The work described in this paper was supported by the U.S. Department of Education through the Teacher Incentive Fund. The opinions expressed are those of the authors and do not necessarily reflect the view of the U.S. Department of Education or the institutional partners. Comments and suggestions are welcome.

In 2012, the U.S. Department of Education awarded the Teacher Incentive Fund STEM Technical Assistance contract to Westat and Horizon Research.

The primary purpose of the Teacher Incentive Fund (TIF) is to support TIF grantees in their implementation efforts through provision of sustained technical assistance and development and dissemination of timely resources.

This work was produced in whole or in part by Westat and Horizon Research with funds from the U.S. Department of Education under contract number GS-23F-8144H/ED-ESE-12-O-0048. The content does not necessarily reflect the position or policy of TIF or the Department of Education, nor does mention or visual representation of trade names, commercial products, or organizations imply endorsement by TIF or the federal government.

Joy Frechtling, Project Director
E-mail: tifstem@westat.com
# Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>1</td>
</tr>
<tr>
<td>Develop a Shared Understanding of the Grant and the STEM Master Teacher Component</td>
<td>4</td>
</tr>
<tr>
<td>Identify and Prioritize Evaluation Questions to Address the STEM Master Teacher Component</td>
<td>8</td>
</tr>
<tr>
<td>Select Data Collection Approaches for the STEM Master Teacher Component</td>
<td>11</td>
</tr>
<tr>
<td>Develop the Evaluation Design and Analysis Plan for Assessing the STEM Master Teacher Component</td>
<td>14</td>
</tr>
<tr>
<td>Disseminate Findings</td>
<td>18</td>
</tr>
<tr>
<td>References</td>
<td>20</td>
</tr>
<tr>
<td>Appendix A: Evaluation Resources</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Logic model</td>
</tr>
</tbody>
</table>
Background

This Evaluation Resource Guide is intended to assist Teacher Incentive Fund (TIF) program grantees in developing their evaluations, focusing on efforts to improve science, technology, engineering, and mathematics (STEM) education. Since the beginning of the TIF program, the U.S. Department of Education (ED) has provided technical assistance (TA) resources to grantees to improve program design and delivery, as well as to design and conduct rigorous, informative evaluation studies. This guide was developed to help grantees think about how they might evaluate the STEM Master Teacher component of their TIF grant. The guide is not meant to be a complete primer on how to do evaluation, but rather a tool for thinking about what you want to know about your STEM Master Teacher component, the critical questions that need to be addressed, and how you might answer them. It builds on and extends other more detailed guides for project evaluation and complements other tools provided through TIF TA contracts. (See the References and Appendix A: Evaluation Resources for some useful guides for evaluation.) While the primary audience is TIF program managers, project implementation teams, and their evaluators, the guide is intended to be useful to others such as funders, program implementers, and local stakeholders. It is one of many services being offered to help grantees develop and support STEM Master Teachers as part of their performance-based compensation systems (PBCSs).

The TIF program is an ambitious effort to improve and to promote high-quality education. It provides opportunities for states, districts, and schools to create new PBCSs that incentivize and reward teachers and principals for attaining improved student achievement. The program goals include the following:

- Improving student achievement by increasing teacher and principal effectiveness;
- Reforming teacher and principal compensation systems so that teachers and principals are rewarded for increases in student achievement;
- Increasing the number of effective teachers teaching poor, minority, and disadvantaged students in hard-to-staff subjects; and
- Creating sustainable performance-based compensation systems.

Initiated in 2006, the program, through its four funding cycles, has made over 120 grants to entities willing to take on this challenge. TIF STEM grantees funded as part of the fourth and latest cohort have a special focus on creating opportunities for STEM teachers to take on new leadership positions and contribute their special talents to supporting improved teaching and learning. Working within the TIF grant’s PBCS, grantees who have
elected to address Absolute Priority 3 of the TIF program (Improving Student Achievement in Science, Technology, Engineering, and Mathematics) are expected to develop a corps of STEM Master Teachers who

1. are rated as effective or highly effective under the local education agency’s (LEA’s) PBCS,
2. are selected based on criteria that are predictive of their ability to lead other teachers,
3. demonstrate effectiveness in one of more STEM subjects, and
4. accept STEM-focused career ladder positions.

Additionally, TIF STEM grantees must identify the unique competencies that characterize effective STEM teachers; use the human capital management system (HCMS) to attract effective teachers to identified hard-to-staff STEM subjects; leverage community support, resources, and expertise; provide financial and nonfinancial resources to attract and retain strong STEM teachers; and work to ensure that all students have access to and participate in rigorous and engaging STEM coursework.

All grantees are expected to conduct evaluations of their programs. Evaluations provide opportunities for collecting data to inform program progress, document program outcomes, and contribute knowledge to the broader education field. While some think of evaluation as an add-on or, even worse, something that is “done to” a grant, those who understand the process know that it can be extremely useful to managers, participants, and other stakeholders, as well as to funders.

The TIF program is an ambitious effort to improve and to promote high-quality education.

It is useful in thinking about evaluation to consider how the collection of data for evaluation purposes may be similar to and different from other uses of data. Frechtling and contributing authors (2010) contrasted four related types of data collection—program description, performance indicators, formative and summative evaluation, and research—in terms of the information obtained and the purposes they are intended to serve.

- **First is project description**, which provides general information. These data are commonly used to describe project activities (e.g., funding levels, total number of participants) and specific project components (e.g., duration of program activity, number of participants enrolled in each activity) and to identify the types of individuals receiving services. Descriptive information may be collected annually or even more frequently to develop a picture of a program’s portfolio.

- **Second are performance indicators**, defined as a collection of statistics used to monitor the ongoing status of a program against a set of targets and metrics. Performance indicators play a critical role in the Government Performance and Results Act (GPRA) and Performance Assessment and Rating Tool (PART) activities. Going beyond project description, performance indicators begin to provide quantitative information that can be measured against a set of goals and objectives.

- **Third are formative and summative evaluations**. These evaluations are intended to gather information to answer a limited number of questions about projects and their outcomes. Evaluations may start with descriptive information and include performance indicators, but typically they go well beyond them. Generally, formative and summative evaluations are intended to assess the extent to which a particular set of interventions, when implemented appropriately, have intended results. They include in-depth data collection activities that may be qualitative as well as quantitative, are intended to support decision making, and range in cost, depending on the questions asked and project complexity.

- **Finally, there are research studies**. Research studies also include descriptive information and provide targeted in-depth exploration of issues, but they differ from evaluations in other ways. Instead of being intended for decision making, research efforts typically are designed to broaden understanding of how a phenomenon or intervention works.
The remainder of the guide discusses five interdependent activities that will help you in planning, conducting, and using your evaluation. It covers the information in the following chapters:

- **Develop a Shared Understanding of the Grant and the STEM Master Teacher Component.** This chapter is intended to assist grantees in refining and clarifying their goals for STEM instructional change, how they are going to reach these goals, and the roles and responsibilities of the STEM Master Teacher.

- **Identify and Prioritize Evaluation Questions to Address the STEM Master Teacher Component.** This chapter discusses identifying what you want to know and how to identify the specific questions to be addressed.

- **Select Data Collection Approaches for the STEM Master Teacher Component.** This chapter talks about different ways of collecting data to answer the questions and discusses some trade-offs to keep in mind.

- **Develop the Evaluation Design and Analysis Plan for the STEM Master Teacher Component.** This chapter discusses study designs and how they might be used at different stages of the TIF STEM grant.

- **Disseminate Findings.** This chapter addresses the importance of sharing evaluation information over time and with different audiences.
Develop a Shared Understanding of the Grant and the STEM Master Teacher Component

It is not unusual in developing a proposal to have some areas addressed in greater detail than others. Once the grant is received, however, it is important to go back and review key aspects of the proposed work to make sure there is a clear understanding of what is expected and to share this understanding with the leadership team and other critical stakeholders. In this chapter, we discuss the steps to establish and solidify this understanding:

- Creating a vision of effective STEM teaching and learning;
- Embedding this vision in a STEM Master Teacher/career ladder model;
- Developing a theory of action to further this vision; and
- Creating a logic model to represent the theory of action graphically.

In the case of TIF STEM grantees, one area requiring further discussion and clarification is likely to be the project’s vision of effective STEM teaching and learning. Clarifying your vision, while important for the evaluation, is probably more important for your project development. For example, does this vision include new pedagogical approaches? Does it include the introduction of engineering concepts? Are STEM subjects to be treated as individual areas of study or is some type of integration seen as desirable? What will high-quality instruction in these courses look like? Is project-based learning to be part of the curriculum? How about experiential learning and practical experiences? What role is envisioned for business or industry partners in the community and are they to be engaged as partners in extending students’ learning experiences? What about STEM education at the primary grades? Given new understandings about the ability of young children ability to grasp scientific concepts at an early age (National Research Council, 2005) what changes might be needed at the elementary level? (A more detailed discussion of creating this vision can be found in the webinar “Building Agreement Around a Vision of Effective STEM Instruction” (https://TIFSTEMcommunity.org).)

Once this vision is clarified, a second step is to examine the implications of this vision for a career ladder for STEM Master Teachers. The team should determine which additional roles and responsibilities are appropriate for the Master Teachers given the expectations for teaching and learning. For example, will they need to take responsibility for the K–12 articulation or selection of new materials? Will they provide professional learning experiences for teachers? Will a step on the career ladder require working across schools, as well as within schools? Will the vision create a position for STEM Master Teachers that involves working with the community, identifying practicum experiences, and monitoring the success of these experiences? How will the HCMS be structured to attract and reward STEM Master Teachers? Is the system built on an assumption that existing staff will fill these roles or is it assumed that new teachers will need to be attracted into the district? What kinds of supports will be needed to prepare the STEM Master Teachers and make them successful? How will the support be tailored to different roles as defined above (providing support for implementing rigorous courses, training other teachers, working with the community)? Beyond student performance, will the evaluation system include any other measures to assess the success of the STEM Master Teacher component? More discussion of career ladders for STEM Master Teachers and utilizing an HCMS to support STEM Master Teachers can be found in the companion briefs Using Career Ladders to Support STEM Master Teachers and
Aligning the Human Capital Management System to Support STEM Master Teachers.

Once the details of the project’s plan and vision are clarified and/or reaffirmed, the next step is pulling them together into a theory of action. A theory of action shows relationships among project components and expected outcomes. It illustrates not only what the separate pieces of the project are, but also how they relate to each other and to ultimate goals. For example, suppose a key aspect of a project’s vision for effective STEM teaching and learning is focusing on gatekeeper courses in middle school, ensuring that instruction is challenging, sensitive to differential student needs, and integrated into real-world problems. The theory of action would need to specify the following:

- **The goals or objectives of the activity**—is it factual or conceptual learning? Career education? Drop-out prevention? Increasing interest in STEM learning? What goals should be considered interim and which ones ultimate? How will you know if you are approaching your desired outcomes, even if it is too soon to assess ultimate goals?

- **The strategies that will be used**—will new curricula need to be identified or developed? Will classroom teachers need professional development in how to conduct project-based learning? What kinds of training and support will be needed by the STEM Master Teachers and who will provide it? Will new partnerships with the community be developed? How will partners be incentivized to donate their time and skills?

- **The roles of STEM Master Teachers**—will STEM Master Teachers lead the development or identification of the curricula and materials? Will they provide professional development and coaching to other teachers? Will they be asked to develop and run model classrooms? Will they be involved in forming relationships with community and industry partners who can help create bridges to the real world?

Being explicit about the project’s theory of action upfront is important, and the initial time allocated to defining the connections between what you want to accomplish and how you plan to get there is usually time well spent. More on developing your theory of action can be found in the webinar “Developing a Theory of Action for STEM Master Teacher Programs” (https://TIFSTEMcommunity.org).

Given the theory of action, the last step in developing a shared understanding of your vision is the development of a logic model. A logic model is a visual depiction of your theory of action, showing the components and the expected connections among them. This logic model helps to document shared understandings and provides an anchor that all team members can relate to as they go about their work. It also provides useful scaffolding for your evaluation, including the development of evaluation questions, identification of measures, and design of your data collection. Although there are a number of different ways to present a logic model, a frequently used format contains the following (Frechtling, 2007; W.K. Kellogg Foundation, 2004):

- **Inputs** are the resources brought to a project, typically funding sources such as TIF funds or intellectual resources such as the experiences of personnel in the current and earlier TIF grants.

- **Activities** are the components of the treatments that a project is providing. They are the actions that will be taken to move toward the project’s goals and outcomes. In the theory of action discussed above, activities would include the development of a career ladder system supported by the HCMS, communication of how this career ladder is intended to implemented and rewarded, training for STEM Master Teachers, professional development for classroom teachers, and development of partnerships with industries in the community.

- **Outputs** are the products of activities. Outputs document the implementation of an activity in simple itemized ways, such as the number of teachers identified as potential STEM Master Teachers, the number of teachers who apply to be part of the STEM Master Teacher corps, and the number who complete training. While an output can in some cases be viewed as an objective of an activity, in most cases we reserve the term objective or outcome for
changes that occur in behaviors. An output, then, is typically considered an indicator that activities are taking place.

- **Outcomes** are results or changes in behaviors or learning. Outcomes reflect the goals and objectives of the project. Outcomes are typically specified in the short and longer term. How to differentiate short- and long-term outcomes is project specific and determined in large part by the complexity of the changes that are sought. In the current example, the acquisition of new skills and knowledge would be a short-term outcome for the STEM Master Teachers, and persistence in STEM courses and improved test scores would be a longer term outcome for the students.

- **Impacts** are broader changes in systems that can be linked to the interventions or activities depicted in the model. Typically such impacts include, but also extend beyond, the specific institutions that have been part of the program being studied. For TIF STEM projects, one broader impact would be the establishment of a critical mass of teachers with high-quality skills in STEM subjects. Additional spin-off impacts might also be expected from the industry partnerships that are established.

- **Contextual factors** are the special characteristics of a grantee setting that need to be considered in understanding how a set of interventions has played out. These factors help clarify the nature of the setting in which the project has occurred and can play a role in determining where findings might and might not be generalizable. Contextual factors of importance might be the overall economic conditions in the surrounding community, the availability of STEM job experiences suitable for high school students, the initial level of preparation of the STEM teaching force, the buy-in of the union, previous STEM programs to enhance STEM teaching and learning, and the academic readiness of high school students in STEM subjects.

Figure 1 (following page) presents a logic model illustrating the theory of action for a TIF STEM grant addressing STEM teaching and learning at high schools serving students from high-risk populations. It lays out three sets of activities: one set relates to the development and definition of a career ladder for STEM Master Teachers, the second set to the identification and preparation of STEM Master Teachers, and the third set to the activities that these STEM Master Teachers undertake as part of their career ladder responsibilities. Also spelled out is a set of short-term or interim outcomes that are expected to occur as a result of the STEM Master Teacher activities. The arrows showing the connections between the activities and the changes that are expected to occur in the short/medium term make explicit how the career ladder program is expected to result in new skills and knowledge for both STEM Master Teachers and STEM classroom teachers. Finally, the last set of arrows linking teacher changes to student changes shows that these teacher changes are expected to result in a range of important student outcomes.

As can be seen, the value of the logic model is its ability to both clearly specify the critical components of one’s intervention (the boxes) and, even more important, the connections among actions and outcomes (the arrows).

A few caveats about this model should also be mentioned.

- To keep it simple, we have not included broader system impacts in this illustration and have focused only the theory of action for the STEM Master Teacher component of the TIF STEM grant. While a separate depiction of this component is useful, it must be kept in mind that it is only one part of the overall TIF grant and should be considered in context of the other critical project components.

- The model appears to portray a one-way progression, starting with inputs and ending in outcomes. Few programs have such uninterrupted progressions. A more realistic (and considerably more complex) depiction of the model would include a number of feedback loops where, for example, information from short-term outcomes feed into a modification of the activities. A logic model should be considered a dynamic document, one that can be expected to change as part of the program implementation process.
Figure 1. Logic model

**Inputs**
- TIF funding
- Local funding
- Experience of leaders with STEM teaching and learners

**Activities**
- Project develops its careers ladder and defines the steps on this ladder
- Career ladder program explained internally and externally
- Project identifies, trains, and supports STEM Master Teachers for the career ladder roles

**Outputs**
- STEM Master Teachers carry out their roles, which may include:
  - Curriculum development
  - Professional development
  - Modeling of lessons
  - Observation and feedback
  - Liaison with industry and the community

**Outcomes**
- Changes in STEM Master Teachers’ ability to:
  - Lead and facilitate the selection of high-quality instructional materials and resources
  - Lead and provide professional development addressing content and pedagogy
  - Work individually with STEM teachers to support their growth
  - Create industry liaisons and monitor practicum experiences

**Context**
- Previous programs to improve STEM teaching and learning, student demographics, local economy, related state or district mandates

**Short/medium**
- Changes in teachers’ instructional practices in STEM courses: Increased use of:
  - Inquiry techniques, student-led discussion
  - Group work
  - Performance assessments

**Long**
- Increased retention in STEM courses
- Improved attendance in STEM courses
- Higher student achievement in STEM subjects
Identify and Prioritize Evaluation Questions to Address the STEM Master Teacher Component

Once you have clarified your vision for the work of STEM Master Teachers and translated your theory of action into a logic model, you are ready to begin identifying and prioritizing the questions your evaluation is designed to address. Evaluation questions are generally divided into two categories: formative and summative (Witham et al., 2011). Formative questions address issues related to program implementation and generally ask whether or not a program is being implemented as planned, or if modified, the rationale for making those modifications. Summative questions address the extent to which goals are being achieved, looking at both progress toward goals and the end state. Evaluations frequently also include questions regarding context and the role that context plays in facilitating or hindering progress and the attainment of goals.

In this chapter, we discuss four steps for developing evaluation questions:

- Using the logic model;
- Gaining input from your stakeholders;
- Developing priorities; and
- Integrating evaluation of the STEM Master Teacher component into the overall program evaluation.

The logic model developed in the previous chapter can be very helpful in identifying the possible set of questions that your evaluation might address. Roughly speaking, the content of the activities boxes can be used to develop formative questions (questions about the implementation of your project), and the content of the outcomes boxes can provide the material for your summative questions. Using the example logic model, we can identify three major formative questions and three major summative questions. The formative questions are as follows:

Q1. Was a career ladder for STEM Master Teachers developed? How were the needs/vision of the career ladder incorporated into the HCMS? Were the positions on the ladder and the responsibilities associated with them defined and communicated? What was the relationship between the roles and responsibilities identified and the grant's vision for STEM instruction?

Q2. What activities were undertaken by the project to identify, train, and support STEM Master Teachers? What barriers were encountered?

Q3. To what extent did the STEM Master Teachers effectively carry out these new roles and responsibilities defined in the career ladder?

Each of these questions could and probably should be further broken down to gain an understanding of not only what happened but why. For example, to address Q2 above, some important additional questions might be the following:

- Did teachers have a sufficient understanding of what it means to be a Master Teacher in the project? Did they understand the career ladder structure?
- Did the incentives and rewards appear to be sufficient? If not, why not?
- Was the training provided sufficient and appropriately tailored to their individual needs and to the defined roles for STEM Master Teachers? Were there some areas that should have been addressed that were not?
Similarly the related summative questions might be these:

**Q1.** To what extent did STEM Master Teachers increase their skills and knowledge?

**Q2.** To what extent did changes occur in classroom teachers’ instructional practices?

**Q3.** To what extent did changes occur in students’ behavior and performance?

Subquestions for Q1 include the following:

- What was the relationship between changes in Master Teachers’ content knowledge and pedagogical content knowledge and the vision of STEM teaching and learning developed by the project? Were some areas harder to address than others? If so, why?

- What kinds of experiences were most effective in helping STEM Master Teachers increase their skills and knowledge? Was a good balance achieved between supports received for leadership and supports received for enhancing their disciplinary and/or content knowledge?

- Was there any consistent relationship between differences in Master Teachers’ backgrounds and characteristics and their success in their roles?

- How well did the teacher evaluation system used to assess the Master Teachers inform and differentiate among participants? Did it focus sufficiently on skills and behaviors needed to be successful in the STEM areas?

Contextual questions examine how factors outside of the program affect program implementation and outcomes. They address both past and present variables that might be expected to influence the development and outcomes of the current project, as well as the generalizability of results to other situations. Examples are as follows:

- To what extent have other programs addressing Master Teachers been undertaken in the district? Have there been earlier attempts to create career ladders and what were their outcomes? Did these previous efforts appear to have impact (positive or negative) on the success of the TIF STEM Master Teacher program?

- What has previously been the vision of effective STEM instruction and how has this vision changed over time?

- What have been the results of any previous attempts to establish relationships with STEM-related industries in the community? Does fertile soil exist for expanding partnerships or are there wounds that need to be healed? What mechanisms exist to expand the partnerships and what still needs to be done in order to broaden partnership networks?

- What resources for STEM professional development are available in the community and what is the district’s relationship with them? What kinds of partnerships have been developed with community colleges and 4-year postsecondary institutions? Are there existing formal agreements with postsecondary institutions both to support students (i.e., dual enrollment) and teachers (i.e., professional development, teacher preparation, etc.)?

- What related state or district initiatives exist and how are they being taken into consideration?

As shown above, the list of potential questions can easily become quite long. While at first this long list may seem intimidating, it is better to be inclusive at the beginning and then narrow down your list based on further considerations.
A first step in prioritizing the questions is to gather input from critical stakeholders. This is not an easy task and will require careful planning, time to listen and record inputs, and a process for providing feedback on how suggestions were addressed. It is helpful to convene a group of your key stakeholders (including ones who have been involved in the logic model development and ones who have not) to solicit their input on the evaluation questions.\(^1\) Potential invitees include school board members, union representatives, parents, and industry representatives engaged in STEM-related activities, as well as key project members and school district administrators. Such a meeting is likely to broaden the number of questions that potentially demand attention, but hearing the voices of such a diverse stakeholder group is likely both to provide useful information to consider, including what stakeholders view as evidence, and to gather broad buy-in to your work. It is also possible that the questions that arise may cause you to make some modifications in your theory of action and plan. Identifying any such areas of interest early on may prevent discord as the project evolves.

Once you have your list of potential questions, the next step is to determine their relative importance. First, what are the grant requirements? Are there certain data that your grant requires you to submit? Second, outside of grant requirements, what are the questions on the “must-have” lists of your most important stakeholder groups? Third, are the questions answerable within the timeframe of your grant? Fourth, is there a research design that will allow you to answer the questions? Fifth, do extant data exist to address the questions? And relatedly, are you asking the appropriate set of questions for the stage of development of your career ladder program? Recognizing that this program will continue to evolve, what is it that you need to know to place yourself in the best possible position to take next steps?

As you think about your questions, it is important to remember that your STEM Master Teacher program is part of a larger effort aimed at enhancing teacher quality and to keep in mind your overall grant evaluation and the questions it is designed to address. Some of the high-priority questions regarding your STEM component will require modifications to your data collection plan; others will not. For example, being a TIF STEM grantee raises some special evaluation questions about how to support and assess teachers who are working to actualize a new vision of how to enhance teaching and learning in targeted STEM areas. If you plan to do classroom observations as one of the strategies for collecting data on the effectiveness of your STEM Master teachers, you will need to make sure that the observers have strong content expertise in STEM subjects, as well as a foundational understanding of good instructional practices and how these practices are particular to the teaching of STEM subjects. Observers would need to be able to assess three areas of knowledge in the context of a particular STEM field: content knowledge, pedagogical content knowledge, and pedagogical knowledge. However, in assessing effectiveness, you could also draw on data being collected for your overall grant evaluation like survey data, interview data, and student assessment data and disaggregate the data to look more closely at your STEM component.

---

\(^1\) While this guide focuses on the STEM Master Teacher component of the evaluation, in actuality such stakeholder convenings would likely address the evaluation of the complete project.
Select Data Collection Approaches for the STEM Master Teacher Component

Once you have prioritized your evaluation questions, the next set of steps in developing your evaluation plan concerns deciding how these questions will be addressed. The steps are as follows:

- Operationalizing the measurement of activities and outcomes;
- Determining the right balance of quantitative and qualitative data;
- Identifying existing data sources; and
- Selecting tools for primary data collection.

Frequently, evaluation questions embed concepts and constructs that are broad or not completely defined and require operationalization, especially in the outcome areas. The first step is clarifying what needs to be measured. For example, when an outcome is identified as “student performance,” there are multiple ways performance could be defined. One important and obvious definition is achievement as measured by annual state accountability tests that typically measure mathematics, science, or reading/language arts. Other possibilities include end-of-course assessments; ACT, SAT, or AP courses assessments; and Student Learning Objectives (SLOs). Other non-test-based definitions are also possible. Operationalizing what will be measured in an area like teacher skill development may be even more challenging. Are pedagogical skills the target? STEM content knowledge? Use of data? The answer obviously depends on what you are trying to do in your project. In developing your evaluation plan, it is important to be clear about the specific areas you expect your project to affect and ensure that there is shared agreement that these areas are seen by stakeholders as being valid and important for what you are trying to achieve.

Presented below are some simple items for you and your team to use to clarify and adequately define what needs to be measured. As you address these items, it is useful to think about your logic model and the critical components and connections that it lays out.

- Briefly describe the purpose of your TIF STEM component.
- What are the objectives for this component?
- Consider the first objective. Can this objective be broken down further? Break it down to the smallest unit.
- Is this objective measurable (can indicators and standards be developed for it)? If not, restate it.
- Using the measurable objective described above, define the criteria for success. (The plural “criteria” is used here as it is likely that more than one criterion will emerge as important.)
- Once you have completed the above steps, consider another objective and repeat the process.
In the context of your evaluation, you might develop the following answers:

• Briefly describe the purpose of your TIF STEM component.

*The purpose of the project is to enhance teacher quality and improve student achievement through the development of a STEM Master Teacher component embedded within a new human capital resource management system.*

• What are the objectives for this component?

*Enhancing the skills and knowledge of these STEM Master Teachers with regard to*

– Supporting STEM classroom teachers,

– Identifying/developing new instructional materials, and

– Creating and monitoring internship experiences for students.

• Consider the first objective. Can this objective be broken down further? Break it down to the smallest unit.

*Supporting STEM classroom teachers can be broken down into the following components:*

– Providing effective professional development around needed topic areas, such as assessing performance in teams using project-based learning or deepening teacher content knowledge;

– Providing modeling of lessons and strategies that address the needs of diverse groups of students; and

– Giving constructive feedback, addressing both the strengths and weaknesses of classroom teacher practices.

• Is this objective measurable (can indicators and standards be developed for it)? If not, restate it.

*Yes, these dimensions are measurable. Two potential indicators are the judgments of professional development providers and STEM Master Teachers. Another possible indicator is changes in the coaching practice of the STEM Master Teachers.*

• Using the indicators described above, define the criteria for success.

*By year 2 of the grant, at least 85 percent of STEM Master Teachers will report that the support they are receiving is meeting their needs and enhancing their skills.*

*By the end of year 4 of the grant, at least 80 percent of the STEM Master Teachers will show improvement in competency areas related to their specific roles and responsibilities.*

Once outcome areas are more completely specified, the next step is determining how to measure them. First, you need to consider whether quantitative or qualitative data are preferred. Generally, quantitative methods are appropriate for establishing relationships between variables and constructs; qualitative methods are suited to exploring these relationships and developing a better understanding of why they may be occurring (Witham et al., 2011). Quantitative approaches include assessments, surveys, and analyses of record data such as attendance, grades, and retention statistics. Such data are typically summarized using numbers—counts or percentages being frequent metrics. Qualitative approaches include document review, responses from in-depth interviews, focus groups, site visit/case study reports, and observations. While some audiences place more trust in quantitative data than qualitative data, evaluations can be strengthened when a mixed-method approach is adopted. Take, for example,
the question of why some teachers seek to become STEM Master Teachers and other highly effective STEM teachers do not. For this type of question, surveys and interviews provide complementary ways to get the information needed to arrive at the answer. Quantitative analysis of survey data could be used to get a broad-based look at whether the responsibilities are appealing and the rewards sufficient. Interviews would provide qualitative information about precisely what is or is not appealing about the choices and how they might be improved.

The second consideration is assessing whether or not these data already exist, and if they do, what the accessibility and quality of these data are. In judging quality, it is important to consider factors such as how the data were collected, what and how much data are missing, what kinds of quality control were undertaken, the extent to which the data are available at the right unit of disaggregation, and how privacy provisions might affect allowable uses of the data. For example, if you need assessment data at the teacher level and the data are only available at the grade level, the extant data may not be usable for your purpose. If you need data at a subtest level and they are only available for the overall test, this reality also raises questions about data usability.

Where no extant data source exists, as is likely to be the case with regard to questions regarding the career ladder program, the next decision is determining the primary data collection method that will be used. Several factors need to be considered—the timeframe for your work, the skills of the evaluation staff, the resources available, the characteristics of the project participants, and the methodological limitations that are inherent in the different data collection approaches. For example, it is likely that during the early stages of evaluation you will want to look closely at how your career ladder program is being received, especially by those who participate in it. Since the number of STEM Master Teachers will in most cases be relatively small, structured interviews would probably be the most productive approach.

If, however, your project is more mature and you are looking at questions involving larger groups of teachers, other data collection choices should be considered. For example, if you want to look at whether teachers have changed their instructional practices, you can ask them (using surveys or interviews) or observe them. If you use the survey or interview approach, you are likely to be able to ask the question of a large sample of teachers and get responses in a fairly short time period. Using these approaches, however, runs the risk of receiving misinformation related to the self-report process. Teachers may report that changes have occurred because they know they were expected to occur (even if they did not) or because they are mistaken in their perceptions. If you use observation, you rule out this source of bias and increase standardization but may run into other problems. That is, with observation, care must be taken to avoid observer bias, staff with appropriate skills and STEM content background need to be available, and resources and time available must be sufficiently large in order to observe an adequate amount of instruction by teachers over time. Another alternative would be to interview principals about whether or not they have noticed changes in instruction and, if so, what these changes might be. While this measure is somewhat less direct and is dependent to a large extent on principals’ skills and knowledge (as well as their having time to closely observe instruction), such data can be triangulated with other data sources to strengthen conclusions. A comprehensive review of the strengths and weaknesses of different data collection approaches can be found in Frechtling (2010).
Develop the Evaluation Design and Analysis Plan for Assessing the STEM Master Teacher Component

Another area to address in developing your evaluation plan is determining an evaluation design. In doing so, it is important to consider the following:

- Suiting evaluation design to nature of the work;
- Aligning the plan with project maturity;
- Assessing the utility/feasibility of different strategies;
- Looking for opportunities to do experimental studies; and
- Identifying appropriate analytic procedures.

One of the biggest challenges in developing your evaluation plan is deciding on the design for your work. Choosing a design involves careful thought about the questions you are trying to answer and the level of development of your idea, as well as the maturity of the project itself. Designs range from descriptive approaches that serve the primary purpose of richly capturing and documenting the implementation of a new strategy or use of a new tool, to design and development approaches that systematically explore different ways of delivering the intervention, to impact studies that are intended to provide a causal linkage between a clearly specified treatment and some desired set of outcomes (U.S. Department of Education and the National Science Foundation, 2013). The approach that is right for you is related to how well the phenomenon is understood, the level of maturity of your particular project, and the practical constraints of the specific context in which the intervention is being carried out.

Typically when an idea is new and being explored or a design is in its early stages, the most useful evaluation approach is a description of what is occurring. The “exploratory approach” can provide valuable information about both the challenges of implementing the approach and whether or not there appears to be some association between the implementation of the approach and the desired outcomes. For example, if you are just beginning to introduce STEM Master Teachers into your instructional plan and are unsure of the best way to do so, carefully describing and documenting the early stages of the introduction of this component would be very useful. Potential questions appropriate for the descriptive approach include the following:

- How were the STEM Master Teachers identified?
- What procedures were used to place them in different career ladder positions? What kind of preparation and training did your grant give them?
- What tasks did they undertake?
- What worked well and what did not?
- Is there any indication that their activities contributed to changes in teaching and learning?

Addressing these formative questions could be done through document review, observations of training sessions, and interviews with project leadership, professional developers and administrators, and the STEM Master Teachers themselves.
“Design studies,” which also rely primarily on description, push this exploration a step further, sequentially exploring variations on an approach. Design studies are developmental in nature; instead of committing to a predetermined structure, design studies allow you to organically explore the best way to approach the establishment of new roles and responsibilities or design new programs. Design studies address questions similar to those listed above, documenting results under the evolving procedures. The data (some parts, if not all) are collected several times in alignment with the change in program features, providing feedback to inform program evolution. Recognizing that neither exploratory studies nor design studies can provide strong evidence of a link between your treatment and changes in outcomes, they can, nonetheless, be very valuable in providing foundational data for moving forward.

Roughly speaking, an experimental design is an empirical approach that includes both treatment and comparison groups.

Once an approach or approaches have been established, and there are some reasons to believe they may be effective, the next step is examining them more scientifically, which means using some form of experimental design.

What is an experimental design? Roughly speaking, an experimental design is an empirical approach that includes both treatment and comparison groups. While other alternatives are possible, education evaluations most frequently rely on two types of experimental designs: randomized controlled trial or quasi-experimental.

- **Random assignment** (sometimes abbreviated as RCT) is an experimental technique for assigning subjects to different treatments (or no treatment). This approach is sometimes referred to as a “true experimental design.” It is believed that by randomizing treatment assignment, the group attributes for the different treatments will be roughly equivalent and therefore any differential outcomes can be attributed to the treatment itself. The RCT is considered the strongest design for assessing causal impact.

- **A quasi-experiment** is an empirical study that is also used to estimate the causal impact of an intervention on its target population. Quasi-experimental research designs share many similarities with RCTs but lack the element of random assignment to treatment or control. Instead, quasi-experimental designs typically allow the researcher to control the assignment to the treatment condition, matching on characteristics hypothesized to have some relationship to the desired outcomes.

Making strong claims about the effectiveness of an intervention program is not possible without an experimental or quasi-experimental design using comparison groups that take into account self-selection, attrition, baseline equivalence of groups, the quality of measurement instruments, and the sufficiency of the statistics reported.

If you are just exploring how to set up a STEM Master Teacher program, it likely that your design will be primarily descriptive. If you are exploring different approaches, you are still in the exploratory arena and again are likely to rely on description. If, however, you are systematically assessing the strengths and weaknesses of different approaches and wish to make claims regarding their relative effectiveness, your design needs to use an experimental or quasi-experimental approach.

Let’s look at two different questions and see what an evaluation addressing effectiveness might look like for each. The first question examines alternative ways of preparing and training STEM Master Teachers. The second looks more closely at the disciplinary knowledge needed for supporting the STEM instructional program.

Say you want to address the first question—preparation and training. Your issue is how to balance training in leadership, defined as guiding and supporting STEM classroom teachers, with training in new instructional practices. Your timeline and budget allow for a special summer session of 90 hours spread out over a 3-week period. How should these 90 hours be apportioned? 45 and 45? 30 hours for leadership and 60 hours...
for instructional practices? The reverse? One way of scientifically examining this is to take your pool of potential STEM Master Teachers, all of whom have similar backgrounds and experience, and assign them randomly to one of the three conditions. Of course, this means you must have a sufficiently large initial pool—20–30 teachers per group at a minimum. Since random assignment should control for other confounding variables, this evaluation design, along with a strong set of measures, should give you a pretty good idea of which preparation and training program works best with regard to supporting STEM classroom teachers. If random assignment is not possible for some reason—perhaps identification of STEM Master Teachers is being done on a rolling basis and it is not possible to wait until all candidates have been identified to offer the training—a second approach would be to compare successive cohorts of trainees, controlling for any differences in initial group status, either by judicious selection or statistical means.

If you have questions about how to determine the STEM-specific knowledge and skills that the STEM Master Teachers need, you could also construct an evaluation to look scientifically at your alternatives. For example, if your vision for effective STEM teaching and learning suggests that those who are likely to be effective in such a role must have expertise in more than one of the STEM disciplines, you may wish to explore how extensive that expertise might need to be. This can be done in a preliminary way by documenting the issues faced by potential STEM Master Teachers who bring different combinations of skills and knowledge to the task, which will allow you to generate some hypotheses about what is needed. In order to reach firmer conclusions about the impact on different skill sets, however, it would be important to confirm or disconfirm this assertion through an experimental design examining the effectiveness of appropriately selected groups of teachers showing the different skill patterns. Teacher (STEM classroom teachers) instructional practices and student outcomes could be compared for participants who had been supported by STEM Master Teachers posing different combinations of disciplinary skills. The understanding gained from this investigation could also provide suggestions for the types of professional development that might be needed to address content area weaknesses. In such a study, however, you would need to carefully control for other factors not related to disciplinary mix that might confound the outcomes—including STEM Master Teachers’ years of teaching experience, familiarity with the student population, and experience with the instructional approaches being used, examples of variables that might contribute to differential outcomes.

In developing your design, it is also important to think about the feasibility of what you are trying to do in terms of time, cost, and contextual constraints such as extant union agreements. Given the reality of TIF STEM projects, conducting true experimental evaluations can be difficult. Three challenges come to mind. First, attempting random assignment could raise tensions among the various stakeholder groups involved. In some cases, maintaining the buy-in of critical stakeholder groups requires a delicate balance that could be disturbed by requiring teachers or STEM Master Teachers to be assigned to instructional situations in this way. Second, even when such assignment is possible, changes that occur in the groups, such as differential attrition, can weaken any conclusions. Third, the number of STEM Master Teachers who participate in any one project is likely to be relatively small, making comparisons difficult. There are no quick fixes to these
problems. However, even if you cannot evaluate your complete STEM Master Teacher component using an experimental approach, there may be opportunities to do mini-studies that explore selected aspects of the component, like assessing the strengths and weaknesses of different ways of presenting a particular curriculum unit.

The design you adopt has implications for the kinds of statistical analyses that will be needed. Complex studies involving multiple comparison groups will require far more sophisticated analytic procedures than simple two-group studies. Studies that track changes in performance over time, involving repeated measures, have different requirements than more delimited one time pre-post data comparisons. In some cases, for example where a rich description of implementation is the focus of the evaluation, statistical analyses may not be appropriate.

In planning your evaluation, it is important to think about how the data will be analyzed early on as part of your planning process. Time and procedures should be established for addressing data quality issues such as dealing with missing data, addressing low response rates, checking for internal consistency and handling of outliers. While modifications may need to be made once you have collected your data, it is important to have an initial plan to guide your work. In creating this plan, it is important to be mindful of the assumptions underlying application of different approaches—for example, sample size requirements and whether a variable is continuous or categorical. The What Works Clearinghouse (http://ies.ed.gov/ncee/wwc/) provides a number of reports detailing criteria for appropriate application of different statistical procedures.
Disseminate Findings

The findings from your evaluation can be used for a number of purposes. Throughout this guide, we have stressed the importance of using data to improve your program. What you find out in the grant’s early stages can be used to modify or fine-tune your plans. This is not, however, the only way evaluation data can be used. Indeed, your data provide a way of communicating to your stakeholders what you are learning and engaging them more fully in your work as it proceeds. In this chapter, we discuss the following:

- Creating a communication strategy;
- Using the findings for program improvement; and
- Creating different products to meet the needs of different audiences.

Previous evaluations of TIF programs have found that the new PBCS are sometimes difficult to understand and that communication is a very important aspect of a successful project (MacAllum, Wells, & Ristow, 2011). It is important, therefore, that TIF STEM projects develop an explicit communications strategy for informing stakeholder audiences about the purpose of the TIF STEM grant, how the new career ladder system is being structured, how well it is working, and how plans are evolving. Ideally, communication begins at the proposal development phase, when plans are being structured and the buy-in of different groups is being sought. School board members, union representatives, teacher and principal associations, and the community should all be part of the conversation. After the grant award, this communication should continue, as it is likely that stakeholders will need help understanding the project as it evolves.

Part of this communication strategy should be sharing with stakeholders what is being found from the evaluation, including both program strengths and areas needing addressing. When results are positive and plans seem to be working well, supporters are reinforced, and skeptics may begin to change their minds. When results suggest changes should be made, being open and proactive about how you plan to address problems that have arisen can gain you supporters and forestall any issues that might come up if problems are made public through some other means. Discussing evaluation findings in an open forum can also lead to the identification of potential solutions that might otherwise be overlooked. These discussions can, additionally, identify any misconceptions that various audiences might hold about your work and provide the opportunity for clarification.

The reporting of evaluation findings should be guided by both your grant reporting requirements and your project development plan. To the extent possible, reporting should be aligned with key decision-making points in your work. If a change is needed in the structuring of your career ladder, when does such a change need to be recognized in order to implement it in a timely fashion? If there is a problem with any of the partnerships you have established, when do you need to reconsider that partnership in order to have the needed opportunities for students in place? If you have questions about the efficacy of some of the instructional strategies that are being used in your STEM classrooms, when do you need to know about them to take the steps needed to make changes? While it is not always possible to have a formal report available at such decision points, less-formal communication venues can be used to ensure that critical information gets to your team when it
is needed. While the term “external evaluator” at one time connoted an individual or group that provided findings at the end of a year or even at the end of multiple years of a grant, that expectation has changed. With the recognition that evaluation information is useful to the project team, as well as a requirement of the funder, new expectations have emerged. It is not just the internal evaluator that is expected to be in conversation with project staff. External evaluators are also frequently expected to interact more regularly. Evaluators are expected to share their findings on a more frequent basis, and project leadership is expected to use these findings for reflection, discussion, and, where necessary, modification to the plans that are in place.

TIF STEM and TIF grantees are also encouraged to share their evaluation findings with each other. The Communities of Practice that have been established at www.TIFcommunity.org and https://TIFSTEMcommunity.org have both set aside areas for discussion of evaluation issues. These communities allow profiling of how different projects are approaching the evaluation of their grant, as well as provide protected space to share challenges and solutions.

Preferred strategies for dissemination of evaluation results have also undergone some important changes. While some audiences may want the typical comprehensive technical report that is provided by an evaluator, more user-friendly communication formats are encouraged. A brief or focused fact sheet may be the best way to reach certain audiences. Findings can be shared via a project website, as well as in written form. And, it is useful to set up public forums for two-way communication at strategic points in a project’s development. Of course, this sets up new requirements for evaluators who frequently are asked to participate in such forums. Not only must the evaluator be able to design and implement an evaluation that is technically sound, but the evaluator must also be able to discuss the findings from a layperson’s perspective. The ability to communicate findings to various audiences is increasingly being seen as a requirement for evaluators both at the individual grant level and where broad, program-level evaluations are the focus.
References


Appendix A: Evaluation Resources


This guide, written specifically for mathematics and science partnership (MSP) projects funded by the U.S. Department of Education, presents a series of criteria to consider in evaluating the rigor of an MSP evaluation report, using a comparison group design. Handling of the topics attrition, baseline equivalence of groups, quality of measurement instruments, and relevant statistics reported is discussed.


This compendium contains resources to help researchers and policy makers review measures used in NCEE evaluations of educational evaluation. The measures cover grades preschool through 12. The compendium discusses criteria and their importance in selecting measures for assessing intervention impacts on student, teacher, and classroom outcomes, and presents profiles and tables that summarize the measures.


The guide was developed to provide education practitioners with user-friendly tools to distinguish practices supported by rigorous evidence from those that are not. The guide contains four sections: a description of the RCT and its role in establishing rigorous evidence, how to evaluate whether an intervention is backed by strong evidence of effectiveness, how to evaluate whether an intervention is backed by possible evidence of effectiveness, and factors to consider in implementing evidence-based programs. The appendices contain a list of websites to consult for information on evidence-based programs and a checklist of evaluating the evidentiary basis for an intervention.


This book provides detailed discussion of how to develop and use a logic model as part of a program or project evaluation. It discusses the relationship between logic modeling and the development of theories of action, showing how these two approaches complement each other. The book addresses the use of logic models across a range of different programs varying in focus, duration, and complexity.

This handbook, developed for projects funded by the National Science Foundation, provides a detailed, step-by-step guide for conducting project evaluation. Written to address a wide range of evaluation questions, the handbook includes detailed discussion of the pros and cons of various data collection methodologies, as well as a discussion of a range of statistical analysis techniques. Special sections address culturally responsive evaluation, multi-site evaluation, and evaluation of projects addressing postsecondary education.


The report emerged from several years of discussion among experts in STEM education and evaluation from the Department of Education and the National Science Foundation. Its purpose is to document and describe the roles of various types of research studies in generating evidence about strategies and interventions for increasing student learning. The research pipeline discussed addresses three types of research: foundational and exploratory/early stage research, design and development research, and efficacy/effectiveness/scale-up research. The document describes 1) the purpose of the research and how it contributes to the evidence base, 2) the theoretical justifications required for doing the type of research, 3) the expectations for research design and expected products, and 4) expectations for review of the products from each type of research.


Funded by the Department of Education as clearinghouse for research, the What Works Clearinghouse has been established to review educational research studies and inform researchers, practitioners, and policy makers about the extent to which various practices are supported by high-quality evidence-based research. The Clearinghouse produces papers on both what has been learned from various kinds of studies and on standards for high-quality research. Included are practice guides, intervention reports, single-study reviews, and a database of reviewed research.


This guide was developed to provide practical assistance to nonprofits engaged in program evaluation. It is intended to provide readers with an orientation to the underlying principles of logic modeling to enhance their program planning, implementation, and dissemination activities. It provides a basic introduction to logic modeling, how to create a logic model, and the various purposes for which it can be used.


This concise guide, developed for the Teacher Incentive Fund (TIF) Program, describes six steps in designing and implementing an evaluation of performance-based compensation systems. These steps are 1) describing the program; 2) developing evaluation questions; 3) using qualitative, quantitative, and mixed-method approaches; 4) the evaluation selection framework; 5) disseminating evaluation results; and 6) managing the TIF program evaluation processes. Supporting information and examples are provided in the appendices.