and informal STEM education programs and has conducted evaluations of programs funded by the National Science Foundation, the William and Flora Hewlett Foundation, Goddard Space Flight Center, and the Jet Propulsion Laboratory, among others. She has a special interest in supporting exemplary and equitable science education for traditionally underserved populations.

to guide program development: Formative evaluations can inform program improvements, while summative evaluations indicate whether programs are meeting their intended outcomes.

Define Activities and Expected Outcomes

In our experience evaluating STEM programs, we have collaborated with scientists, engineers, program developers, educators, and public outreach providers who bring unique knowledge, talents, and perspectives to the design and delivery of OST programs. Invariably, these individuals are united in their vision: They want to share the excitement of scientific discovery with the people, young and old, who participate in their programs. Using evaluation tools early in program planning enables them to transform that vision into clearly articulated and attainable outcomes for target audiences.

In the development phase, evaluators work with program planners to develop SMART goals: outcomes that are specific, measurable, attainable, realistic, and timely. Bodilly and Beckett's (2005) meta-analysis of OST programs found that programs with clearly defined goals and outcomes had greater success than those whose goals and outcomes were poorly articulated. Success also depends on aligning program planning and activities with goals and outcomes (Huang et al., 2009). This coherence provides a clear line of sight from program purpose to actualization.

In our experience, common short-term outcomes include increasing participants' awareness of and interest in STEM and STEM careers, knowledge of STEM concepts, and program-related skills. Common intermediate outcomes include improving participants' STEM self-efficacy and their application of their new or enhanced knowledge and skills, as shown in such behaviors as continued program participation, enrollment in STEM courses, and choice of STEM majors. Long-term outcomes often include increasing academic learning and achievement in STEM content areas and, ultimately, encouraging STEM career choices. These outcomes reflect the priorities of STEM funding agencies such as NASA (National Aeronautics and Space Administration, 2011) and the National Science Foundation (2011). With well-articulated outcomes, evaluators can develop an evaluation plan and data collection methods that align with these outcomes and corresponding program activities.

During program planning, logic models provide a road map of intended program outcomes so that activities are coherent, focused, and aligned. A logic model depicts a program's theory of change through:

- Inputs: funding, facilities, and resources
- Activities: what and when
- Outputs: numbers of participants, sessions, events, and materials developed
- Outcomes: short-term, intermediate, and long-term effects on target audiences (W. K Kellogg Foundation, 2004)

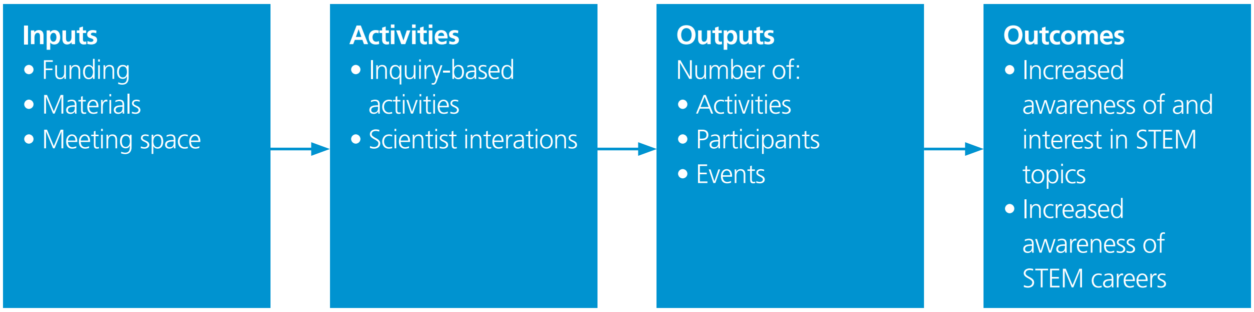
Figure 1 shows a simplified logic model based on NASA Goddard Space Flight Center's Big Explosions and Strong Gravity (BESG) program, a one-day event that engages Girl Scouts in activities with astronomers in the Washington, DC, area. The BESG's theory of change posits that, *if* Girl Scouts engage with scientists in inquiry-based activities and conversation, *then* they will increase their awareness of and interest in STEM topics and careers.

As they develop the logic model, OST program developers must clarify processes for program development and implementation and make cause-and-effect connections about how the program moves from activities to outputs and outcomes. Once the theory of change is laid out, evaluators can decide on the best design and methods to answer questions about program delivery and outcomes (Chen, 1990; Rossi, Lipsey, & Freeman, 2003; Weiss, 1995).

Promote Continuous Learning and Reflection on Practice

Once programs are underway, evaluation creates a feedback loop that guides program decisions and improvements, thereby engaging STEM OST program developers and pro-

Figure 1. Logic Model for the BESG Program



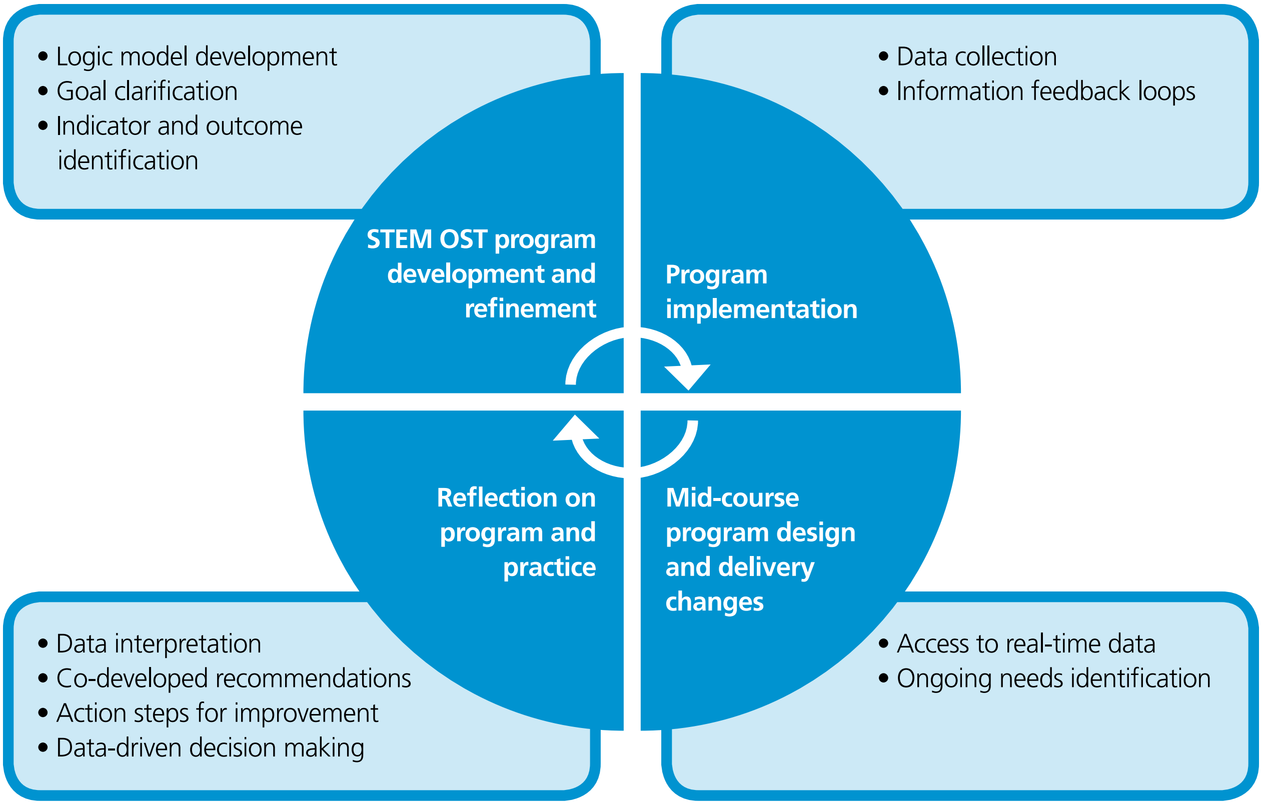

• Logic model development
• Goal clarification
• Indicator and outcome identification
• Data collection
• Information feedback loops
STEM OST program development and refinement
Program implementation
Reflection on program and practice
Mid-course program design and delivery changes
• Data interpretation
• Co-developed recommendations
• Action steps for improvement
• Data-driven decision making
• Access to real-time data
• Ongoing needs identification

Figure 2. Evaluation as a Continuous Learning Process

viders in a continuous learning process, illustrated in Figure 2. At this stage, evaluators have developed or selected data collection instruments, such as surveys, interview and observation protocols, and assessment tools, that align with intended program outcomes. Data collection is ongoing, with formative data providing timely information to inform program modifications (Gray, 1993; Reisner, 2005). Real-time data provide information on program implementation "from the trenches," tapping the perspectives of those who deliver and participate in the STEM program.

For example, Mid-continent Research for Education and Learning (McREL) collaborated with a team of evaluators to develop the two-week Cosmic Chemistry summer program, which aims to improve interest and achievement in chemistry among rising ninth- and tenth-grade students. During two summers, facilitators implemented hands-on activities and interactions with scientists focused on the engaging context of NASA's Genesis mission. To understand how Cosmic Chemistry was implemented and how well its lessons reflected the intended OST best practices, we observed the program in action during both weeks of implementation each summer. Our observations, together with daily facilitator logs, gave evaluators and program developers real-time data, which suggested mid-course modifications to help facilitators implement the program as intended. For example, based on facilitator feedback from the first summer, the developers revised the facilitator's guide to include tips on differentiating instruction and on sense-making activities. The changes were implemented and evaluated during the second summer.

Provide Evidence of Impact and Recommend Improvements

During the last stage of the continuous learning process shown in Figure 2, summative evaluation findings provide information on how well the STEM OST program has achieved its objectives; the findings also document any unintended outcomes. Evaluators analyze data, interpret findings, and work with program planners to develop actionable recommendations for program improvement. Because program developers and providers sometimes bring specialized STEM content knowledge to OST programs, they should be involved in interpreting evaluation findings so that recommendations are relevant, feasible, and specific enough to guide improvement. Evaluation becomes a critical reflective tool for informing the next cycle of program delivery. Summative evaluations can provide evidence of effectiveness to justify continued funding or support proposals for new funding.

Effective Practices for Designing STEM OST Evaluations

Program developers, providers, and evaluators must consider several factors that influence which evaluation designs and data collection methods will be most appropriate for particular STEM OST programs. Effective evaluation practices take into account a program's intended outcomes, phase of development, duration, and budget. These considerations are relevant whether the program is small or large, with evaluation methods being scaled accordingly.

Align Evaluations with Intended Outcomes

As previously described, a logic model is a tool that helps program providers clearly define intended outcomes representing a program's theory of change. It articulates the changes that should result if program providers implement the program as intended. Evaluators use this causal chain ("If we do *x*, then *y* will result") to design evaluations that will support program providers in showing that the program is the cause of any outcomes achieved. Evaluators use logic models to develop evaluation questions that align with a program's intended implementation process and with its short-term, intermediate, and long-term outcomes. Taking into account a program's phase of development and duration, the evaluator frames evaluation questions so they are feasible to answer. The evaluation questions then drive the data collection methods and analytical approach.

STEM OST programs often have long-term outcomes that cannot realistically be measured during the evaluation period. Sometimes they anticipate outcomes that cannot be attributed solely to the OST program. Student outcomes associated with the school day provide a good example. Based on a research review of OST programs, a panel of experts funded by the U.S. Department of Education recommended that OST programs should address content and skills that align with school-day instruction (Beckett et al., 2009). Research suggests that students have a greater potential for experiencing significant learning outcomes and achievement when OST programs connect to school goals (Beckett, 2008; Cooper, Charlton, Valentine, & Muhlenbruck, 2000; McLaughlin & Phillips, 2008).

In our experience, STEM OST program developers align much of their content with what students are expected to know and be able to do as part of their school learning. For example, focusing on short-term outcomes such as students' STEM interest and attitudes is expected to motivate students to enroll in more STEM courses, explore science careers with guidance counselors, and engage in additional learning opportunities. By aligning content with standards, such as the Common Core State Standards for mathematics or the Next Generation Science Standards, OST programs intend for students to apply their learning to coursework during the school day in order to enhance academic achievement, a long-term outcome. When feasible and appropriate, evaluation can serve an important role in measuring the extent to which short-term student outcomes from STEM OST programs transfer to the school day.

Consider a Program's Phase of Development

STEM OST programs that are just beginning will have different evaluation needs than will well-established programs. An effective evaluation design supports a program's growth through various phases from development to refinement to completion (Rossi et al., 2003). Programs cannot be expected to attain longer-term outcomes during development or early implementation.

Before a STEM OST program is even implemented, a variety of evaluation practices can help with program development. During the development phase, evaluation questions ask, "What do you want to do, with whom, and to what end?" Logic models provide a road map to help ensure that activities are coherent and align with program goals (Chen, 1990; McLaughlin & Jordon, 2005). While program materials are in development, program staff might use evaluation methods such as focus groups and interviews to get immediate feedback from target users. This "proof of concept" activity allows developers to make design changes before a program is rolled out. The development phase is also an appropriate time to conduct an informal or formal needs assessment to ensure that program activities will meet the needs of those who stand to benefit (Davidson, 2005). Once a full version of the program is developed, evaluators can facilitate expert review or quality assurance processes by establishing review criteria and feedback forms. These processes help developers to ensure that STEM program content is accurate and consistent with current thinking and practice.

Before a STEM OST program is even implemented, a variety of evaluation practices can help with program development. During the development phase, evaluation questions ask, "What do you want to do, with whom, and to what end?"

Effective evaluation practices for relatively new STEM OST programs involve conducting a pilot study that measures program implementation, creates information feedback loops to inform ongoing revisions, and assesses initial participant reactions and short-term outcomes. Evaluation questions during the implementation phase include "How are providers implementing the program? What additional support do they need? How do participants perceive the quality and utility of the program? What could be changed to better align the program with the intended outcomes?" At the beginning, evaluators and providers focus on building capacity to deliver the program. Data collection methods such as training feedback forms and observations provide information on the consistency of training delivery across multiple sites; whether the training was delivered as intended; and attendees' perceptions of the quality and utility of the training, their level of preparation to implement what they learned, and their recommendations for improvement (Carroll et al., 2007).

From this point, evaluations move into measuring how providers implement STEM OST programs using such data collection methods as online implementation logs, surveys, observations, focus groups, and interviews. These methods can provide program developers with continuous descriptive feedback on variations in implementation, barriers and supports to implementation, implementation fidelity, additional training needs, and perceptions of effects on students (Century, Rudnick, & Freeman, 2010). Student interviews, focus groups, and surveys can provide formative information on how students are responding to the program, how it is affecting them, and what they think would make the program better.

After pilot studies, programs are often revised before scaling up for wider implementation or undergoing another round of small-scale implementation, sometimes referred to as field testing. At this point, the emphasis shifts from measuring implementation to measuring intended outcomes. Evaluation focuses on collecting baseline and post-participation data related to short-term, intermediate, and long-term student outcomes. Implementation measures assess whether STEM OST programs are implemented with fidelity and whether students receive the intended dosage.

Once a program shows promising evidence of student outcomes and has been finalized, it is ready for more rigorous evaluation designs that measure differences in outcomes between students who participate in the STEM OST program and those who participate in a comparison program or receive no intervention at all. Evaluation questions in this phase ask, "Did the program meet its goals? To what degree, and for which participants?" In assessing OST outcomes, particularly academic outcomes, measures must focus on both specific and more general components (Geiger & Britsch, 2003). For example, the evaluation of the Cosmic Chemistry summer program during feasibility testing included an assessment of student understanding of the specific standards addressed in the program. For an outcome evaluation of Cosmic Chemistry, we would use both an assessment of standards aligned with the program and a more general measure of chemistry achievement to understand the program's broader effects on participant learning.

Select Evaluation Methods Appropriate for the Program's Duration

STEM OST developers and providers should clearly define outcomes that are feasible and appropriate given a program's scope and expected reach. In many respects, these expectations relate directly to the amount of time intended audiences spend in the program. For example, the BESG single-day event for Girl Scouts aims to affect student awareness of and interest in science and science careers, whereas the two-week Cosmic Chemistry program is designed to affect student science interest and academic learning. More intensive programs, such as a yearlong afterschool program, might be designed to affect students' science understanding and ultimately their achievement on a state science test.

Figure 3 illustrates the relationship between program duration and common STEM OST program outcomes. As program duration increases, so does the likelihood that the program can achieve longer-term outcomes. Research on summer school programs shows that programs lasting 60–120 hours are more effective at achieving academic outcomes than programs lasting less than 60 hours (Cooper, et al., 2000). A meta-analysis of OST math and reading programs found positive effects on outcomes for programs ranging from 44 to 210 instructional hours (Lauer et al., 2003). Obviously, a program that exposes students to STEM content for 44 hours or more does not alone increase student achievement unless it also provides high-quality, engaging, and developmentally appropriate instruction. However, when deciding which outcomes can reasonably be expected and measured, evaluators should consider program duration.

Effective evaluation practices include selecting appropriate data collection methods for the program's duration and intended outcomes. The following examples from our own experience illustrate how effective evaluation practices can be applied to STEM OST programs of various durations. We find that, irrespective of duration, program developers and providers want both formative feedback to guide improve-

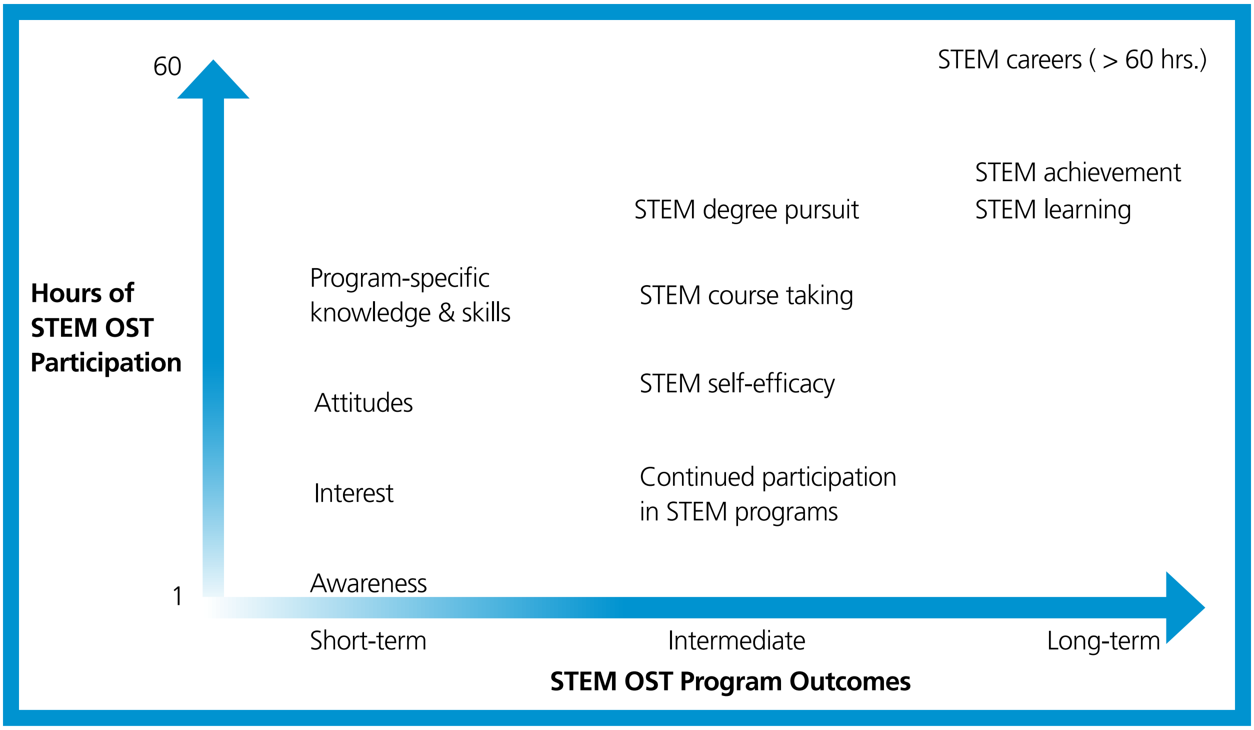

60
1
Hours of STEM OST Participation
STEM careers (> 60 hrs.)
STEM achievement
STEM learning
STEM degree pursuit
Program-specific knowledge & skills
STEM course taking
STEM self-efficacy
Attitudes
Continued participation in STEM programs
Interest
Awareness
Short-term
Intermediate
Long-term
STEM OST Program Outcomes

Figure 3. Relationship Between STEM OST Program Duration and Program Outcomes

ments and summative feedback on outcomes. Accordingly, we tailor evaluation designs and data collection methods to yield both types of feedback and take into account how program duration influences the nature of that feedback.

Short Duration and Single-Day Events

In our experience, most short-duration STEM OST events focus on increasing participant awareness of and interest in STEM-related content or careers. Involving participants in data collection activities can be challenging because of the limited time. Data collection tools must be easily accessible and brief. Depending on the purposes of the evaluation, the methods might include short event surveys or postcards, participant exit polls, or event observations.

One short-duration event we evaluated is the Family Science Night (FSN) series at the Smithsonian's National Air and Space Museum, coordinated and presented by the Universities Space Research Association. FSN invites students and their families to attend evening events lasting a few hours that feature talks by scientists and engineers, an IMAX movie on space exploration, and an after-hours tour of the museum. FSN's intended outcomes include increasing participant interest in space science and raising awareness of space science topics, the work of NASA scientists, and NASA careers. With a limited budget, our evaluation included short, paper-based surveys for students and adults. The surveys allowed us to collect participants' demographic data, their perceptions of the quality of the event, its effect on their interest and learning, their interest in related follow-up activities, and, for adults only, their reason for attending the event. Because the events were promoted through and supported by schools, we conducted follow-up telephone interviews with school liaisons to understand how FSN was integrated into school activities or curricula and to learn how the liaisons perceived the program and its effects on students. Combined, the student surveys, adult surveys, and telephone interviews gave program planners useful formative data for improving the events and relevant summative data on participants' space science awareness and interest outcomes.

The evaluation of the BESG one-day events, whose logic model is depicted in Figure 1, involved brief paper-based student and adult leader surveys, which included items on awareness and interest outcomes, participant demographics, the perceived quality of activities, and suggestions for improvement. Underpinning these efforts was the intention of Goddard Education and Public Outreach (EPO) providers to transition the program away from conducting local events and toward providing materials so groups outside the DC area could conduct their own BESG events with local scientists and resources. As the intent and reach of the program evolved, the evaluation evolved with it to encompass new questions addressing how well new BESG facilitators could plan for and conduct their own events. To understand to what extent BESG was portable, we created one facilitator survey to measure the effectiveness of the training and another on event planning and implementation. We conducted telephone interviews with scientists, educators, and Girl Scout liaisons to understand how well the materials provided by the Goddard EPO team helped them conduct successful events. Over the course of two years, the evaluation provided useful information to BESG planners, who modified the schedule and activities based on evaluation findings. The continuous learning process and a final report enabled program planners to compare findings from early events to those from later events, which had been modified in response to the earlier findings.

Longer Programs

STEM OST programs that engage students for longer periods of time, such as afterschool, Saturday, or summer programs, hold greater potential for affecting intermediate and long-term outcomes than do short-duration programs (Cooper et al., 2000). The intensity or frequency of delivery among longer-duration programs can vary: Afterschool delivery is distributed over weeks or months during a school year, while summer programs are condensed into a few consecutive weeks. Compared to evaluations of short-duration programs, evaluations for longer programs can employ more rigorous designs with a greater variety of data collection methods. These methods might include longitudinal student surveys, implementation logs, student journals, case study interviews and observations, and student achievement measures.

Compared to evaluations of short-duration programs, evaluations for longer programs can employ more rigorous designs with a greater variety of data collection methods. These methods might include longitudinal student surveys, implementation logs, student journals, case study interviews and observations, and student achievement measures.

As part of our ongoing work with Goddard EPO, we conducted an evaluation of the A.C.E. (Astronomical Cosmic Exploration) of Space afterschool club for Girl Scouts. A.C.E. of Space engages girls in hands-on learning opportunities, "girl-given" group presentations, "girl-driven" activities, meetings with successful female scientists and professionals, and tours of NASA facilities. Because the program met once a month for an academic year, we were able to measure changes in girls' interest in space science and STEM careers, their vision of themselves as scientists, and their understanding of STEM topics. Girls completed a pre- and post-participation interest survey containing 23 items—some ranking statements on a Likert scale and some open-ended—to measure intended program outcomes and participant perceptions. Additionally, girls kept journals on their club activities and responded to reflection questions each month on what they had learned, what they found exciting about the month's event, how A.C.E of Space activities related to their own lives, and how interested they were in space science and space science careers. The surveys and journal reflections allowed us to examine gains in space science interest and skills over an extended period of time. With a modest budget, the evaluation provided abundant formative feedback to improve program design and delivery throughout implementation, as well as summative feedback on measurable outcomes.

A summer program like Cosmic Chemistry also allows for study of longer-term outcomes, in this case students' understanding of chemistry and their motivation to study science. Evaluation team members at McREL and Magnolia Consulting assessed Cosmic Chemistry students with a pre- and post-participation chemistry assessment aligned with the standards covered by the program. We also administered a survey of motivation and perceived competence before and after the program, and then again during the following school year, to examine effects on student interest, motivation, and self-efficacy in science and chemistry. In addition to assessing specific chemistry content objectives, we also administered daily facilitator implementation logs and conducted daily observations to measure implementation of best OST practices, including setting high expectations, motivating students, and building background knowledge. The condensed program delivery—60 hours over a two-week period—allowed us to increase the intensity of our data collection. Had the program been delivered in non-consecutive sessions, the cost of traveling to sites to conduct the same number of observations would have been prohibitive. Findings from the pilot study provided formative data to the development team for program modification, while findings from the subsequent field test during the second summer provided summative information on program effects.

Provide the Most Rigorous Designs Possible Under the Allocated Budget

Taking into account stakeholder information priorities, intended outcomes, phase of program development, and program duration, evaluators develop evaluation designs that give STEM OST program providers the most "bang for the buck." This is no easy task, as there are trade-offs between design and budget. Typically, the more rigorous the evaluation study—that is, the more the evaluation design allows providers to make *causal* claims about program effectiveness—the more expensive it is. Done right, providing this level of rigor usually involves costly randomized control trials or quasi-experimental designs that include a control group to measure whether differences between treatment and control group outcomes can be attributed to the program. This type of design, with its corresponding budget, is most appropriate for well-established STEM OST programs of long duration that have already used evaluation for planning, feedback, and improvement (Rossi et al., 2003).

More often than not, evaluation budgets for STEM OST programs are meager at best, yet the programs come with the same information needs and priorities as programs with larger evaluation budgets. So how do program providers get the information they need, given their limited funds?

More often than not, evaluation budgets for STEM OST programs are meager at best, yet the programs come with the same information needs and priorities as programs with larger evaluation budgets. So how do program providers get the information they need, given their limited funds? Using the following recommendations, STEM OST program providers can become better-informed consumers, working with evaluators to maximize evaluation offerings and minimize costs.

- Prioritize which program outcomes are most appropriate and important to evaluate based on the phase of program development and funder information needs (Stecher & Davis, 1987).
- Create a long-term evaluation plan that identifies how program outcomes will be measured over time, rather than all at once. Use logic models to justify prioritizing short-term outcomes over intermediate or long-term outcomes (Reisner, 2005).
- Use data collection methods, such as online surveys and social media, that are less expensive to implement than site interviews, focus groups, and observations. Instead of site visits, conduct phone interviews or focus groups to collect in-depth formative feedback about user perceptions.
- Keep survey instruments brief. The longer the survey, the more time is required for data analysis and reporting, thus increasing the budget.
- Learn from evaluations of similar programs (Geiger & Britsch, 2003). Identify existing instruments that align closely with program outcomes, such as those provided through the Harvard Family Research Project OST Program Research and Evaluation Database (Wimer, Bouffard, & Little, 2008).
- Collect data from small samples of participants during early phases of program development, and then expand to include larger numbers as the program matures.
- Use informal data reports to give developers access to pertinent, timely data for program improvement without having to expend resources on formal implementation reporting.

Effective Practices for Communicating Results

A utilization-focused approach to evaluation emphasizes how stakeholders will use the findings (Patton, 2008). Program developers and providers, participants, and funders might each have different needs for information about the STEM OST program being evaluated; effective evaluation reporting should address these needs (Torres, Preskill, & Piontek, 2005). As with curriculum development, evaluators often use a sort of backward-mapping technique that begins with the end in mind, determining how evaluation findings will be used, for what purposes, and by whom.

Comprehensive evaluation reports can address the needs of many stakeholders. Reader-friendly reports include an executive summary; provide visual representations of data, such as charts, graphs, and summary tables; omit technical jargon; are well-organized and concisely written; include recommendations for improvement; and append supporting and detailed technical information (Torres et al., 2005). However, evaluators can also provide more tailored information based on specific stakeholders' intended use of the results.

Program developers are interested in recommendations for improvement and data that will drive decision making. They also want to know if they have achieved the outcomes they set out to accomplish. Data reports generated from online surveys and informal debriefs (in person or by phone) can provide real-time feedback to guide mid-course decisions during implementation of STEM OST programs. This information not only provides timely formative feedback, but also can function as a tool for monitoring student progress toward intended outcomes. One way to increase the likelihood that program developers will use evaluation results is to engage them in interpreting findings and co-developing recommendations or responding to evaluators' recommendations (Cousins, 2003; Patton, 2009). Engaging program developers in the reporting process will help them identify action steps in response to recommendations. Verbal presentations of study results allow for meaningful dialogue about data interpretation, recommendations, and program improvements.

STEM OST practitioners, the ones who deliver the programs, seek how-to information and methods for ensuring successful implementation. They want reports that emphasize lessons learned and implications for future practice. Additionally, reports that capture the experiences, perceptions, and voices of participants can tell a compelling story about the importance of effective practices. For example, vignettes or descriptive narratives based on qualitative data can be an effective way to help facilitators to "see" important nuances in implementation and instructional pedagogy.

Funders want to know if their investment results in expected outcomes. Future funders seek evidence of effectiveness or promising practices that are worth funding. Various presentations of evaluation findings can help connect funders to the people who benefit from their investment. A concise description of evaluation findings, such as an executive summary or oral presentation, can be an effective way of highlighting program effects and outcomes. Videos of participants sharing how their STEM OST experience affected them are also compelling. Younger participants might show how a STEM OST experience affected them by drawing, for example, what they understand about plant life cycles or how they feel about science.

Effective evaluations meet the needs of STEM OST program stakeholders. They take into account a program's intended outcomes and purpose, phase of development, duration, information priorities, and budget limitations. The more funders and consumers of STEM OST evaluations understand effective evaluation practices, the more relevant, timely, and useful the evaluation results will be in helping programs to achieve their goals. Evaluations designed with these considerations in mind ensure that programs operate in an information-rich environment, to the benefit of all who participate.

References

Beckett, M. K. (2008). *Current-generation youth programs: What works, what doesn't, and at what cost?* Santa Monica, CA: RAND Corporation.

Beckett, M., Borman, G., Capizzano, J., Parsley, D., Ross, S., Schirm, A., & Taylor, J. (2009). *Structuring out-of-school time to improve academic achievement* (NCEE 2009–012). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Bodilly, S., & Beckett, M. (2005). *Making out-of-school time matter: Evidence for an action agenda.* Santa Monica, CA: RAND Corporation.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science,* 2(40), 1–9. doi:10.1186/1748-5908-2-40

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Education, 31*(2), 199–218. doi:10.1177/1098214010366173

Chen, H. (1990). *Theory-driven evaluations.* Thousand Oaks, CA: Sage.

Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monograph Series for the Society for Research in Child Development, 65*(1).

Cousins, J. B. (2003). Utilization effects of participatory evaluation. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), *International handbook of educational evaluation* (pp. 245–265). Boston, MA: Kluwer Academic.

Davidson, J. E. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation.* Thousand Oaks, CA: Sage.

Frechtling, J. (2010*). The 2010 user-friendly handbook for project evaluation.* Washington, DC: National Science Foundation.

Geiger, E., & Britsch, B. (2003). *Out-of-school time program evaluation: Tools for action.* Portland, OR: Northwest Regional Educational Laboratory.

Gray, S. T. (Ed.). (1993). *Leadership is: A vision of evaluation.* Washington, DC: Independent Sector.

Huang, D., Cho, J., Mostafavi, S., Nam, H. H., Oh, C., Harven, A., & Leon, S. (2009). *What works? Common practices in high functioning afterschool programs across the nation in math, reading, science, arts, technology, and homework* (CRESST Report 768). Los Angeles:

University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H., Snow, D., & Martin-Glenn, M. L. (2003). *The effectiveness of out-of-school-time strategies in assisting low-achieving students in reading and mathematics: A research synthesis.* Aurora, CO: Mid-continent Research for Education and Learning.

McLaughlin, B., & Phillips, T. L. (2008). *Meaningful linkages between summer programs, schools, and community partners: Conditions and strategies for success.* Baltimore, MD: National Center for Summer Learning.

McLaughlin, J. A., & Jordan, G. B. (2005). Using logic models. In J. Wholey, H. Hatry, & K. E. Newcomer (Eds.) *Handbook of practical evaluation* (2nd ed., pp. 7–32). San Francisco, CA: Jossey-Bass.

National Aeronautics and Space Administration. (2011). *2011 NASA strategic plan.* Retrieved from http://www.nasa.gov/pdf/516579main_NASA2011StrategicPlan.pdf

National Science Foundation. (2011). *Empowering the nation through discovery and innovation: NSF strategic plan for fiscal years 2011– 2016.* Washington, DC: Author.

Patton, M. Q. (2008). *Utilization-focused evaluation*. Newbury Park, CA: Sage.

Patton, M. Q. (2009). *Developmental evaluation.* New York, NY: Guilford Press.

Reisner, E. R. (2005). *Using evaluation methods to promote continuous improvement and accountability in after-school programs: A guide.* Washington, DC: Policy Studies Associates.

Rossi, P. H., Lipsey. M. W., & Freeman, H. E. (2003). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.

Stecher, B. M., & Davis, W. A. (1987). *How to focus an evaluation.* Newbury Park, CA: Sage.

Torres, R. T., Preskill, H., & Piontek, M. E. (2005). *Evaluation strategies for communicating and reporting: Enhancing learning in organizations* (2nd ed.). Thousand Oaks, CA: Sage.

W. K. Kellogg Foundation. (2004). *Using logic models to bring together planning, evaluation, and action: Logic model development guide.* Retrieved from http://www.wkkf.org/knowledge-center/resources/2006/02/wk-kellogg-foundation-logic-model-development-guide.aspx

Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts* (pp. 65–92). Washington, DC: Aspen Institute, Roundtable on Comprehensive Community Initiatives for Children and Families.

Wimer, C., Bouffard, S., & Little, P. (2008). *Measurement tools for evaluating out-of-school programs: An evaluation resource.* Cambridge, MA: Harvard Family Research Project.