Evaluation Matters

Getting the Information You Need From Your Evaluation

A guide for educators to build evaluation into program planning and decision-making, using a theory-driven, embedded approach to evaluation

U.S. DEPARTMENT OF EDUCATION

EVALUATION MATTERS: GETTING THE INFORMATION YOU NEED FROM YOUR EVALUATION

A GUIDE FOR EDUCATORS TO BUILD EVALUATION INTO PROGRAM PLANNING AND DECISION-MAKING, USING A THEORY-DRIVEN, EMBEDDED APPROACH TO EVALUATION

Prepared by:

Susan P. Giancola
Giancola Research Associates, Inc.

Prepared for:

U.S. Department of Education
Office of Elementary and Secondary Education
School Support and Rural Programs
DRAFT 2014

This publication was prepared for the U.S. Department of Education under Contract Number ED-07-CO-0098 (Contracting Officer's Representatives: Kenneth Taylor, Sharon Horn, and Vickie Banagan) with Kauffman & Associates, Inc. The views expressed in this publication do not necessarily reflect the positions or policies of the U.S. Department of Education. For the reader's convenience, this publication contains information about and from outside organizations, including hyperlinks and URLs. Inclusion does not constitute endorsement by the Department of any outside organization or the products or services offered or views expressed. Nor is any endorsement intended or implied of the consulting firm "Evaluation Matters." In fact, this publication was not prepared with help from or in consultation with, in any manner, that firm.

U.S. Department of Education
Arne Duncan
Secretary

Office of Elementary and Secondary Education
Deb Delisle
Assistant Secretary

School Support and Rural Programs
Jenelle V. Leonard
Director

DRAFT January 2014

This publication is in the public domain, except for the 1.1 Implementation Rubric in Appendix B, for which the William & Ida Friday Institute for Educational Innovation at North Carolina State University kindly granted permission to reproduce herein. Authorization to reproduce *Evaluation Matters* in whole or in part—except for the 1.1. Implementation Rubric—is granted. Any further use of the 1.1 Implementation Rubric in Appendix B is subject to the permission of the William & Ida Friday Institute (for more information, email Jeni Corn, director of program evaluation programs, Friday Institute, at jeni_corn@ncsu.edu). The citation for *Evaluation Matters* should be: U.S. Department of Education, Office of Elementary and Secondary Education, School Support and Rural Programs, *Evaluation Matters: Getting the Information You Need From Your Evaluation*, Washington, D.C., 2014.

To obtain copies of this publication,

Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 22207, Alexandria, VA 22304.
Or **fax** your request to 703-605-6794.
Or **email** your request to edpubs@edpubs.ed.gov.
Or **call** in your request toll-free to 1-877-433-7827 (1-877-4-ED-PUBS). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY), should call 1-877-576-7734. If 877 service is not yet available in your area, call 1-800-872-5327 (1-800-USA-LEARN).
Or **order online** at http://edpubs.gov.

On request, this publication is available in alternate formats, such as Braille, large print, audiotape, or compact disk. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-260-0818. In addition, if you have difficulty understanding English, you may request language assistance services for Department information that is available to the public. These language services are available free of charge. If you need more information about interpretation or translation services, please call 1-800-USA-LEARN (1-800-872-5327) (TTY: 1-800-437-0833), or email the content contact below.

Content Contact:
Nancy Loy, Project Officer
Phone: 202-205-5375; Email: Nancy.Loy@ed.gov

Contents

Acknowledgements ..... **iv**

Before You Get Started ..... **vi**

Introduction **1**
- What Is the Purpose of the Guide? 1
- Why Evaluate and What Do I Need to Consider? 2
- Where Do I Start? 6
- How Is the Guide Organized? 7

Embedding Evaluation Into the Program **9**
- STEP 1: DEFINE – What Is the Program? 9
- STEP 2: PLAN – How Do I Plan the Evaluation? 24
- STEP 3: IMPLEMENT – How Do I Evaluate the Program? 53
- STEP 4: INTERPRET – How Do I Interpret the Results? 61
- STEP 5: INFORM and REFINE – How Do I Use the Evaluation Results? 68

Appendix A: Embedded Evaluation Illustration – READ* **74**
- Program Snapshot 74
- Step 1: Define the Program 74
- Step 2: Plan the Evaluation 80
- Step 3: Implement the Evaluation 94
- Step 4: Interpret the Results 107
- Step 5: Inform and Refine – Using the Results 120

Appendix B: Embedded Evaluation Illustration – NowPLAN* **124**
- Program Snapshot 124
- Step 1: Define the Program 124
- Step 2: Plan the Evaluation 132
- Step 3: Implement the Evaluation 133
- Step 4: Interpret the Results 145
- Step 5: Inform and Refine – USING the Results 145

Appendix C: Evaluation Resources **162**
- Evaluation Approaches 162
- Program Theory and Logic Modeling 163
- Research and Evaluation Design, Including Reliability and Validity 163
- Threats to Validity 164
- Budgeting Time and Money 164
- Ethical Issues 164
- Data Collection, Preparation, and Analysis 165
- Evaluation Pitfalls 165
- Interpreting, Reporting, Communicating, and Using Evaluation Results 166

Appendix D: Evaluation Instruments for Educational Technology Initiatives **167**

Appendix E: Evaluation Templates **173**

Appendix F: Lists of Tables and Figures **175**
- List of Tables 175
- List of Figures 176

Acknowledgements

This guide was created with the valuable input and advice from many individuals. Some individuals provided input into shaping the initial conceptual framework of the guide, some in editing portions of the guide, and some in reviewing draft versions of the guide.

Kathleen Barnhart, Principal Education Consultant, Illinois State Board of Education

Barbara DeCarlo, Retired Principal and Teacher

Beverly Funkhouser, Adjunct Professor, University of Delaware

Rick Gaisford, Educational Technology Specialist, Utah State Office of Education

Robert Hampel, Interim Director, School of Education, University of Delaware

Vic Jaras, Education Technology Director, Iowa Department of Education

Karen Kahan, Director of Educational Technology, Texas Education Agency

Tonya Leija, Reading Recovery Teacher Leaders, Spokane Public Schools

Melinda Maddox, Director of Technology Initiatives, Alabama Department of Education

Daniel Maguire, District Instructional Technology Coach, Kennett Consolidated School District

Jeff Mao, Learning Technology Policy Director, Maine Department of Education

Jennifer Maxfield, Research Associate, Friday Institute for Educational Innovation, North Carolina State University

Brandy Parker, Graduate Research Assistant, Friday Institute for Educational Innovation, North Carolina State University

Shannon Parks, State Education Administrator, Technology Initiatives, Alabama Department of Education

Barry Tomasetti, Superintendent, Kennett Consolidated School District

Bruce Umpstead, Educational Technology Director, Michigan Department of Education

Carla Wade, Technology Education Specialist, Oregon Department of Education

Brent Williams, Director, Educational Technology Center, Kennesaw State University

Thanks also to ***Jeni Corn,*** Director of Evaluation Programs, with the William & Ida Friday Institute for Educational Innovation at North Carolina State University for obtaining approval to use the 1:1 Implementation Rubric in the *Evaluation Matters* guide.

I would like to extend a special thank you to ***Jenelle Leonard,*** Director of School Support and Rural Programs (SSRP), Office of Elementary and Secondary Education (OESE) at the U.S. Department of Education, for being the driving force in the creation of this guide.

In addition, I would like to thank ***Andy Leija, Kelly Bundy, Kim Blessing, Janelle McCabe, and Anna Morgan*** with Kauffman & Associates, Inc. for their continued support during the creation of the guide.

And finally, I would like to especially thank ***Nancy Loy*** (SSRP/OESE) at the U.S. Department of Education for her constant assistance and support throughout the development and writing of the guide, from brainstorming ideas to reading multiple drafts to facilitating review of the evaluation guide.

Portions of this guide were adapted from *An Educator's Guide to Evaluating the Use of Technology in Schools and Classrooms* prepared by Sherri Quiñones and Rita Kirshstein at the American Institutes for Research for the U.S. Department of Education in 1998 (Nancy Loy, Project Officer).

Before You Get Started

Some who use this guide, especially those who are unfamiliar with evaluation or educational program design, may decide to read it cover to cover. However, most who read the guide will likely use it as a compendium and a companion with which they will travel to those portions that are relevant to their current needs. To facilitate this use, there are several features that will aid you in your navigation through the guide.

Click on the I note icon to go to excerpts from **Appendix A: Embedded Evaluation Illustration – READ*** that appear throughout the text to illustrate each step of the evaluation process. If you find the excerpts interspersed within text to be distracting, you may want to skip them in the main text and instead read the example in its entirety in Appendix A. There you will find a detailed example of a theory-driven, embedded program evaluation from its inception through the use of its first-year results. **Appendix B: Embedded Evaluation Illustration – NowPLAN*** provides another example. Both examples set out in this guide are provided solely for the purpose of illustrating how the principles in this guide can be applied in actual situations. The programs, characters, schools, and school districts mentioned in the examples are fictitious.

Click on the R note icon to see additional resources on a topic included in **Appendix C: Evaluation Resources**.

Introduction

What Is the Purpose of the Guide?

Who Is this Guide For?

This guide is written for educators. The primary intended audience is state- and district-level educators (e.g., curriculum supervisors, district office personnel, and state-level administrators). Teachers, school administrators, and board members also may find the guide useful. It is intended to help you build evaluation into the programs and projects you use in your classrooms, schools, districts, and state. This guide will also provide a foundation in understanding how to be an informed, active partner with an evaluator to make sure that evaluation provides the information you need to improve the success of your program, as well as to make decisions about whether to continue, expand, or discontinue a program.

No previous evaluation knowledge is needed to understand the material presented. However, this guide may also be useful for experienced evaluators who want to learn more about how to incorporate theory-based evaluation methods into their programs and projects.

In addition to using the guide to embed evaluation within your program, the guide will be useful for

- *State education agencies* during preparation of program and evaluation guidelines within Requests for Proposals (RFPs), in order to facilitate uniform assessments of proposals and for districts to know how their proposals will be assessed.
- *School districts* in responding to RFPs or in writing grant proposals, in order to set clear expectations for what a program is intended to accomplish and how the evaluation will be embedded within the program to measure changes as a result of the program.
- *Teams* of *educators* to show value added for a program, in order to build program support and provide budget justification.
- *Program staff* to tell the story of a program using data.
- *Organizations* for evaluation training and professional development

How Is this Guide Different From Other Evaluation Guides?

There are many evaluation guidebooks, manuals, and tool kits readily available. So, what makes the material presented in this guide different from other evaluation guides? This guide is written with you, the educator, in mind. It outlines an evaluation approach that can be built into your everyday practice. It recognizes the preciousness of time, the need for information, and the tension between the two. The theory-driven, embedded approach to evaluation is not an additional step to be superimposed upon what you do and the strategies you use but rather a way to weave evaluation into the design, development, and implementation of your programs and projects.

> The term *program* is used broadly in this guide to represent activities, small interventions, classroom-based projects, schoolwide programs, and district or statewide initiatives.

This guide will help you to embed evaluation within your program in order to foster continuous improvement by making information and data the basis upon which your program operates. The step-by-step approach outlined in this guide is not simply a lesson in "how to evaluate" but rather a comprehensive approach to support you in planning and understanding your program, with a rigorous evaluation included as an integral part of your program's design.

In Appendices A and B, you will find two examples of educators building evaluation into their everyday practices. Through a narrative about programs, characters, schools, and school districts that are fictitious, each example is designed to illustrate how the principles in this guide can be applied in actual situations. While embedded evaluation can be used for any type of program you may be implementing, these illustrations specifically focus on programs that involve infusing technology into the curriculum in order to meet teaching and learning goals.

Why Evaluate and What Do I Need to Consider?

Why Evaluate?

Evaluation is important so that we can be confident the programs we are using in our schools and classrooms are successful. A common criticism regarding evaluation is that it takes time and resources that could be dedicated to educating students. However, evaluation, done properly, can actually result in better quality practices being delivered more effectively to enhance student learning.

You would not hire new teachers without regularly monitoring and mentoring to help them improve their skills and foster student success. Would you adopt and maintain a new curriculum full scale without being sure that student learning improved when you tested the

new curriculum? What if student learning declined after implementing a new curriculum? How would you know whether the curriculum did not work well because it was a faulty curriculum, or because teachers were not trained in how to use the curriculum, or because the curriculum was not implemented properly? Building evaluation into your educational programs and strategies enables you to make midcourse corrections and informed decisions regarding whether a program should be continued, expanded, scaled down, or discontinued.

> Evaluation enables you to identify and use better quality practices more effectively to improve learning outcomes.

A primary purpose of evaluation is to make **summative decisions**. You can use summative evaluation results from rigorous evaluations to make final, outcome-related decisions about whether a program should be funded or whether program funding should be changed. Summative decisions include whether to continue, expand, or discontinue a program based on evaluation findings.

Another important purpose of evaluation is to make **formative decisions**. You can use formative evaluation data from rigorous evaluations to improve your program while it is in operation. Formative evaluation examines the implementation process, as well as outcomes measured throughout program implementation, in order to make decisions about midcourse adjustments, technical assistance, or professional development that may be needed, as well as to document your program's implementation so that educators in other classrooms, schools, or districts can learn from your program's evaluation.

Who Should Do the Evaluation?

Once you have decided to evaluate the implementation and effectiveness of a program, the next step is to determine who should conduct the evaluation. An evaluation can be conducted by someone internal to your organization or someone external to your organization. However, the ideal arrangement is a partnership between the two, i.e., forming an evaluation team that includes both an internal and an external evaluator.

> Preferably, evaluation is a partnership between staff internal to your organization assigned to the evaluation and an experienced, external evaluator.
>
> Such a partnership will ensure that the evaluation provides the information you need for program improvement and decision-making. It also can build evaluation capacity within your organization.

An **internal evaluator** may be someone at the school building, district office, or state level. For evaluations that focus on program improvement

and effectiveness, having an internal evaluator on your evaluation team can foster a deeper understanding of the context in which the program operates. Involving people inside your organization also helps to build capacity within your school or district to conduct evaluation. An internal evaluator should be someone who is in a position to be objective regarding program strengths and weaknesses. For this reason, choosing an internal evaluator who is *responsible* for the program's success is not recommended and may compromise the evaluation. In order to maintain objectivity, an internal evaluator should be external to the program. However, while staff internal to the program itself should not be part of the evaluation team, they should certainly partner with the evaluation team in order to ensure that the evaluation informs the program during every phase of implementation.

It is good practice to have an **external evaluator** be part of your evaluation team. Using an external evaluator as a "critical friend" provides you with an extra set of eyes and a fresh perspective from which to review your design and results. Professional evaluators are trained in the design of evaluations to improve usability of the findings, and they are skilled in data collection techniques such as survey design, focus group facilitation, conducting interviews, choosing quality assessments, and performing observations. An experienced evaluator can also help you analyze and interpret your data, as well as guide you in the use of your results. Further, when you are very close to the program being evaluated, objectivity or perceived objectivity may suffer.

Partnering with an external evaluator can improve the credibility of the findings, as some may question whether an evaluator internal to an organization can have the objectivity to recognize areas for improvement and to report results that might be unfavorable to the program. For some programs, you may choose to use an evaluator who is external to your organization to be the sole or

The choice of who conducts your evaluation should depend upon the anticipated use of the results and the intended audience, as well as your **available resources**.

For some programs, while an external evaluator is preferred, funding an evaluator who is external to your organization may not be feasible. In such cases, partnering with an evaluator who is internal to your organization, yet external to your program, might work well.

Other potentially affordable evaluation options include putting out a call to individuals with evaluation experience within your community who might be willing to donate time to your program; contacting a local university or community college regarding faculty or staff with evaluation experience who might work with you at a reduced rate; asking your local university if there is a doctoral student in evaluation who is looking for a research opportunity or dissertation project; or exploring grant opportunities that fund evaluation activities.

primary evaluator. An external evaluator may be a researcher or professor from your local university or a professional evaluator from a private evaluation firm.

> The focus of embedded evaluation is to enable educators to build and implement high-quality programs that are continuously improving, as well as for educators to know when to discontinue programs that are not working.

The choice of who conducts your evaluation should depend upon the anticipated use of the results and the intended audience, as well as your available resources. If evaluation results are to be used with current or potential funding agencies to foster support and assistance, contracting with an external evaluator would be your most prudent choice. If the evaluation is primarily intended for use by your organization in order to improve programs and understand impact, an evaluation team comprised of an internal and an external evaluator may be preferred. Connecting with someone external to your organization to assist with the evaluation and results interpretation will likely enhance the usability of your evaluation and the credibility of your evaluation findings. Evaluation as a partnership between an internal evaluator and an external evaluator is the ideal arrangement to ensure the utility of the evaluation and its results.

For some programs, while an external evaluator might be preferred, funding an evaluator who is external to your organization may not be feasible. In such cases, partnering with an evaluator who is internal to your organization, yet external to your program, might work well. For instance, staff from a curriculum and instruction office implementing a program might partner with staff from another office within the district, such as an assessment or evaluation office, to conduct the evaluation.

If resources are not available for an external evaluator and there is no office or department in your organization that is not affected by your program, you may want to consider other potentially affordable evaluation options. You could put out a call to individuals with evaluation experience within your community who might be willing to donate time to your program, contact a local university or community college regarding faculty or staff with evaluation experience who might work with you at a reduced rate, ask your local university if there is a doctoral student in evaluation who is looking for a research opportunity or dissertation project, or explore grant opportunities that fund evaluation activities.

What Is Embedded Evaluation?

The embedded evaluation approach presented in this guide is one of many approaches that can be taken when conducting an evaluation. **Embedded evaluation** combines elements from several approaches, including theory-based evaluation, logic modeling, stakeholder evaluation,

and utilization-focused evaluation. See **Appendix C: Evaluation Resources** for resources with additional information on evaluation approaches.

Further, it is important to note that evaluation is not a linear process. While the steps of embedded evaluation may appear as if they are linear rungs on a ladder culminating with the final step, they are not rigid steps. Rather, embedded evaluation steps build on each other and depend upon decisions made in prior steps, and information learned in one step may lead to refinement in a previous step. The steps of embedded evaluation are components of the evaluation process that impact and influence each other. What you learn or decide in one step may prompt you to return to a previous step for modifications and improvements. Just as programs are ongoing, evaluation is dynamic.

> Evaluation is a dynamic process. While embedded evaluation leads the evaluator through a stepped process, these steps are not meant to be items on a checklist. Information learned in one step may lead to refinement in a previous step. The steps of embedded evaluation are components of the evaluation process that impact and influence each other.

The dynamic nature of evaluation and the interconnectedness of an embedded evaluation with the program itself may seem amiss to researchers who prefer to wait until a predefined time to divulge findings. And inarguably, having a program stay its course without midcourse refinements and improvements would make cross-site comparisons and replication easier. However, embedded evaluation is built upon the principle of continuous program improvement. With embedded evaluation, as information is gathered and lessons are learned, the program is improved. The focus of embedded evaluation is to enable educators to build and implement high-quality programs that are continuously improving, as well as to determine when programs are not working and need to be discontinued. The overall purpose of designing a rigorous, embedded evaluation is to aid educators in providing an effective education for students.

Where Do I Start?

Just as the first step in solving a problem is to understand the problem, the first step in conducting an evaluation is to **understand what you want to evaluate**. For the purposes of this guide, what you want to evaluate is referred to as the "program." It is important to note that the term *program* is used broadly in this guide to represent small interventions, classroom-based projects, schoolwide programs, and districtwide or statewide initiatives.

> The first step in evaluation is to understand what it is you want to evaluate.

You can use the evaluation process that is presented in this guide to define and evaluate a small project, as well as to understand and evaluate the inner workings of large programs and initiatives. Regardless of the size or type of program, understanding the program is not only the first step in evaluation. It also is the most important step. Defining why your program should work and making the theory that underlies your program explicit lay the foundation upon which you can accomplish program improvement and measure program effectiveness.

How Is the Guide Organized?

Steps to Embed Evaluation Into the Program

This guide presents a framework to aid you in embedding evaluation into your program planning, design, and decision-making. You will be led step-by-step from documenting how and why your program works to using your evaluation results (see **Figure 1: Embedded Evaluation Model**). The framework is based on the following five steps:

STEP 1: DEFINE – What is the program?

STEP 2: PLAN – How do I plan the evaluation?

STEP 3: IMPLEMENT – How do I evaluate the program?

STEP 4: INTERPRET – How do I interpret the results?

STEP 5: INFORM (a) and REFINE (b) – How do I use the results?

Throughout the guide, the boxed notes highlight important evaluation ideas. As mentioned earlier,

I notes provide excerpts from **Appendix A: Embedded Evaluation Illustration – READ*** to illustrate the process of designing an evaluation from understanding the program to using results, and

R notes indicate that additional resources on a topic are included in **Appendix C: Evaluation Resources**.

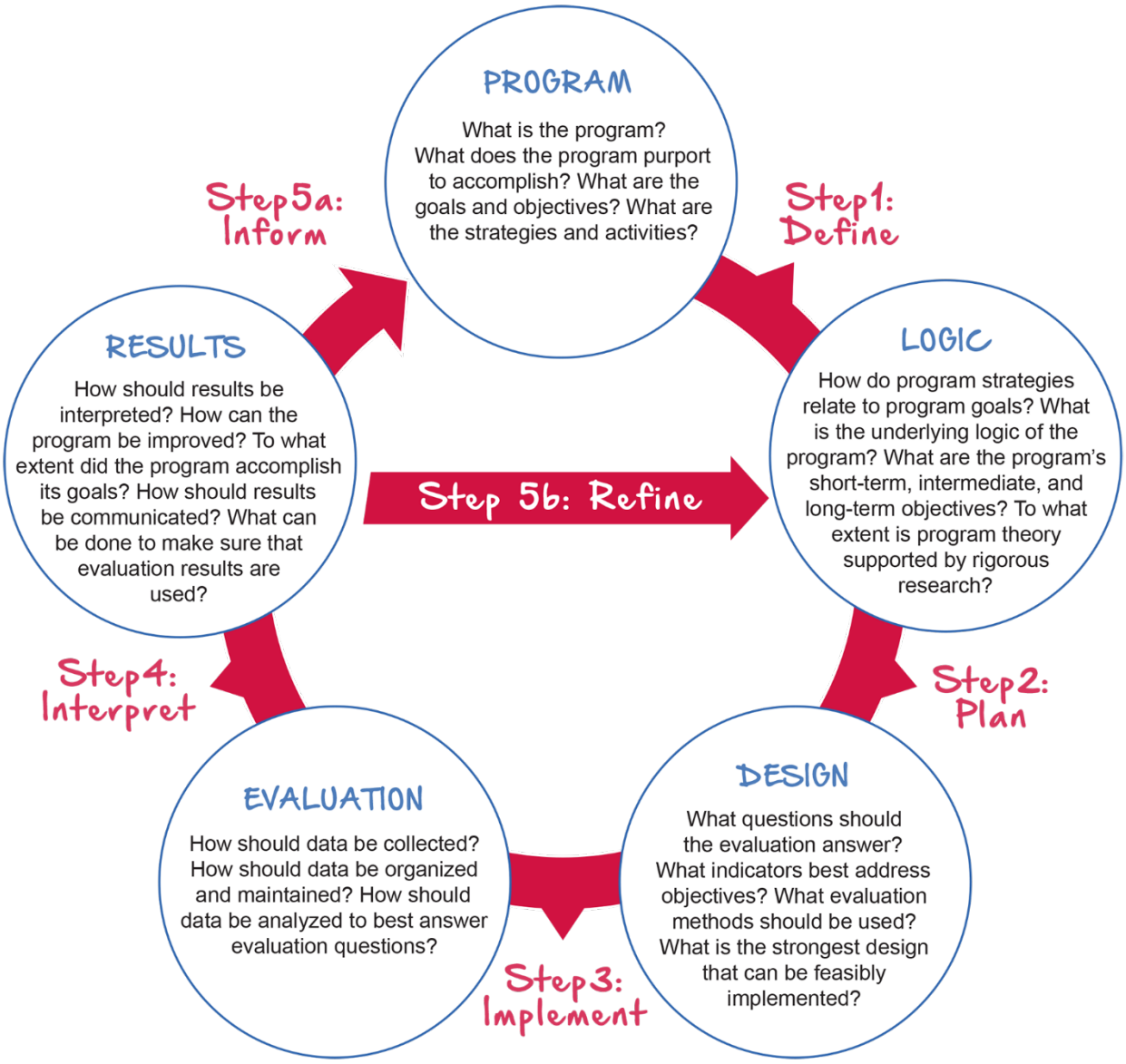
Appendices

Appendices A and B provide examples of theory-driven, embedded evaluations of two programs that involve infusing technology into the curriculum in order to meet teaching and learning goals. These examples are provided solely for the purpose of illustrating how the principles in this guide can be applied in actual situations. The programs, characters, schools, and school districts mentioned in the examples are fictitious. The examples include methods and tools to aid you as you build evaluation into your programs and projects and become an informed, active partner with the evaluator.

The illustration in **Appendix A: Embedded Evaluation Illustration – READ*** is of a districtwide reading program that uses technology to improve literacy outcomes and to assess reading progress. The illustration in **Appendix B: Embedded Evaluation Illustration – NowPLAN*** focuses on a building-level evaluation of a statewide strategic technology plan. This example builds evaluation into the everyday practice of educators in order to improve instruction and monitor strategic planning components.

Appendix C: Evaluation Resources and **Appendix D: Evaluation Instruments for Educational Technology Initiatives** include evaluation resources and information about instruments that you may find useful for your evaluations. **Appendix E: Evaluation Templates** includes a logic model template you can use to define your program and an evaluation matrix template to use to plan your evaluation. Finally, **Appendix F: Lists of Tables and Figures** appears at the end of the guide.

Figure 1: Embedded Evaluation Model



Embedding Evaluation Into the Program

STEP 1: DEFINE – What Is the Program?

How Can I Find Out More About the Program? (Understanding the Program)

The first step to conducting your evaluation is to understand what you want to evaluate. Whether you are evaluating a new program or a program that you have been using for some

For the past 5 years, reading scores in the Grovemont School District have been declining. The curriculum supervisor, Mrs. Anderson, has tried many strategies to improve reading skills. However, scores continue to decline. Mrs. Anderson has been searching for curricular and assessment materials that are better aligned with state reading standards and that provide ongoing standards-based assessment data. Mrs. Anderson found a program called READ (Reading Engagement for Achievement and Differentiation) that looked promising. After reviewing research on the program and documentation from the vendor as well as numerous discussions and interviews with other districts that had implemented the program, Mrs. Anderson and the district superintendent decided to present the READ program to the school board, in order to gain approval for funding the program for Grades 3-5.

At last month's meeting, the school board voted to partially fund the READ program. Due to recent state budget cuts, the school board was only able to fund the program at 50% for 2 years. At the end of the 2 years, the board agreed to revisit its funding decision. The board required an evaluation report and presentation due in September of each year.

Before starting to plan the READ program, Mrs. Anderson invited one teacher from each of the district's six elementary schools, the district reading coach, one of the district's reading specialists, and the district technology coordinator to join the READ oversight team. This 10-member team was charged with planning the READ program and its evaluation. The team asked an evaluator from the local university to conduct the READ evaluation and to attend oversight team meetings.

Note: The examples set out in this guide are provided solely for the purpose of illustrating how the principles in this guide can be applied in actual situations. The programs, characters, schools, and school districts mentioned in the examples are fictitious.

time, it is still important to begin from the basics in understanding how a program works. Do not rely on what you already know about the program or what you believe the program is intended to accomplish. Instead, take what you know, and build upon it with information from multiple sources. By doing this, you will have a full understanding of the program including multiple perspectives and expectations, as well as basic underpinnings and complex inner workings.

So, how do you find out more about the program? If you have experience with the program, you should first document what you know. You may want to investigate if any rigorous **previous evaluations** have been conducted of the program. If well designed and well carried out, previous evaluations can provide useful information regarding how a program operates.

Another good source from which you can learn more about the program is **existing documentation**. Documents such as technology plans, curriculum materials, strategic plans, district report cards, user manuals, and national, state, or district standards may have useful information for understanding your program and the context in which it will be implemented. Further, you may want to talk with people who are most familiar with the program, such as vendors and people from other districts that have implemented the program. Consider

The oversight team asked the external evaluator, Dr. Elm, to help them plan the evaluation. Dr. Elm suggested that the oversight team build evaluation into its program as the team is designing it. By embedding evaluation into the program, information from the evaluation would be available to guide program implementation. Evaluation data would both drive program improvement and be the foundation for future decisions regarding whether the program should be continued, expanded, scaled down, or discontinued.

The oversight team members invited Dr. Elm to lead them through the process of building evaluation into their program planning. Dr. Elm explained that the first step is to gain a thorough understanding of the program. In doing this, Mrs. Anderson shared the materials she had already reviewed with the oversight team. In addition, the oversight team contacted four school districts that had used the READ program successfully in order to learn more about the program. To develop a thorough and shared understanding of the context in which the READ program would be implemented, the team reviewed the state's reading standards, the district's strategic plan, the district's core learning goals and curriculum maps in reading, and the district's technology plan. The team also examined reading grades and state reading assessment scores for the district as a whole, as well as by school, English Language Learner (ELL) status, and special education status for the past 5 years.

conducting **interviews and group discussions** to learn more about their insight into the program, how it operates, and what goals it is intended to achieve.

Why Should the Program Work? (Explaining the Program Theory)

Once you have a good understanding of the program, the next step is for you to document more thoroughly what you know about the program. The first component in explaining the program is to describe the program's **goals and objectives**. Goals should reflect a shared understanding among program stakeholders as to what the program should achieve. What is the program intended to accomplish? How would you know if it worked? If the program were a success, what would have happened? What would have changed?

The next step, stated Dr. Elm, is to define the program by explaining the program theory. Explaining the program theory will include what the program is intended to accomplish, as well as how and why the program is expected to work. Dr. Elm recommended that the team complete the program theory in three parts: (a) defining the program's long-term goals, (b) delineating the program's strategies and activities, and (c) explaining how and why the team believes the program's activities and strategies will result in the desired outcomes.

Your program may have one or two goals, or your program may have many goals. For some programs, the primary goal may be to improve student learning. For others, primary goals might be to affect teacher content knowledge and teacher practice. Goals may have to do with behavior, safety, involvement, or attitudes. The first piece in explaining the program is to list the overall goals of your program or initiative. Goal statements should be broad and general and should reflect the overall intent of your program or a shared vision of what your program is supposed to accomplish. Objectives tend to be more specific and are often short term or intermediate term. If objectives are known, record them. However, at this point in program planning, broad goal statements are sufficient.

Based on their review of documentation and research as well as discussions and interviews with other districts that have implemented the program, and from meetings with district administration and school staff, Mrs. Anderson and the oversight team set the following long-term goals for READ:

1. Increased student engagement in reading
2. Improved student reading skills

Once you have documented what the program is intended to accomplish, the next component is to document your program's **strategies and activities**. How will the program accomplish these goals? What strategies will be used to achieve your goals? What activities will need to be put in place for the program? Does the program have activities that occur in the classroom, in another setting at school, at home, or in a combination of these settings?

The READ oversight team examined program materials to determine the primary components of the READ program. They determined that the READ program had three strategies: classroom lessons, homework, and assessments. Each of these strategies required certain activities in order to be successful. For instance, teachers would need professional development on how to integrate the READ classroom lessons into their instruction, as well as how to use the READ assessment data. Students would also need training in how to use the READ system in the classroom and at home.

Strategies might include activities such as professional development, technology access, and the use of curricular materials. Strategies might be ongoing throughout the program or drawn on at various stages during the program's operation. Listing all strategies and activities used in your program is important to explain later on how and to what extent your program's goals were met.

After careful review of the READ program and the district's particular program needs, the oversight team outlined the following primary strategies and activities for the READ program:

1. Interactive, standards-based classroom lessons (using the READ software with interactive classroom technologies and individual handheld mobile devices for each student).
2. Standards-based reading assessments (Internet-based, formative READ assessments of student reading skills administered using the READ software).
3. Standards-based reading homework (Internet-based using READ software).
4. Teacher professional development on integrating READ into classroom instruction (using an interactive wireless pad).
5. Teacher professional development on using READ assessment data for classroom lesson planning.
6. Student training on using READ (in the classroom and at home).

At this point, you have documented your program's goals and objectives, as well as the strategies and activities that will be conducted as part of the program to meet these goals. The next component is to **relate program strategies and activities to program goals**. Why should the program work? Why do you think implementing this set of strategies and activities will result in the goals you have set? The linkages between program strategies and program goals are assumptions as to why the program should work.

During a planning meeting focusing on why READ strategies and activities should result in the desired long-term goals, the oversight team brainstormed the underlying assumptions that were necessary for READ to work. The evaluator, Dr. Elm, facilitated the discussion among the oversight team members, leading them through the process of linking the program's activities and strategies to the long-term goals. Dr. Elm asked each member of the team to record *why* and *how* they thought each strategy or activity would lead to increased student engagement and improved student reading skills. Team members shared their reasoning with the group.

These underlying assumptions, taken together, are the basis of the program's theory. That is, **the program's theory is your theory as to why the program should work**. Perhaps you believe that employing a set of curricular materials and providing professional development to teachers in the use of these materials will result in improved differentiation of instruction and ultimately increased student learning. Or an assumption might be that having students respond to teacher questions using tablet computers or handheld devices will improve student engagement and participation as well as student learning. Documenting the relationship between your program's strategies and its goals explains your program design and is the basis for embedding evaluation into your program.

Dr. Elm led a discussion with the oversight team in which they examined each team member's ideas regarding why the program should work. Focusing on these ideas but not limited by them, the team members formulated, as a group, the underlying assumptions that were necessary to relate READ strategies and activities to long-term goals. During the discussion, team members were able to build on each other's ideas in order to construct a comprehensive theory that was supported by the group. As a result of their discussion, the team put forward seven assumptions forming the basis of READ's program theory.

The following seven assumptions form the basis of READ's program theory:

1. **Interactive, standards-based classroom lessons** (using READ software) will increase student interaction during learning, which will lead to increased exposure to standards-based learning opportunities.
2. **Standards-based reading assessments** (using READ software) will increase the availability of formative standards-based data on reading performance, which will lead to increased teacher use of standards-based reading assessment data and then improved differentiation of instruction.
3. **Standards-based reading homework** (using READ software) will increase student exposure to standards-based learning opportunities.
4. **Teacher training** on integrating READ into their classroom instruction will increase teacher use of READ, which will lead to improved integration of READ into classroom instruction. Teacher training on using READ assessment data for classroom lesson planning will increase teacher use of formative standards-based reading assessment data. Both will lead to improved differentiation of instruction.
5. **Student training** on using READ in the classroom will increase student interaction during learning. Student training on using READ at home will increase student use of READ at home. Both will lead to increased student exposure to standards-based learning opportunities.
6. Increased student interaction in the classroom and improved differentiation of instruction will result in **increased student engagement**.
7. Increased student exposure to standards-based learning opportunities, improved differentiation of instruction, and increased student engagement will result in **improved reading skills**.

Most programs rely upon certain **contextual conditions** being met and **resources** being readily available in order to operate the program. If your program assumes that a certain infrastructure is in place or that certain materials are available, you should identify and list these conditions and resources when planning your evaluation.

The oversight team also identified contextual conditions and resources that are necessary to the success of READ:

1. Program funding for READ, as well as necessary equipment to support infrastructure needs.
2. Program funding for external evaluation assistance.
3. Technology infrastructure at school:
    a. Classroom computer with Internet access
    b. Interactive technologies in each classroom
    c. Interactive, wireless pad for convenient, mobile teacher operation of computer
    d. 25 student handheld mobile devices per classroom for interactive learning
4. Availability of professional development for teachers on:
    a. Using interactive equipment in the classroom with the READ software; ongoing technical assistance from technology coordinator
    b. Integrating the READ software into their instruction
    c. Using READ assessment data for classroom lesson planning and differentiation of instruction
5. Availability of student training on how to use interactive equipment in the classroom, as well as how to use the READ software at home.
6. Student access to technology at home (computer with Internet connection).

You have now documented your program, including the strategies and activities that will be part of the program and the goals you hope to accomplish. You have also documented your assumptions as to why the strategies should result in achieving the program's goals. As mentioned in the previous paragraph, these assumptions explaining why the program should work are the basis of the program's theory. Before defining your program any further, this would be a good place to pause for a moment and reflect on the program design that you have

documented. Ask yourself again why you think your assumptions of the program should work. Are your assumptions based on a solid research foundation? That is, do you have reason to believe based on the results from past evaluations or research conducted by others that the program will work? Or are your assumptions based on emerging knowledge in the field or your own experience? Do you believe that your assumptions are based on strong evidence or are they just hypotheses?

Understanding the basis of the program's theory is important to designing a rigorous evaluation. Assessing program implementation should always be central to your evaluation design. However, the less evidence there is to support the program's theory, the more carefully you will want to monitor the implementation of your program and gather early and intermediate information of program effectiveness. If there is evidence from methodologically sound past evaluations that is contrary to your proposed theory, you will want to think carefully about what is different regarding your program to cause you to think it will work. In such cases, documenting alternative theories may prove useful to you in understanding and interpreting program results. It is important to note that there is nothing wrong with a sound, well-documented theory that has little existing information to support its effectiveness, as the information you obtain from your evaluation may be the foundation of innovation. See the section of **Appendix C** for more on program theory.

How Does the Program Work? (Creating the Logic Model)

You have completed the most important part of program design and evaluation; you have defined your program, documenting why your program should work. Next is the process of refining the program design and evaluation: How does the program work? Using the program's theory and underlying assumptions as the foundation, you will begin to create a model that depicts your program's inner workings.

What Is Logic Modeling?

A logic model lays out your program's theory by explaining how you believe your program works. Your logic model will set short-term and intermediate objectives that you can check throughout the evaluation to determine the extent to which your program is working as envisioned. Your logic model is the cornerstone of your program and its evaluation, and you should continually use it to check progress throughout the program, to help you discover problems with your program, and to make necessary corrections and improvements while your program is in operation.

> Logic modeling is a process, and the model created through the process will be the foundation of your program and its evaluation.

Logic modeling is a process, not simply an end result. While you will create a logic model through the process—a model that will be a critical component of your program's operation and evaluation—the power is in the process. The process of logic modeling has many uses, from designing a new project to fostering shared ownership of a plan to teaching others how a program is intended to work. We will touch on those uses that are important to evaluation. Additional resources are provided in the section of Appendix C if you would like to learn more.

Putting a new idea into practice is change, and change takes time. Logic modeling can facilitate change by building a shared vision and ownership among stakeholders from the outset, but only if creating the logic model is a shared process. This does not mean that you need to include every stakeholder in every phase of your logic modeling. The initial creation of your logic model works best if done by a small group. However, once this group creates a draft, including others in the process will likely improve your model and the program's subsequent implementation.

Including teachers in the logic modeling process can help to ensure that teachers are working toward a common goal and that all teachers understand and support what the program is trying to accomplish. Including parents can help to foster a culture in which parents understand and embrace what the teachers are trying to accomplish with their children, so parents can, in turn, support these efforts at home. Including students invites them to be active participants in the program planning and understanding process. Further, including administrators and school board members is critical to creating a shared understanding and mutual support of the program and its goals. Finally, the inclusion of stakeholders is not a one-time effort to garner support but rather an ongoing partnership to improve your program's design and operation.

> A logic model explains how you expect a program's strategies and activities to result in the program's stated goals and objectives.

The logic modeling process should include the person or people who will have primary responsibility for the program, as well as those who are critical to its success. Because the logic model you are creating will be used for evaluation purposes, your model will not simply describe your program or project, but it will also provide indicators that you will use to measure your program's success throughout its operation. For this reason, it would be helpful to ask someone with evaluation expertise to be part of your logic modeling group. Once you have your logic modeling team assembled, the following paragraphs will step you through the process of creating your model.

How Do I Create a Logic Model?

At the heart of your logic model are the linkages between what you do as part of your program and what you hope to accomplish with the program. The linkages explain how your program works, and they include your program's short-term and intermediate objectives. Short-term

and intermediate objectives are critical to improving the implementation of your program, as well as to establishing the association, supported by data, that your program's activities are theoretically related to your program's goals. Without short-term and intermediate indicators that reflect the program's underlying theory, your evaluation would be a black box with inputs (strategies and activities) and outputs (goals and objectives). The logic model is a depiction of the inside of the box, allowing you to monitor your program's operation and enabling you to make assertions about the success of the strategies that are part of your program.

If your program theory is well defined, you may find that creating the logic model is a breeze. If your program theory still needs more explanation of how your program should work, the process of creating your logic model will aid you in further refining it. Logic modeling is an opportunity to really think through the assumptions you laid out in your program's theory, to consider again what resources and supports you will need to implement your program effectively, and to lay out what you plan to achieve at various stages during your program's operation.

These are the primary components of a logic model:

1. Long-term objectives or outcome goals
2. Program strategies and activities
3. Early (short-term) objectives
4. Intermediate objectives
5. Contextual conditions

Your logic model will be a living model, in that the theory underlying your model and the indicators informing your model are not static but should be changed as your understanding changes. You will start with your program theory, and your logic model will represent this theory. However, as information is obtained through the program's implementation and evaluation, you will need to revise and improve the model so that it is always an accurate representation of your program. The logic model is your road map and should reflect your initial understanding of the program, as well as the knowledge you learn during your program's operation.

These are the primary components of a logic model, in order of development:

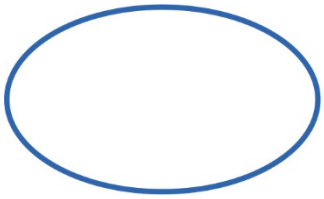
1. Defining **long-term objectives/outcome goals**.
2. Delineating **program strategies and activities**.
3. Detailing **early (short-term) objectives**.
4. Outlining **intermediate objectives**.
5. Listing necessary contextual conditions or resources (**context**).

While you may decide to depict your logic model using various shapes, in this guide:

- Strategies and activities will be denoted by rounded rectangles.

- Early (short-term) and intermediate objectives will be denoted by rectangles.

- Long-term goals will be represented by elongated ovals.

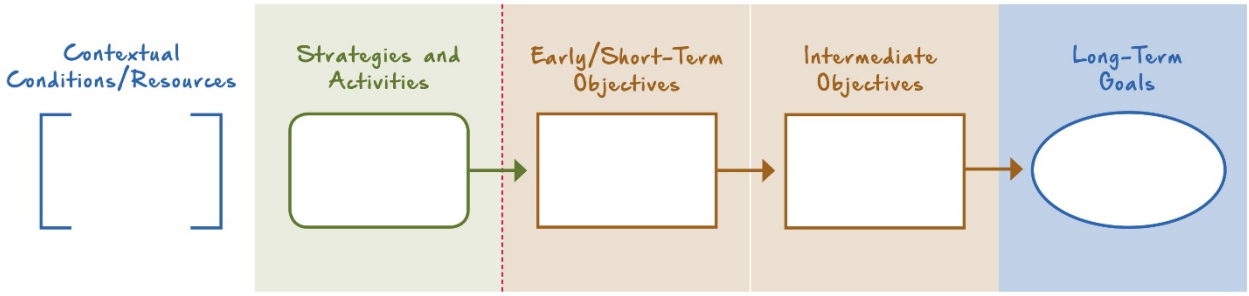


Remember that there is no magic to the shapes. You should use whatever shapes make the most sense to you! The substance is in the connections between your shapes, as these connections represent your program's theory.

Start by stating your long-term goals on the right-hand side of your logic model. Move to the left and give your intermediate and early or short-term objectives, followed by your strategies and activities on the left-hand side. Including contextual conditions and resources on your model is a helpful reminder of what needs to be in place for your program to operate. If you decide to add contextual conditions or resources to your model, you can list them on the far left-hand side of your model (before your strategies and activities).

The headings of your model might look like those in Figure 2.

Figure 2: Possible Logic Model Headings



Once you have listed your contextual conditions and necessary resources, strategies and activities, short-term and intermediate objectives, and long-term goals, it is time to translate your program's theory (set of assumptions) into your logic model. Think carefully about what needs to occur in the short term, intermediate, and long term. Map out your assumptions, carrying each strategy through to a long-term goal. Some strategies may share short-term and intermediate objectives, and some objectives may branch out to one or more other objectives. Check to be sure that all strategies ultimately reach a goal and that no short-term or intermediate objectives are dead-ends (meaning that they do not carry through to a long-term goal). Every piece of your model is put into place to achieve your long-term goals. As mentioned earlier, it is your road map, keeping you on track until you reach your destination. Seeing a fully completed logic model may be helpful at this point. Please refer to **Figure 3: READ Logic Model** in Appendix A (and reproduced on page 24), and **Figure 5: NowPLAN-T Logic Model** in Appendix B for examples.

> A logic model can be used to explain your program and its evaluation to others, as well as to track your program's progress.

Keep in mind that creating your logic model offers another opportunity for you to examine whether important activities are missing. Does it make sense that the program strategies and activities would result in your short-term and intermediate objectives and long-term goals for the program? Are additional strategies needed? Are some strategies more important than others? If so, note this in your program definition and theory. In addition, the logic modeling process can help you to refine your program's theory. As you think through the assumptions that link strategies and activities to goals, you may decide that the logic model needs more work and may want to include additional or different objectives. It is important to use the logic modeling process to reaffirm or refine your program's theory, as the model will be the basis of your program's design and evaluation. Your logic model will have many uses, including documenting your program, tracking your program's progress, and communicating your program's status and findings. As mentioned previously, your model can

also be used to foster a mutual understanding among your stakeholders of what your program looks like, as well as what you intend for the program to accomplish.

At this point in the evaluation design, Dr. Elm recommended that the READ oversight team create an evaluation subcommittee, named the E-Team, comprised of 3 to 5 members. The evaluation subcommittee was formed as a partnership and a liaison between the READ program staff and the external evaluator, and was tasked with helping to design the evaluation and with monitoring the evaluation findings shared by the READ external evaluator. Mrs. Anderson appointed two oversight committee members (the district reading coach and one of the district reading specialists) to the E-Team. She also asked the district supervisor for assessment and evaluation to serve on the E-Team and to be the primary internal contact for the READ external evaluator. Finally, she invited Dr. Elm to serve as the chair of the E-Team and to serve as the lead, external evaluator of the READ program. As the external evaluator, Dr. Elm would conduct the evaluation and share findings with the E-Team and oversight team. The four-member E-Team's first task was to create the READ logic model.

This guide touches on the basics of logic models. Logic models can be simple or quite sophisticated, and can represent small projects as well as large systems. If you would like to know more about logic models or logic modeling, a few good resources are included in the section in **Appendix C**.

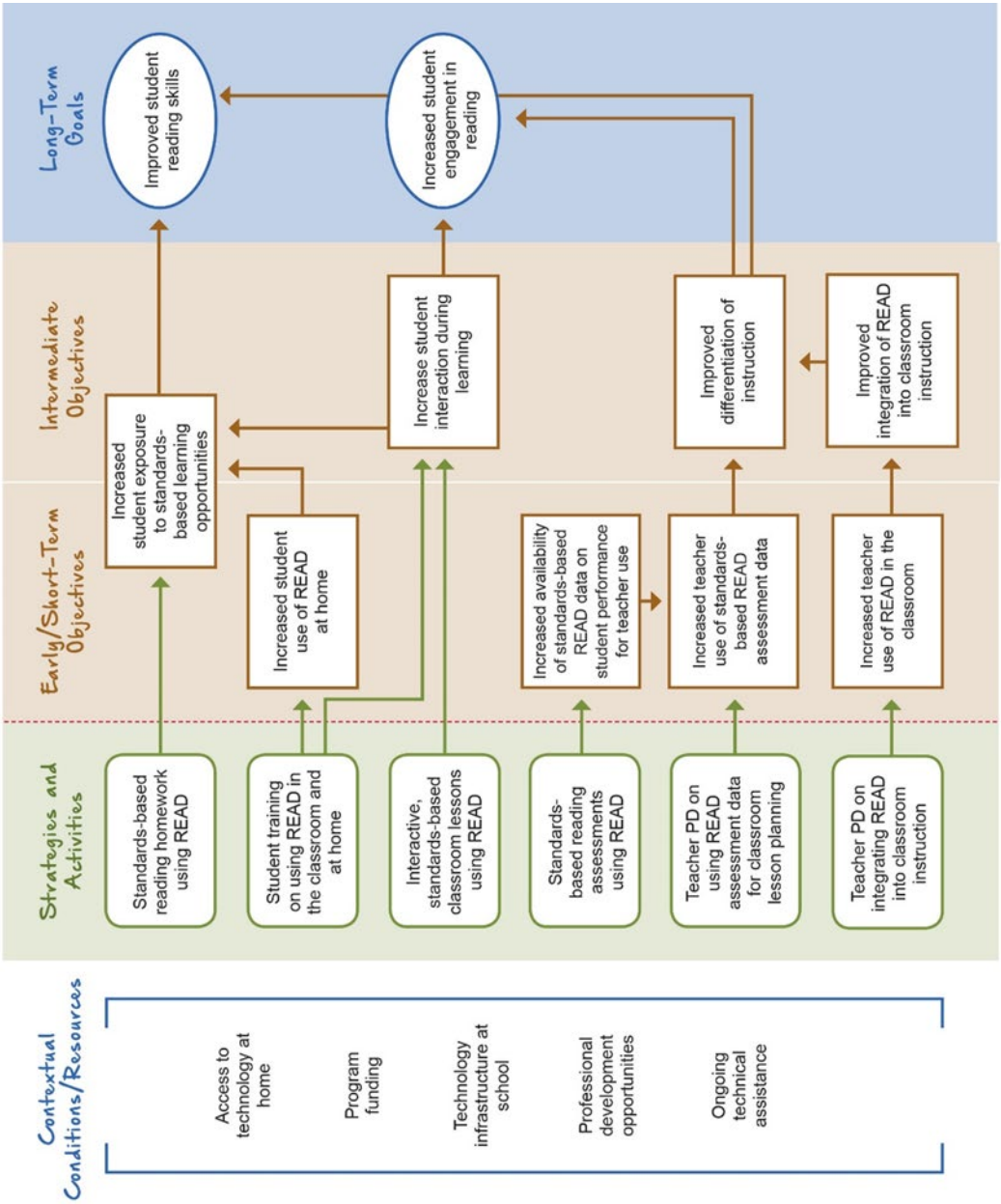
Why Is Understanding the Program Important?

As stated earlier, understanding your program by defining your program's theory is the most important step in program design and evaluation. The logic model that you create to depict your program's theory is the foundation of your program and your evaluation. Once you have a draft logic model, you can share the draft with key program stakeholders, such as the funding agency (whether it be the state education agency, the district, the school board, or an external foundation, corporation, or government entity), district staff, teachers, and parents. Talking through your model with stakeholders and asking for feedback and input can help you improve your model as well as foster a sense of responsibility and ownership for the program. While your program may be wonderful in theory, it will take people to make it work. The more key stakeholders you can substantively involve in the logic model development process and the more key people who truly understand how your program is intended to work, the more likely you will succeed.

Using the program definition developed by the oversight team, the E-Team worked to create a logic model. The E-Team started with the long-term goals on the right side of the model. The E-Team listed the contextual conditions and resources on the left. Just to the right of the context, the E-Team listed the strategies and activities. Next, the E-Team used the oversight team's assumptions to work through the early/short-term and intermediate objectives.

Finally, following and updating your logic model throughout your program's operation, as well as recording the degree to which early (short-term) and intermediate objectives have been met, enable you to examine the fidelity with which your program is carried out and to monitor program implementation. Logic modeling as an exercise can facilitate program understanding, while the resulting logic model can be a powerful tool to communicate your program's design and your program's results to stakeholders. Stakeholders, including the funding agency, will want to know the extent to which their resources – time and money – were used effectively to improve student outcomes.

This is a reduced size of the full logic model for the READ program. Appendix A provides the full-size logic model in **Figure 3: READ Logic Model.**



STEP 2: PLAN – How Do I Plan the Evaluation?

What Questions Should I Ask to Shape the Evaluation?

While many evaluations ill-advisedly begin with creating evaluation questions, the first step should always be understanding the program. How can you create important and informed evaluation questions until you have a solid understanding of the theory that underlies a program? Because you have already created a logic model during the process of understanding your program, generating your evaluation questions is a natural progression from the model.

Your evaluation questions should be open-ended. Avoid yes/no questions, as closed-ended responses limit the information you can obtain from your evaluation. Instead of asking "does my program work?" you might ask:

- To what extent does the program work?
- How does the program work?
- In what ways does the program work?
- For whom does the program work best?
- Under what conditions does the program work best?

Evaluation questions tend to fall into three categories taken from your logic model: measuring the implementation of strategies and activities, identifying the progress toward short-term and intermediate objectives, and recognizing the achievement of long-term program goals. The following paragraphs will lead you through a process and some questions to consider while creating your evaluation questions.

At the next READ planning meeting, the E-Team shared the draft logic model with the full oversight team. Oversight team members reviewed the model and felt comfortable that it represented the assumptions and logic as they had agreed on at their last meeting. No changes were needed to the logic model at this time. Next, the E-Team and the oversight team used the logic model to develop evaluation questions for the READ program.

Evaluating Implementation of Activities and Strategies

How do you know if your program contributed toward achieving (or not achieving) its goals if you do not examine the implementation of its activities and strategies? It is important for your evaluation questions to address the program's activities and strategies. Education does not take place in a controlled laboratory but rather in real-world settings, which require that you justify why you believe the program strategies resulted in the measured outcomes. Your program's underlying theory, represented by your logic model, shows the linkages between the strategies and activities and the goals. The evaluation of your program's operation will set the stage to test your theory. And more importantly, asking evaluation questions about how your strategies and activities were applied can tell you the degree to which your program had the opportunity to be successful.

> Use your logic model to guide you as you create your evaluation questions.
>
> Your questions regarding strategies and activities address the degree to which your program had the opportunity to be successful. Questions in this category may also address contextual conditions and resources.
>
> Questions addressing your early and intermediate objectives are important in determining if your program is on track toward meeting its long-term goals.

It is never a good idea to measure outcomes before assessing implementation. If you find down the road that your long-term goals were not met, is it because the program did not work or because key components of it were not applied properly or at all? Suppose you find that your long-term goals were successfully met. Do you have enough information to support that your program contributed to this success? It is a waste of resources to expend valuable time and money evaluating program outcomes if important program components were never put into place. While you will likely want to create evaluation questions that are specific to your program's activities and strategies, a fundamental evaluation question at this stage is: What is the fidelity with which program activities have been implemented?

> Using each of the strategies and activities listed on the left-hand side of the logic model, the E-Team worked with the READ oversight team to develop evaluation questions. For each strategy or activity, they developed questions addressing whether the strategy or activity had been carried out, as well as questions addressing some contextual conditions and resources necessary for program implementation.

The READ E-Team and oversight team created six **evaluation questions** to assess READ **strategies and activities**:

Strategies and Activities	Evaluation Questions
Interactive, standards-based classroom lessons (using the READ software with interactive classroom technologies and individual handheld mobile devices for each student).	To what extent did teachers have access to the necessary technology in the classroom to use READ in their instruction?
Standards-based reading assessments (Internet-based, formative assessments of student reading skills administered within the READ software).	To what extent were READ assessments made available to students and teachers? Examine overall, by school, and by grade level.
Standards-based reading homework (Internet-based using READ software).	To what extent did students have access to READ at home? Examine overall and by grade level, race, gender, and socioeconomic status.
Teacher professional development on integrating READ into classroom instruction (using an interactive wireless pad).	To what extent did teachers receive professional development on how to integrate READ into their classroom instruction?
Teacher professional development on using READ assessment data for classroom lesson planning.	To what extent did teachers receive professional development on how to incorporate READ assessment data into their classroom lesson planning?
Student training on using READ (in the classroom and at home).	To what extent were students trained in how to use READ?

Note: These questions are intended to evaluate the degree to which the program had the opportunity to be successful, as well as to determine if additional program supports are needed for successful implementation.

Evaluating Progress Toward Short-Term and Intermediate Objectives

Evaluating your program's opportunity to be successful is the initial step toward determining your program's success. The second category of evaluation questions will address how your program is working. That is, how do you know if your program is on track to meeting its long-term goals? Measuring progress toward short-term and intermediate objectives plays a significant role in determining how your program is working. By examining progress, you can

catch early problems with the program and remediate them before they become critical impediments to your program's success. Program staff can use interim evaluation findings to plan, shape, and improve the program prior to the evaluation of final outcomes. It is much easier and more cost-effective to uncover problems or issues early in your program's implementation. Your evaluation should strive to provide program staff with the necessary information for them to be able to understand the degree to which the program is on course so that they can make midcourse adjustments and refinements as needed.

> While you created your logic model right to left (starting with your long-term goals), it is often easier to craft your evaluation questions left to right. Begin with your early (short-term) objectives and work your way toward your intermediate objectives, and then long-term goals.
>
> There does not need to be a one-to-one correspondence between objectives and evaluation questions. Some evaluation questions may address more than one objective, while some objectives may have more than one evaluation question.
>
> At this point in your evaluation design, it is important to brainstorm evaluation questions and not be hindered by resource concerns. Prioritizing your questions will come later. When prioritizing your evaluation questions, you will decide based on resource constraints and feasibility which questions your evaluation can adequately address.

Your evaluation questions at this stage should focus on your program's specific short-term and intermediate objectives. However, an overarching evaluation question at this stage might be: To what extent is the program on track to achieving long-term goals? Use your logic model to guide you as you create your evaluation questions pertaining to early and intermediate objectives, just as you used the strategies and activities from your logic model to create your first set of evaluation questions.

While you created your logic model right to left (starting with your long-term goals), it is often easier to craft your evaluation questions left to right. Begin with your early (short-term) objectives and work your way toward your intermediate objectives, and then long-term goals. Some evaluation questions may address more than one objective, while some objectives may have more than one evaluation question. That is, there does not need to be a one-to-one correspondence between objectives on your logic model and evaluation questions. However, you should have at least one evaluation question that addresses each objective. Later, the evaluation team can prioritize evaluation questions. In doing this, it is possible that you will decide, based on your priorities and resource constraints, not to address certain questions and objectives in your evaluation.

Next, the E-Team worked with the READ oversight team to create several evaluation questions addressing READ **early/short-term** objectives and **intermediate** objectives:

Early/Short-Term and Intermediate Objectives	**Evaluation Questions**
Increased student use of READ at home (early/short-term).	How often did students receive READ homework assignments? To what extent did students complete READ homework assignments? **Note frequency and duration of use.
Increased teacher use of READ in the classroom (early/short-term).	In what ways and how often did teachers use READ in the classroom with students? **Note frequency, duration, and nature of use.
Increased student exposure to standards-based learning opportunities (early/short-term).	To what extent did students complete READ homework assignments? How often did teachers use READ in the classroom with students?
Increased availability of standards-based, formative READ assessment data on student reading performance (early/short-term).	How often did teachers access READ student assessment data? **Note frequency and type of access.
Increased teacher use of standards-based READ assessment data (early/short-term).	In what ways did teachers use READ student assessment data?
Increased student interaction during learning (intermediate).	To what extent and how did students interact during classroom instruction when READ was used? **Note frequency and type of interaction.
Improved integration of READ into classroom instruction (intermediate).	In what ways and to what extent did teachers integrate READ into their classroom instruction? **Note the quality with which READ was integrated into classroom instruction by teachers.
Improved differentiation of instruction (intermediate).	In what ways and to what extent did teachers use READ assessment data to plan and differentiate instruction? **Note what data were used and how data were used in instructional planning.

Evaluating Progress Toward Long-Term Goals

Finally, a third set of evaluation questions should focus on the program's long-term goals. While evaluation findings at this stage in your program's operation can still be used to improve the program's operation, assessment of long-term goals is typically used for summative decision-making. That is, results from the measurement of progress toward long-term goals are often used to make decisions about whether program funding should be extended and if a program should be continued, expanded, scaled down, or discontinued. Your questions will be specific to your program's goals, though they should address the following: To what extent does the program work? For whom does the program work best? Under what conditions does the program work best?

Finally, the E-Team and the READ oversight team created **evaluation questions** addressing READ **long-term goals**:

Long-Term Goals	Evaluation Questions
Increased student engagement in reading.	To what extent and in what ways did READ foster student engagement during reading lessons?
Improved student reading skills.	To what extent did READ improve student learning in reading? • To what extent did student learning improve after READ was implemented? • To what extent did learning outcomes vary with teacher use of READ in the classroom? • To what extent did learning outcomes vary with teacher use of READ assessment data to plan and differentiate instruction? • How did student performance on the READ assessments correlate with student performance on state assessments? • In what ways did learning outcomes vary by initial reading performance on state assessments? • In what ways did learning outcomes vary by grade level? • In what ways did learning outcomes vary by special education status and English language proficiency? • In what ways did learning outcomes vary with the frequency of READ use at home?

If you do not have the resources to focus on all of your evaluation questions, you may need to prioritize. When prioritizing evaluation questions, it is important to have at least some measurement in all three categories: implementation, short-term/intermediate objectives, and long-term goals.

What Data Should I Collect?

Now that you have developed your logic model and decided on your evaluation questions, the next task is to plan how you will answer those questions. Your logic model is your road map during this process. Just as you used the key components of your logic model as a guide to develop your evaluation questions, your evaluation questions will drive the data that will be collected through your evaluation.

The answers to your evaluation questions will give you the information you need to know in order to improve your program and to make critical program decisions. The following paragraphs will take you through the process of creating indicators for your evaluation questions that relate to program strategies and activities, short-term objectives, intermediate objectives, and long-term goals. Your indicators will dictate what data you should collect to answer your evaluation questions.

Indicators are statements that can be used to gauge progress toward program goals and objectives. An **indicator** is a guide that lets you know if you are moving in the right direction. Your indicators will be derived from your evaluation questions; for some evaluation questions, you might have multiple indicators. Indicators are the metrics that will be tied to targets or benchmarks, against which to measure the performance of your program.

Indicators and targets should be specific, measurable, agreed upon, realistic, and time-bound (SMART). For instance, suppose you are evaluating a teacher recruitment and retention program. You may have an

Indicators can be derived from evaluation questions and are used to measure progress toward program goals and objectives. An evaluation question may have one or more indicators.

Targets provide a realistic time line and yardstick for your indicators. Indicators and targets should have the following characteristics:

An indicator is SMA:

- Specific
- Measurable
- Agreed upon

And a target is RT:

- Realistic
- Time-bound

Together, indicators and targets are SMART.

objective on your logic model that states “to increase the number of highly qualified teachers in our school district” and a corresponding evaluation question that asks “to what extent was the number of highly qualified teachers increased in our school district?” However, we know there are several ways that a “highly qualified teacher” can be defined, such as by certifications, education, content knowledge, etc. The indicator would specify the definition(s) that the evaluator chooses to use and the data element(s) that will be collected. For example, to be specific and measurable, the indicator might be twofold: “increasing number and percentage of teachers who are state certified” and “increasing number and percentage of teachers who hold National Board certification.” At this point, it would also be wise to consider whether the indicators you choose are not only measurable, but also available, as well as agreed upon by the evaluation team and program staff.

Next, clarify your indicator by agreeing upon a realistic and time-bound target. Thus, a **target** is a clarification of an indicator. A target provides a yardstick and time line for your indicator, specifying how much progress should be made and by when in order to determine to what extent goals and objectives have been met. Targets for the above example might include: “by 2015, all teachers in the school district will be state certified” and “by 2018, 50 percent of district teachers will have National Board certification.” For some programs, it is possible that reasonable targets cannot be set prior to the program’s operation. For instance, consider a program that is intended to improve writing skills for seventh graders, and the chosen indicator is a student’s score on a particular writing assessment. However, the evaluation team would like to see baseline scores for students prior to setting their target. In this case, a pretest may be given at the start of the program and, once baseline scores are known, targets can be determined.

With the evaluation questions that the READ oversight team and E-Team had created, the E-Team was ready to expand on each with indicators and accompanying targets. Using the logic model as its guide, the E-Team created an evaluation matrix detailing the logic model component, associated evaluation questions, indicators, and accompanying targets. Two examples are provided below. All indicators for the READ project are provided in Appendix A.

1. To what extent were READ assessments made available to students and teachers? (activity)
   Indicator: Increased number of students and teachers with access to READ assessments.
   Targets: By the start of the school year, all teacher accounts will have been set up in READ. By the end of September, all student accounts will have been set up in READ.
2. In what ways and to what extent did teachers integrate READ into their classroom instruction? (intermediate objective)
   Indicator: Improved integration of READ lessons into classroom instruction, as measured by teacher scores on the READ implementation rubric (rubric completed through classroom observations and teacher interviews).
   Targets: By April, 50% of teachers will score a 3 or above (out of 4) on the READ implementation rubric. By June, 75% of teachers will score a 3 or above and 25% of teachers will score a 4 on the READ implementation rubric.

Evaluation Matrix

Now that you have created evaluation questions with accompanying indicators and targets for each component of your logic model, how do you organize that information into a usable format for your evaluation? One method is to use an **evaluation matrix**. An evaluation matrix represents your logic model components, evaluation questions, indicators, and targets by your logic model strategies and activities, early and intermediate objectives, and long-term goals. Table 1 shows an example shell. **Table 26: Evaluation Matrix Template** is provided in Appendix E. Information for completing the data source, data collection, and data analysis columns will be covered next in the guide.

Table 1: Evaluation Matrix Example Shell

	Logic Model Component	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
Strategies and Activities/ Initial Implementation							
Early/Short-term and Intermediate Objectives							
Long-term Goals							

In an effort to organize their logic model and associated information, the E-Team created an Evaluation Matrix. At this stage, the Evaluation Matrix included the READ logic model components, evaluation questions, indicators, and targets by the READ logic model strategies, early/short-term and intermediate objectives, and long-term goals. A copy of the READ Evaluation Matrix starts at **Table 7: Evaluation Matrix Addressing Strategies and Activities During the Initial Implementation—Indicators and Targets**.

How Should I Design the Evaluation?

Evaluation Design

You are most likely evaluating your program because you want to know to what extent it works, under what conditions or with what supports it works, for which students it works best, and how to improve it. You have spent the last step defining the program and what you mean when you say it "works." A strong evaluation design can help you to rule out other plausible explanations as to why your program may or may not have met the expectations you set through your indicators and targets. How many programs are continued with little examination of how they are benefiting the students? How often do we "experiment" in education by putting a new program into the classroom without following up to see if there was any benefit

(much less, any adverse effect)? When do we make our decisions based on data, and how often do we accept anecdotal stories or simple descriptions of use as though they were evidence of effectiveness (because we have nothing else on which to base our decisions)? Evaluation can provide us with the necessary information to make sound decisions regarding the methods and tools we use to educate our students.

Evaluation should be built into your program so you can continually monitor and improve your program—and so you know whether students are benefiting (or not). Your evaluation also should help you determine the extent to which your program influenced your results. Suppose you are evaluating a mathematics program and your results show that student scores in mathematics, on average, increased twofold after your program was put into place. But upon further investigation, you find that half of the students had never used the program, and that the students who used the program in fact had much lower scores than those who did not. What if you had not investigated? This program may have been hindering, rather than helping, student learning.

> Common evaluation designs include:
>
> - single-group designs
> - comparison group designs
> - randomized controlled experiments
>
> Strong comparison group designs are often referred to as *quasi-experimental designs*.
>
> Randomized controlled experiments are also called *true experiments* or *randomized controlled trials* (RCTs).

The questions and example in the above paragraph are intended to show that while evaluation is important, it is a good evaluation (one that gives you valid information as to how your program is working) that really matters. Evaluation relies on attribution. And, the more directly you can attribute your evaluation findings to the program activities you implemented, the more meaningful your findings will be—and the more useful your findings will be to you as you work to improve your program.

Some evaluation designs provide you with stronger evidence of causality than others. So, how do you choose the strongest possible design and methods to answer your evaluation questions, taking into account any constraints that you may have? This will partly depend upon the extent to which you have control over your implementation setting and other, similar settings.

Single-Group Designs

If you are implementing a project in only one of the schools in your district, your evaluation may focus on a single group—one school. In a **single-group design**, one group participates in the program and that same group is evaluated. While a single-group design is the simplest

evaluation design, it is also the weakest evaluation design. This is because there may be many competing explanations as to why your evaluation produced the results it did. If your evaluation showed promising results, could it be because of something else that was going on at the same time? Or perhaps the participants would have had the same results without the program?

Using your logic model along with the single-group design can help to improve the credibility of your findings. For instance, suppose you are working with an evaluator to examine a new program in your classroom or school focused on improving reading comprehension among third graders. If the evaluation results are promising, the principal has agreed to incorporate the funding for the program into the ongoing budget. If you do not have another classroom or school against which to compare progress (i.e., you have a single-group design), you can explain how the program operates by using your logic model and the data collected at each stage of operation. You can give evidence showing that the program's activities were put into place, use data from your early and intermediate objectives to show change in teacher practice and student progress, and present your long-term outcomes showing how reading comprehension changed. While you cannot claim that your program caused the change in reading comprehension, you can use your logic model and its associated indicators to demonstrate a theoretical association between your program and long-term outcomes.

> Using your logic model to guide your evaluation will strengthen your evaluation design.
>
> Other ways to strengthen your design include:
>
> - measuring indicators multiple times,
> - sampling, and
> - studying the program longitudinally.

Comparison Group Designs

If you are able to have more than one group participate in your evaluation, typically you can improve the usability of your findings. For instance, one teacher could use the program in the classroom in one year, and another the next year—and you could compare the results not only within the evaluation classroom from one year to the next, but also between the two classrooms in the evaluation. Using multiple groups, referred to as a **comparison group design**, can help you rule out some of the other competing explanations as to why your program may have worked. The comparison group is the group that does not use the program being evaluated. However, the groups must be comparable. Comparing test scores from a district that used a new program to test scores from another district that did not use the new program would not yield meaningful information if the two districts are not comparable.

The strength of your evaluation design will vary with how closely matched your comparison group is with the group that will be implementing your program. Convenience groups, such as a district chosen because it neighbors your district, will likely not yield results that are as meaningful as would a comparison district that is purposefully chosen to match your school district based on multiple key indicators that you believe might influence your outcomes, such as gender, ethnic and socioeconomic composition, or past test performance.

Just because a good comparison group does not readily exist for your program, do not give up on the possibility of finding or creating one. Use some creativity when creating your evaluation design and identifying comparison groups. If you are implementing a project across a district, you may have flexibility such that you could vary the timing of the implementation in order to create a comparison group. For instance, several schools could implement the program in one year, leaving the remaining schools as a comparison group. If your evaluation results are promising, the comparison schools can be brought on board in the following year.

Strong comparison group designs are often referred to as *quasi-experimental designs*. When considering a comparison group, seek to identify or create a group that is as similar as possible, especially on the key indicators that you believe might influence your results, to the group that will be implementing your program. However, the only way to make certain, to the extent possible, that groups are equivalent is through random assignment. Random assignment is discussed in the following section.

Experimental Designs

The gold standard of evaluation design is the *true experiment*. Comparison group designs, discussed in the above paragraph, attempt to approximate, to the extent possible, a true experiment. In an **experimental design**, participants are randomly assigned to the program or to a nonprogram control group. True experiments are also referred to as *randomized controlled experiments* or *randomized controlled trials* (RCTs).

In a true experiment, participants are randomly assigned to either participate in the program or an alternative condition (such as a different program or no program at all). Theoretically, the process of random assignment creates groups that are equivalent across both observable and unobservable characteristics. By randomly assigning program participants, you can rule out other explanations for and validity threats to your evaluation findings. See the **Research and Evaluation Design, Including Reliability and Validity** section and the **Threats to Validity** section in **Appendix C** for resources addressing random assignment and threats to validity.

For some programs, random assignment may align well with program resources. For instance, for programs that do not have the resources to include all students from the start, randomly assigning students or classrooms to the program would address in a fair manner who participates in the program and would allow you to draw causal conclusions from your

evaluation findings. As mentioned in the section on comparison groups, be creative when designing your evaluation. You might find that, with a little resourcefulness at the design stage, you can implement a stronger evaluation than you originally thought. For example, instead of purposefully assigning students or teachers to the program or allowing participants to self-select, you might consider a lottery at the start to determine who will participate. In such cases, if results are promising for the first cohort who participates, additional resources could be sought to expand the program to all students and classrooms.

Enriching Your Evaluation Design

Whether you have chosen to evaluate using a single-group, comparison-group, or experimental design, there are several methods and approaches you can use to **enrich your evaluation design**. Such methods are added supports in your evaluation design that can increase the usefulness of your results and credibility of your findings, make your evaluation more manageable, and expand upon information obtained throughout program implementation. These methods include using repeated measures, longitudinal data, and sampling. Logic modeling too can enrich your evaluation, as it can be used to construct a reasoned explanation for your evaluation findings. Supplementing your evaluation design with a case study could also enrich your evaluation design by providing in-depth information regarding implementation and participant experiences.

Using **repeated measures**, collecting the same data elements at multiple time points, can also help to strengthen your evaluation design. If the program you are evaluating is intended to improve critical thinking skills over a specified time period (e.g., 1 year), taking repeated measurements (perhaps monthly) of indicators that address critical thinking skills will not only provide you with baseline and frequent data with which to compare end-of-year results, but will also enable program staff to use midterm information in order to make midcourse corrections and improvements.

Using **longitudinal data**, data collected over an extended period of time, can enable you to follow program participants long-term and examine post-program changes. Longitudinal data can also enable you to examine a program's success using a time series analysis. For example, suppose your district made the change from half-day to full-day kindergarten 5 years ago, and you are asked whether the program positively affected student learning. The district has been using the same reading assessment for kindergarteners for the past 10 years. The assessment is given in September and May of each year. You examine the September scores over the past 10 years and find that there has been little variability in mean scores. Mean scores by gender, ethnicity, and English Language Learner (ELL) status have been fairly steady. You conclude the kindergarteners have been entering school at approximately the same mean reading level for the past 10 years.

Next, you examine the May reading scores for the past 10 years. You notice that for the first 5 years, the mean end-of-year scores (overall and by subgroup) were significantly greater than the September scores, but varied little from year to year. However, for the past 5 years, the May scores were about 15 percent higher than in the previous 5 years. The increase by gender and ethnicity was similar and also consistent over the past 5 years, while reading scores for ELL students were over 30 percent higher in the spring, after the full-day program was instituted. After ruling out other possible explanations for the findings to the extent possible, you conclude that the full-day kindergarten program appears to have been beneficial for all students and, particularly, for the district's ELL students.

If your program has many program participants or if you lack the funds to use all of your participants in your evaluation, **sampling** to choose a smaller group from the larger population of program participants is an option. Random sampling selects evaluation participants randomly from the larger group of program participants, and may be more easily accepted by teachers, parents, and students, as well as other stakeholder groups. Whether you are using random sampling or purposeful sampling, you should select a sample group that is as representative as possible of all of your participants (i.e., the population). Typically, the larger the sample you use, the more precise and credible your results will be. For more information on **Research and Evaluation Design, Including Reliability and Validity** and **Threats to Validity**, see Appendix C.

Using **logic modeling** in your evaluation can also help to strengthen the credibility of your findings. By examining the implementation of your strategies and activities as well as the measurement of progress on your early, intermediate, and long-term indicators, your logic model can provide you with interim data that can be used to adjust and improve your program during its operation. As described with the reading comprehension example in the single-group design section, logic modeling can help to show a theoretical association between the strategies and outcomes in even the weakest of evaluation designs.

Finally, **case studies** are in-depth examinations of a person, group of people, or context. Case studies can enrich your understanding of a program, as well as provide a more

Evaluation reporting should be ongoing. While formal evaluation data and reports may be issued once or twice a year, informal updates should be provided to program staff on a regular basis.

It is a good idea to explicitly identify (on your logic model) when evaluation updates will occur and to delineate these evaluation milestones in the evaluation contract with your evaluator.

Frequent and ongoing evaluation updates give a program the best opportunity to monitor and improve implementation, as well as to help maintain stakeholder support.

accurate picture of how a program operates. See the **Evaluation Methods and Tools** section for more information on case studies.

Building Reporting into Your Evaluation Design

You do not need to wait until the end of the evaluation to examine your goals. In fact, you should *not* wait until the end! Just as our teachers always told us that our grades should not be a surprise, your evaluation findings should not be a surprise. You should build reporting into your evaluation design from the very start.

It works well to align your evaluation's schedule with your program's time line. If you aim to have your program infrastructure in place by the end of summer, monitor your logic model indicators that address this activity prior to the end of the summer (to verify that the program is on track), and again at the end of summer (or early fall). If the infrastructure is not in place on schedule or if it is not properly operating, program staff need to know right away to minimize delays in program implementation (and so you do not waste time measuring intermediate indicators when early indicators tell you the program is not in place). Likewise, do not wait until the end of the year to observe classrooms to determine how the program is used. Frequent and routine observations will provide program staff with valuable information from which they can determine whether additional professional development or resources are needed.

Grovemont School District had 80 third- through fifth-grade classrooms across six elementary schools (28 third-grade classrooms, 28 fourth-grade classrooms, and 24 fifth-grade classrooms). District class size for grades three through five ranged from 22 to 25 students per classroom. Because of state budget cuts and reduced funding for the program, the E-Team knew that Mrs. Anderson and the READ oversight team would have to make some difficult choices about how to structure and evaluate their program.

Some members of the oversight team wanted to implement the program in fifth grade only for the first year, and then reexamine funds to see if they might be able to expand down to fourth grade in Year 2. Others voted to start the program at two of the six elementary schools and then try to include an additional school in Year 2.

Dr. Elm and the E-Team recommended that they consider partially implementing the program at all six schools and across all three grades. Dr. Elm explained that they would receive much better information about how their program was working and, more importantly, how it could be improved, if they were able to compare results from those classrooms that were using the program with those that were not. Dr. Elm knew that students at all of the schools in Grovemont School District were randomly assigned to teachers during the summer before each school year. However, Dr. Elm explained that in order to minimize initial differences between those classrooms that participate in READ and those that do not, they should consider randomly assigning half of the classrooms to continue with the existing district curriculum while the other half would supplement their existing curriculum with the READ program.

Dr. Elm also recommended that they first divide the classrooms by school and grade level so that each school and grade would have one half of the classrooms assigned to the program. Teachers whose classrooms were not assigned to the program would be assured that if the program proved successful, they would be on board by Year 3. However, if the program did not have sufficient benefits for the students, it would be discontinued in all classrooms after Year 2. Dr. Elm concluded that building a strong evaluation into their program would provide them with credible information as to how their program was working and that having data to direct their program adjustments and improvements would give the program the best opportunity to be successful.

The READ oversight team agreed to think about this idea and reconvene in 1 week to make a decision. The E-Team also distributed the evaluation matrix it had created based on the READ logic model. The E-Team asked the oversight team to review the matrix and provide any feedback or comments.

The following week, the E-Team and READ oversight team reconvened to decide how to structure the program and to work on the evaluation design. Mrs. Anderson had spoken with the district superintendent about the evaluator's suggestion of implementing READ in half the district's third- through fifth-grade classrooms, with the promise that it would be expanded to all classrooms in Year 3 if the program was successful. Although logistically it would be easier to implement the program in two or three schools or one or two grades than to implement it in half the classrooms in all schools and at all grades, the superintendent understood the benefit of the added effort. The evaluation would provide higher quality data to inform decisions for program improvement and decisions regarding the program's future.

Mrs. Anderson shared the superintendent's comments with the oversight team and evaluation subcommittee. Like the superintendent, team members felt conflicted by the choice between simpler logistics or a stronger evaluation design. Dr. Elm understood the dilemma all too well, but as an evaluator and an educator, she believed that a strong evaluation would result in improved program implementation and improved program outcomes.

Dr. Elm recognized that implementing the program in all classrooms in one grade level across the district would offer the weakest evaluation design and the least useful information but would likely be the simplest option logistically. Another option would be to start the program in all classrooms at two or three schools. In such a case, the other schools could be used as comparisons. For this reason, Dr. Elm explored the comparability of the six elementary schools in case the team decided to go that route. Five of the elementary schools had somewhat comparable state test scores in reading, while the sixth school had lower state test scores, and the difference was statistically significant. In addition, schools one through five had similar (and fairly homogenous) populations, while school six had a much lower socioeconomic student population and a much higher percentage of ELL students. Because the district was interested in how the program worked with ELL students, the team knew that the evaluation needed to include school six. However, if school six were used in a three-school implementation, the team would not have a comparable school against which to benchmark its results.

While not the simplest option, the oversight team decided that its best option would be to structure the program in such a way as to maximize the quality of the information from the evaluation. The team chose to build a strong evaluation into the READ program design to provide the formative information needed for program improvement and valid summative information for accountability.

Follow progress on your logic model indicators carefully along the way, so you continually know how your program is doing and where it should be modified. And when the time does come to examine results in terms of your long-term goals, your logic model is critical to explaining your findings. While you may not be able to rule out all competing explanations for your results, you can provide a plausible explanation based on your program's logic that your program activities are theoretically related to your program findings.

Finally, as mentioned above, the strength of your evaluation design, or the **design rigor**, directly impacts the degree to which your evaluation can provide the program with valid ongoing information on implementation and long-term goals regarding the success of the program. A strong evaluation design is one that is built to provide credible information for program improvement, as well as to rule out competing explanations for your summative findings. A strong evaluation design coupled with positive findings is what you might hope for, but even a strong evaluation that provides findings showing dismal results from a program provides valuable and important information. Evaluation results that help you to discontinue programs that do not work are just as valuable as findings that enable you to continue and build upon those programs that do improve student outcomes.

Evaluation Methods and Tools

You have almost completed your evaluation design. The most difficult part is over—you have defined your program and built your evaluation into your program's logic model. Using your logic model as a road map, you have created evaluation questions and their related indicators. You have decided how your evaluation will be designed. Now, **how will you collect your data?** You may have thought about this during the discussion on creating indicators and setting targets. After reading through the following paragraphs on methods that you might use in your evaluation, revisit your indicators to clarify and refine the methods you will use to measure each indicator.

Based on the READ oversight team's decision about how to structure the program, Dr. Elm and the E-Team drafted the following evaluation design. They presented the design at the next oversight team meeting. The oversight team voted to approve the design as follows:

Design: Multiple-group, experimental design (students randomly assigned to classrooms by the school prior to the start of the school year and classrooms randomly assigned to the READ program group or a non-READ comparison group)

Program group (READ): 40 classrooms (22 to 25 students per classroom)

Comparison group (non-READ): 40 classrooms (22 to 25 students per classroom)

Classrooms will be stratified by grade level within a school and randomly assigned to either the READ program group or a comparison group. The READ and non-READ groups will each include 14 third-grade classrooms, 14 fourth-grade classrooms, and 12 fifth-grade classrooms.

Enriching the evaluation design: Program theory and logic modeling will be used to examine program implementation as well as short-term, intermediate, and long-term outcomes.

Although there are many evaluation methods, most are classified as qualitative, quantitative, or both. **Qualitative methods** rely primarily on noncategorical, free responses or narrative descriptions of a program, collected through methods such as open-ended survey items, interviews, or observations. **Quantitative methods**, on the other hand, rely primarily on discrete categories, such as counts, numbers, and multiple-choice responses. Qualitative and quantitative methods reinforce each other in an evaluation, as qualitative data can help to describe, illuminate, and provide a depth of understanding to quantitative findings. For this reason, you may want to choose an evaluation design that includes a combination of both qualitative and quantitative methods, commonly referred to as **mixed-method**. Some common evaluation methods are listed below and include assessments and tests; surveys and questionnaires; interviews and focus groups; observations; existing data; portfolios; and case studies. Rubrics are also included as an evaluation tool that is often used to score, categorize, or code interviews, observations, portfolios, qualitative assessments, and case studies.

Assessments and tests (typically quantitative but can include qualitative items) are often used prior to program implementation (pre) and again at program completion (post), or at various times during program implementation, to assess program progress and results. Results of assessments are usually objective, and multiple items can be used in combination to create a

Reliability and validity are important considerations when selecting and using instruments such as assessments and tests (as well surveys and questionnaires).

Reliability is the consistency with which an instrument assesses (whatever it assesses). Reliability may refer to any of the following elements:

- The extent to which a respondent gives consistent responses to multiple items that are asking basically the same question in different ways (internal consistency reliability).
- The extent to which individuals' scores are consistent if given the same assessment a short time later (test-retest reliability).
- The extent to which different raters give consistent scores for the same open-ended response or different observers using an observation protocol give consistent scores for the same observation (inter-rater reliability).

(See next page for information on validity.)

subscale, often providing a more reliable estimate than any single item. If your program is intended to improve learning outcomes, you will likely want to use either an existing state or district assessment or choose an assessment of your own to measure change in student learning. However, before using assessment or test data, you should be sure that the assessment adequately addresses what you hope your program achieves. You would not want the success or failure of your program to be determined by an assessment that does not validly measure what your program is intended to achieve.

Surveys and questionnaires (typically quantitative but can include qualitative items) are often used to collect information from large numbers of respondents. They can be administered online, on paper, in person, or over the phone. In order for surveys to provide useful information, the questions must be worded clearly and succinctly. Survey items can be open-ended or closed-ended.

Open-ended survey items allow respondents to provide free-form responses to questions and are typically scored using a **rubric**. Closed-ended items give the respondent a choice of responses, often on a scale from 1 to 4 or 1 to 5. Surveys can be quickly administered, are usually easy to analyze, and can be adapted to fit specific situations.

Validity refers to how well an instrument measures what it is supposed to or is claims to measure. An assessment is not simply valid or not valid but rather valid for a certain purpose with a certain population. In fact, the same assessment may be valid for one group but not for another. For example, a reading test administered in English may be valid for many students but not for those in the classroom who are ELL.

Traditional views of validity classify the validity of a data collection instrument into three types: content validity, construct validity, and criterion-related validity.

Content validity addresses whether an instrument asks questions that are relevant to what is being assessed.

Construct validity is the degree to which a measure accurately represents the underlying, unobserved theoretical construct it purports to measure.

Criterion-related validity refers to how well a measure predicts performance. There are two types of criterion-related validity—concurrent and predictive. Concurrent validity compares performance on an assessment with that on another assessment. For example, how do scores on the statewide assessment correlate with those on another nationally normed, standardized test? Predictive validity indicates the degree to which scores on an assessment can accurately predict performance on a future measure. For instance, how well do SAT scores predict performance in college?

A fourth type of validity that is sometimes noted is consequential validity. Consequential validity refers to the intended and unintended social consequences of using a particular measure, for example, using a particular test to determine which students to assign to remedial courses.

When choosing an assessment or creating your own assessment, you should investigate the technical qualities of reliability and validity to be sure the test is consistent in its measurement and to verify that it does indeed measure what you need to measure.

Building your survey in conjunction with other methods and tools can help you understand your findings better. For instance, designing a survey to explore findings from observations or document reviews can enable you to compare your findings among multiple sources. Validating your findings using multiple methods gives the evaluator more confidence regarding evaluation findings.

Using a previously administered survey can save you time, may give you something to compare your results to (if previous results are available), and may give you confidence that some of the potential problems have already been addressed.

Two notes of caution, however, in using surveys that others have developed: (a) be sure the instrument has been tested and demonstrated to be

Rubrics are guidelines that can be used objectively to examine subjective data. Rubrics as an evaluation tool provide you with a way to identify, quantify, categorize, sort, rank, score, or code portfolios, observations, and other subjective data.

Rubrics are used to score student work, such as writing samples or portfolios, as well as to examine classroom implementation of a program. When rubrics are used to examine behavior or performance, observers rely on the rubric definitions to determine where the behavior or performance lies on the rubric scale. Rubrics are typically scaled 1 to 4 or 1 to 5, with each number representing a level of implementation or a variation of use.

Observers or rubric scorers must be highly trained so that scoring is consistent among scorers (referred to as inter-rater reliability) and over multiple scoring occasions.

Rubrics can also be used to facilitate program implementation. Providing those implementing a project or program with a rubric that indicates variations in implementation, as well as what the preferred implementation would look like, can help to promote fidelity of implementation. For instance, just as students are provided with a scoring rubric before they complete a writing assignment (so they know what is expected and what constitutes an ideal response), teachers or administrators could be provided with a rubric regarding how to use or operate a program or how to conduct an activity.

reliable, and (b) be sure the survey addresses your evaluation needs. It is tempting to use an already developed survey without thinking critically about whether it will truly answer your evaluation questions. Existing surveys may need to be adapted to fit your specific needs.

Interviews and focus groups (qualitative) are typically conducted face-to-face or over the phone. You can create an interview protocol with questions to address your specific information needs. The interviewer can use follow-up questions and probes as necessary to clarify responses. However, interviews and focus groups take time to conduct and analyze. Due to the time-consuming nature of interviews and focus groups, sample sizes are typically small, and research costs can be expensive.

Observations (usually qualitative but can be quantitative) can be used to collect information about people's behavior, such as teacher's classroom instruction or students' active engagement. Observations can be scored using a rubric or through theme-based analyses, and multiple observations are necessary to ensure that findings are grounded. Because of this, observational techniques tend to be time- consuming and expensive, but can provide an extremely rich description of program implementation.

When collecting, storing, and using data from any source (including surveys, interviews, observations, existing data, etc.), it is important to keep participant information and responses confidential.

Ethical considerations should be first and foremost in the minds of an evaluator. Participant privacy is more important than evaluation needs or evaluation results.

Existing data (usually quantitative but can be qualitative) are often overlooked but can be an excellent and readily available source of evaluation information. Using existing data such as school records (e.g., student grades, test scores, graduation rate, truancy data, and behavioral infractions), work samples, and lesson plans, as well as documentation regarding school or district policy and procedures, minimizes the data collection burden. However, despite the availability and convenience, you should critically examine the quality of existing data and whether they meet your evaluation needs.

Portfolios (typically qualitative) are collections of work samples and can be used to examine the progress of your program's participants throughout your program. Work samples from before (pre) and after (post) program implementation can be compared and scored using rubrics to measure growth. Portfolios can show tangible and powerful evidence of growth and can be used as concrete examples when reporting program results. However, scoring can be subjective and is highly dependent upon the strength of the rubric and the training of the portfolio scorers.

Case studies (mostly qualitative but can include quantitative data) are in-depth examinations of a person, group of people, or context. Case studies can include a combination of any of the methods reviewed above. Case studies look at the big picture and investigate the interrelationships among data. For instance, a case study of a school might include interviews with teachers and parents, observations in the classroom, student surveys, student work, and test scores. Combining many methods into a case study can provide a rich picture of how a program is used, where a program might be improved, and any variation in findings from using different methods. Using multiple, mixed methods in an evaluation allows for a deeper understanding of a program, as well as a more accurate picture of how a program operates and its successes. See Appendix C for resources on **Data Collection, Preparation, and Analysis**, as well as **Research and Evaluation Design, Including Reliability and Validity**.

Table 2 presents an overview of evaluation methods and tools used to collect data, noting advantages and disadvantages.

Table 2: Evaluation Methods and Tools: Overview

Methods and Tools	Basic Information	Advantages	Disadvantages
Assessments and Tests	• Usually quantitative but can be qualitative • Can be administered online or in person • Can be administered individually or in groups	• Multiple items may be used in combination to create a subscale, often providing a more reliable estimate than any single item. • Can be used pre- and post-program implementation to measure growth	• If assessment is not aligned well with the program, data may not be a meaningful indicator of program success. • If reliability and validity are not adequate, the data will be poor quality, and inaccurate conclusions may be drawn.
Surveys and Questionnaires	• Typically quantitative but can be qualitative • Can be administered in person, over the phone, online, or through the mail	• In-person surveys can be a quick method to collect data. • If conducted with a captive (in-person) audience, response rates can be high. • Electronic or Internet-based surveys can save time and costs with data entry and can improve data quality by reducing data entry errors.	• Due to postage costs and multiple mailings, mail surveys can be expensive. • Response rates of mail surveys can be low. • If upon data analysis it is found that questions were not worded well, some data may be unusable.
Interviews	• Qualitative method • Can be conducted in person or over the phone	• Follow-up questions can be used to obtain more detail when needed. • Follow-up probes can be used to determine how interviewees are interpreting questions. • Nonverbal communication during in-person interviews aids in response interpretation.	• Time-consuming to conduct • Time-consuming to analyze data • Limited number of participants • Can be expensive, depending on the number of people interviewed

Methods and Tools	Basic Information	Advantages	Disadvantages
Focus Groups	• Qualitative method • Multiple people can be interviewed at the same time.	• Follow-up questions can be used to obtain more detail when needed. • Follow-up questions can be used to determine how interviewees are interpreting questions. • Participants can build on each other's responses. • Often more cost effective than interviews • Nonverbal communication during in-person focus groups can aid in response interpretation.	• Group setting may inhibit participants from speaking freely. • Difficult to coordinate schedules with multiple people • Participants may focus on one topic, limiting exploration of other ideas. • Requires a skilled facilitator • Time-consuming to analyze data
Observations	• Typically qualitative but can be quantitative • Can be done in person, via videotape, through one-way glass, or from a distance	• Provides a good sense of the use of program • Allows the researcher to gain a full understanding of the environment of participants • Helps to provide a context for interpreting data	• Sometimes need many observations to gain a realistic sense of the use of a program • Time-consuming to observe, thus expensive • Time-consuming to analyze • Participant behavior may be affected by observer presence
Existing Data	• Can be qualitative or quantitative • Might include school records (electronic or paper based), work samples, lesson plans, or existing documentation (such as meeting minutes or attendance sheets)	• Low burden on participants to provide data • Relatively inexpensive to collect • Electronic data may facilitate analysis • Interpretation of existing data is often objective. However, interpretation of existing data such as documents or meeting minutes can be subjective.	• May not correspond exactly to evaluation needs • May be incomplete or require additional interpretation • May need special permission or consent to access and use • If not electronic, may be time- consuming to analyze

Methods and Tools	Basic Information	Advantages	Disadvantages
Portfolios	• Primarily a qualitative method • Can be captured and stored electronically	• Can provide a representative cross-section of work • If portfolio work is used pre-program and post-program, data can be used to examine growth.	• Scoring of qualitative work is often subjective. • Objectivity of results relies on strength of scoring rubric and training of scorers. So, reliability and validity should be considered.
Case Studies	• Primarily a qualitative method • Can include both qualitative and quantitative data • Can include a mixture of many methods, including interviews, observations, existing data, etc.	• Provides a multi-method approach to evaluation • Often allows a more in-depth examination of implementation and change than other methods	• Analyses of data can be subjective • Expensive to conduct and analyze; as a result, sample sizes are often small
Rubrics	• Quantitative method • Guidelines to objectively examine and score subjective data such as observations, portfolios, open-ended survey responses, student work, etc. • See **Rubrics** sidebar on page 47 for more information.	• Powerful method to examine variations of program implementation • Well-defined rubrics can be used not only for evaluation purposes but also to facilitate program implementation.	• Objectivity of results relies on strength of scoring rubric and training of scorers

Constraints

All programs have constraints during their implementation. Constraints might be contextual in that you may not have the support needed to fully evaluate your program. Or you may have resource constraints, including financial or time constraints. **Feasibility** is important to consider while designing your evaluation.

> Feasibility is important to remember when planning your evaluation. A good evaluation must be doable. Remember, a small evaluation is better than no evaluation, because basing program decisions on some information is better than basing decisions on no information. Considering the feasibility of carrying out your evaluation is critical when planning your evaluation. Be sure to plan within your organizational constraints.

A good evaluation must be doable. The design for a rigorous, comprehensive evaluation may look great on paper, but do you have the time available and financial resources necessary to implement the evaluation? Do you have adequate organizational and logistical support to conduct the evaluation the way you have planned?

Every evaluation has constraints, and if you do not consider them at the outset, your thoughtfully planned evaluation may be sidelined to no evaluation. Remember, a small evaluation is better than no evaluation, because basing program decisions on some information is better than basing decisions on no information. Considering the feasibility of carrying out your evaluation is critical to planning your evaluation. Be sure to plan within your organizational constraints.

You can also use your logic model to represent your evaluation time line and evaluation budget. The time frame of when and how often you should measure your short-term, intermediate, and long-term objectives can be noted directly on the logic model, either next to the headings of each or within each objective. Likewise, the cost associated with data collection and analysis can be recorded by objective.

By examining time line and budget by objective, evaluation activities that are particularly labor intensive or expensive can be clearly noted and planned for throughout the program's implementation and evaluation. The **Budgeting Time and Money** section in Appendix C includes several resources that may help you with considerations when budgeting time and money for an evaluation.

The READ E-Team decided on data collection methods, including the data sources, for each evaluation question and associated indicators. Two examples are provided below.

1. In what ways and to what extent did teachers integrate READ into their classroom instruction?
   - → A READ rubric will be used to measure teacher implementation of READ in the classroom.
   - → The rubric will be completed through classroom observations and teacher interviews.
   - → The READ implementation rubric will be on a 4-point scale, with a 4 representing the best implementation.
   - → Data will be collected monthly, alternating between classroom observations one month and interviews the following month.
2. To what extent did READ improve student learning in reading?
   - → The state reading assessment will be used to measure student learning in reading. It is administered in April of each academic year, beginning in second grade.
   - → READ assessment data will be used as a formative measure to examine student reading performance.
   - → State reading scores and READ assessment data will be disaggregated and examined by quality of teacher use (using the READ implementation rubric), frequency of home use, initial reading performance, grade level, gender, ethnicity, special education status, and English language proficiency.
   - → Previous year state reading assessment scores will be used as a baseline against which to measure student reading improvement.
   - → Reading scores on the state assessment will be analyzed in relation to scores on the READ assessments in order to determine the degree to which READ assessments correlate with the state reading assessment.

For a full list of evaluation questions, data sources, and data collection methods, see the READ Evaluation Matrix tables 10, 11, and 12 in Appendix A, **Step 3: Implement the Evaluation**. The READ Evaluation Matrix includes the READ logic model components, evaluation questions, indicators, targets, data sources, and data collection methods by the READ logic model strategies and activities, early/intermediate objectives, and long-term goals. The data analysis column in the READ Evaluation Matrix will be completed in Step 3.

STEP 3: IMPLEMENT – How Do I Evaluate the Program?

Ethical Issues

Because evaluation deals with human beings, ethical issues must be considered. Evaluation is a type of research—evaluators research and study a program to determine how and to what extent it works. You likely have people (perhaps teachers or students) participating in the program, people leading the program, people overseeing the program, and people relying on the program to make a difference. It is the responsibility of the evaluator to protect people during evaluation activities. An evaluator must be honest, never keeping the truth from or lying to participants. You should be clear about the purpose of the program and its evaluation. Respect for participants always comes before evaluation needs.

> Policies and procedures regarding informed consent and ethics to safeguard study participants must be followed *before* any data are collected.

Prior to collecting any data, check with your administration to see what policies and procedures are in place for conducting evaluations. Is there an Institutional Review Board (IRB) at the state, district, or school level that must be consulted prior to conducting an evaluation? Does your state, district, or school have formal Human Subjects Review procedures that must be followed? Does the evaluator need to obtain approvals or collect permission forms? Policies and procedures to safeguard study participants must be followed and permissions must be received before any data are collected. For resources on federal requirements regarding Institutional Review Boards or the protection of human subjects in research, see the **Ethical Issues** section in Appendix C.

Many programs are implemented as part of the school curriculum or as a districtwide or statewide initiative. In such cases, participants may by default participate in those programs as part of their education or work. However, if data are collected or used in the program evaluation, the participants have the right to consent or refuse to have their information used in the evaluation. In some situations and for some data, participants may have consented prior to the evaluation for their information to be used for various purposes, and their consent may extend to your evaluation. If you think this may be the case for your evaluation, be sure the evaluators verify it with your administration. In other instances, especially when data will be newly collected for the evaluation, the evaluator should obtain **informed consent** from participants before data collection begins. Depending upon the nature of your study and your institution, informed consent may be obtained through permission forms or through a formal human subjects review process.

As part of the evaluator's responsibility to protect people, information obtained through and used by the evaluation must be kept confidential. Individual identities should be kept private and access to evaluation data should be limited to the evaluation team. Evaluators should protect privacy and ensure **confidentiality** by not attaching names to data and also by ensuring that individuals cannot be directly or deductively identified from evaluation findings. An exception to this may be case studies or evaluations that use student work as examples. For these evaluations, you should take care that your informed consents and written permissions explicitly state that participating individuals or organizations consent to being identified in evaluation reports, either by name or through examples used in the report.

Finally, you must be especially careful not to blur the lines between the two roles of program staff and evaluation team when it comes to privacy and confidentiality. This is one of the reasons that it is prudent to have the external evaluator on your evaluation team collect, manage, and analyze your data. If your data are particularly sensitive or if evaluation participants were promised complete confidentiality, using an external evaluator to handle all data collection and management needs would be the practical and pragmatic choice, as well as the ethical preference. See the **Ethical Issues** section in Appendix C for resources on ethical considerations and obligations of evaluation.

How Do I Collect the Data?

Your data collection approach will depend upon your evaluation method. Table 3 includes an overview of data collection procedures for various evaluation methods.

Table 3: Evaluation Methods and Tools: Procedures

Methods and Tools	Procedures
Assessments and Tests	• Review the test to be sure that what it measures is consistent with the outcomes you hope to affect. • Review the test manual to be sure the test has adequate reliability and validity. (See **reliability and validity** sidebars on pages 45 and 46 for more information.) • Be sure that test proctors are well trained in test administration.
Surveys and Questionnaires	• Develop the survey questions or choose an existing survey that addresses your evaluation needs. • Pilot test the survey to uncover and correct problems with survey items and questions as well as to plan data analyses. • Decide in advance on a target response rate as well as the maximum number of times you will administer the survey or send the questionnaire. • Examine reliability and validity. (See **reliability and validity** sidebars on pages 45 and 46 for more information.)

Methods and Tools	Procedures
Interviews	• Develop an interview protocol, highlighting key questions. • Include question probes to gather more in-depth information. • Limit how long the interview takes so that participants will be more willing to participate (and make sure to tell participants how much time will be needed for the interview). • Obtain permission to digitally record so that you can concentrate on listening and asking questions. (The recording can be transcribed and analyzed after the interviews.)
Focus Groups	• As with an interview, develop a focus group protocol that includes key questions. • Limit group size. (Using six to eight participants tends to work well, though a skilled facilitator may be able to increase the size.) • Purposefully organize focus groups that include participants who can build upon and benefit from each other's ideas, providing for a richer discourse. • Purposefully organize focus groups that include participants who will feel comfortable speaking their opinion in the group. • Obtain permission to digitally record so that you can concentrate on listening and asking questions. (The recording can be transcribed and analyzed after the focus groups.)
Observations	• Design the observation protocol and rubrics (if you will be analyzing data with rubrics). Remember to consider the environment and atmosphere, dispositions, pedagogy, curriculum, etc. when designing your protocol and rubrics. • Observers should try to be as unobtrusive as possible so as to not influence the environment they are observing. • See the **rubrics** row below in this table for pointers on design and consistency in scoring.
Existing Data	• Review existing data for applicability and accuracy. Caution: Simply because data exist does not mean that they are complete or accurate.
Portfolios	• Choose artifacts to be included in the portfolio. • Design the scoring rubric in advance. • See the **rubrics** row below in this table for pointers on design and consistency in scoring.
Case Studies	• Case studies might involve a combination of the above methods.

Methods and Tools	Procedures
Rubrics	• Design the scoring rubrics before examining the qualitative data. • Describe the best response or variation in detail. • Decide on the number of variations or categories. (It works well to use four or five.) • For each variation, describe in detail what the response or variation would look like. Typically, the best response is at the top of the scale. For example, on a scale of 1 to 4, the best response would be a 4. A variation with many but not all components of the best might be a 3. A variation with a few components of the best response might be a 2, while a variation with little to no components of the best response would be a 1. • Train raters or observers how to score using the rubric. Use several raters to score the same responses, observations, or student work using the rubric. Compare scores to examine inter-rater reliability. Discuss scoring among raters to improve consistency.

How Should I Organize the Data?

Confidentiality and individual privacy are of primary importance during all aspects of the evaluation.

An evaluator should safeguard that private information is not divulged in conversations regarding the program; during data collection, organization, and storage; and through evaluation reporting.

During data collection, procedures should be put in place to protect **privacy** and to provide data **security**. For instance, if data can be tied to individual respondents, assign each respondent an identification number and store data according to that number. Often when data are collected at multiple times during the evaluation (e.g., pre and post) or when data sources need to be individually connected (e.g., student demographic data and assessment data), a secondary data set can be created to match identification numbers with respondents. If this is the case, this secondary data set should be encrypted and kept highly confidential (i.e., stored in a locked office and not on a shared server), so that individual information cannot be accessed intentionally or inadvertently by others. It is also good practice to control and document who has access to raw evaluation data.

You should also document your data sets. Having good documentation increases the credibility of your evaluation should questions be asked regarding your findings. It is sound practice to keep a record of what data were collected, when they were collected, and how respondents and other participants were chosen. This documentation also should include any definitions that might be necessary in order to interpret data, as well as interview protocols or survey instruments that were used. Documentation of data collected and how data were stored will be useful if you should want to reanalyze your data in the future, if someone asks you questions

about your data, or if someone would like to replicate your evaluation. See the **Data Collection, Preparation, and Analysis** section in Appendix C for resources on data preparation and creating codebooks to organize and document your data.

How Should I Analyze the Data?

The purpose of analyzing your data is to convert all of the raw data that you have collected into something that is meaningful. Upon organizing your data, you may find that you are overwhelmed with the data you have available and wonder how you will make sense of it. **Start with your logic model and evaluation questions**. List the indicators and associated targets you have outlined for each evaluation question. Use what you have set up during your evaluation design to organize your analysis. Take each evaluation question one at a time, examine the data that pertain to the indicator(s) you have identified for the evaluation question, and compare the data collected to your targets.

Analyzing your data does not have to be daunting. Often when people think of data analysis, they assume complicated statistics must be involved. In reality, there are two things to keep in mind:

- Not all data analysis involves statistics.
- Even if statistics are involved, they should be at the level that the intended audience will understand.

Analysis methods differ by the type of data collected. If the information to be analyzed includes **quantitative data**, some type of statistical analysis will be necessary. The most common way statistics are used in evaluation is for descriptive purposes. For example, if you want to describe the number of hours students spent using a computer at home or at school, you would calculate either the average number or the percentage of students who use computers for a specified period of time. Or, you may want to compare the results of one group of students (e.g., at-risk students) to another group to see if technology influences different groups differently. In this case, you may want to use the same statistics (e.g., means and percentages), but report separate results by group.

You may also want to use a simple test of significance (e.g., *t*-test) to see if the differences in means are statistically significant (i.e., unlikely to differ by chance). Whether you use simple descriptive statistics or tests of significance and how you want to group your information depend on the type of information you have collected and your evaluation questions. For more complex data sets or in-depth analyses, more sophisticated statistical techniques, such as regression analysis, analysis of variance, multilevel modeling, factor analysis, and structural equation modeling can be used.

If the information to be analyzed involves **qualitative data**, such as data collected from open-ended survey questions, interviews, case studies, or observations, data analysis will likely involve one of two methods. The first is to develop a rubric to score your interview or observational data. Remember, if at all possible, the rubric should be developed in advance of data collection. Once data are scored using the rubric, you can use quantitative analyses to analyze the resulting numerical or categorical data.

A second method to analyze qualitative data is to create a protocol to aid you in data analysis. Such protocols typically call for an iterative process of identifying and understanding themes, organizing data by emerging themes, coding data by theme, and making assertions or conclusions based on these themes. Often, example responses or descriptions taken from the data are used to support the assertions. As with quantitative data, it is important when reporting qualitative data not to inadvertently reveal an individual's identity. All assertions and findings should be "scrubbed" to be sure that someone reviewing the report cannot deductively identify evaluation participants. See Appendix C for more information on **Data Collection, Preparation, and Analysis**.

When developing a rubric to code qualitative data:

- Decide on the number of variations or categories. (It works well to use four to five categories.)
- Describe the best response in detail.
- For each subsequent variation, describe what the response would look like. For example, on a scale of 1 to 4, the best response would be a *4*. A variation with many but not all components of the best might be a *3*. A variation with a few components of the best response might be a *2*, while a variation with little to no components of the best response would be a *1*.

The READ external evaluator collected a mix of quantitative and qualitative data to address evaluation questions. Qualitative data collected through observations and interviews were coded using the READ implementation rubric and analyzed using descriptive statistics, including means and frequency distributions. Student reading assessment data were analyzed by testing for statistical significance, comparing mean test scores between groups of students and over time. An example is provided below. The full READ Evaluation Matrix start in Appendix A, at **Table 10: READ Evaluation Matrix—Strategies and Activities/Initial Implementation**.

1. **Logic Model Component**: Improved integration of READ into classroom instruction (intermediate objective).
2. **Evaluation Question**: In what ways and to what extent did teachers integrate READ into their classroom instruction?
3. **Indicator**: Improved integration of READ lessons into classroom instruction.
4. **Targets**: By April, 50% of teachers will score a 3 or above (out of 4) on the READ implementation rubric. By June, 75% of teachers will score a 3 or above on the READ implementation rubric.
5. **Data Source**: READ implementation rubric (developed by the E-Team and administered by Dr. Elm).
6. **Data Collection**: Rubric completed through alternating, monthly classroom observations and teacher interviews.
7. **Data Analysis**: Rubric scores aggregated into frequency distributions and means; change over time to be analyzed.

All data collected through the evaluation were managed and stored by Dr. Elm, the external evaluator. The computer used for storage and analysis was located in a locked office. Only the external evaluator had access to the raw data. Data were backed up weekly to an external drive, which was kept in a locked drawer. To protect teacher and student privacy, identification numbers were assigned to all participants. Teacher and student names were not recorded with the data.

READ online records regarding student and teacher use, rubric data, and survey data were only accessible by the external evaluator. Results that were released were only in aggregation and had no identifying information. All evaluation data were secured and kept confidential to protect individual privacy.

Managing the Unexpected and Unintended

Just as with life, sometimes the unexpected happens. Perhaps you find that you were unable to collect all the data you had outlined in your design. Or maybe the existing data that you were relying on are not accessible. Or the data were available but the quality was not as good as you had expected (e.g., too much missing information or recording errors). Or possibly you were unable to get enough program participants to respond to your survey or agree to an interview. Don't panic. Go back to your evaluation questions. Reexamine your indicators and measures. Is there another measure that can be used for your indicator? Is there another indicator you can use to address your evaluation question? Think creatively about what data you might be able to access or collect. You may find that you are not able to answer a certain evaluation question or that your answer to that question will be delayed. Or you may find that you can answer your question, sort of, but not in the best way. In any case, document what happened, explain what alternatives you are pursuing, and simply do the best you can. Evaluation does not occur in a sterile laboratory but within the course of everyday practice. Your evaluation might be less than ideal at times and you will undoubtedly face challenges, but in the long run, some information is better than no information. See Appendix C for resources on **Evaluation Pitfalls**.

STEP 4: INTERPRET – How Do I Interpret the Results?

How Do I Examine and Interpret My Results?

In the end, the important part of collecting and analyzing your information is not the statistics or analytical technique but rather the conclusions you draw. The process of coming to a conclusion can vary from goal to goal and objective to objective. One of the most difficult tasks is defining vague goals and objectives, such as "sufficient training" or "adequate progress." However, you have gone to great lengths to understand your program and plan your evaluation, and you have already developed targets for your indicators. Because of this, your interpretation of results will likely be more straightforward and less cumbersome.

Examination of evaluation results should be ongoing. It is not wise to wait until the end of an evaluation to analyze your data and interpret your results. For instance, if your evaluation results from implementation reveal that program activities were not put into place, continuing with the measurement of short-term and intermediate objectives is likely a waste of your resources. Similarly, if the evaluation of intermediate objectives reveals that outcomes are not as envisioned, an important question would be whether the program should be modified, scaled back, or discontinued. Do results indicate that the program is not working as expected? Or do results reveal that the program's theory is invalid and needs to be revisited? Was the program implemented as planned? Is it reasonable to think that making a change in the program could improve results? These are important questions to consider before moving on to the measurement of progress toward long-term goals.

> Examination of data and interpretation of findings should be ongoing. Do not wait until the end of the evaluation!
>
> In order to use evaluation for program improvement, communication of findings should be regular, continuous, and timely.

The READ evaluation subcommittee, E-Team, examined the evaluation results and determined the following. Use of these findings will be discussed in Step 5.

Summative

1. First-year results indicate that state reading scores for READ students are higher than those for non-READ students. The gains are especially compelling for classrooms in which READ was used regularly and with fidelity, where increases in state reading scores were over three times that of non-READ students.
2. Students in classrooms where READ was used regularly and with fidelity increased their reading scores on the state assessment by twice that of students in READ classrooms where READ was used minimally.
3. Students of teachers who used READ assessment data as intended to differentiate instruction increased their reading scores on the state assessment by twice as much as students of teachers who did not use READ assessment data as intended.
4. Student scores on READ assessments had a significant and strong positive correlation with student scores on the state reading assessment, indicating the state reading and the READ assessments are likely well aligned and that READ assessment data are likely a good indicator of performance on the state reading assessment.

Formative

5. State reading assessment data could not be analyzed by home use of the READ program because only one classroom implemented the home component.
6. At the start of the year, teacher use of READ was promising and the program met its targets. However, as the program progressed and as more teachers were pressed to improve their use of READ, several targets were not met. READ student assessment data were not used as regularly by teachers as the classroom component of READ.

A full accounting of evaluation results by logic model component and evaluation question is provided in Appendix A, starting at **Table 13: READ Evaluation Results—Strategies and Activities/Initial Implementation**.

Interpretation should address the relationship between implementation and long-term goals. Presuming that your program was implemented and ongoing results were promising, to what extent did the program accomplish its long-term goals?

During interpretation, consider how the program worked for different groups of participants and under different conditions. You may also want to examine how long-term outcomes vary with implementation, as well as with results from short-term and intermediate indicators.

Results should be examined in relation to the proposed program's theory. Do evaluation findings support the program's theory? Were the assumptions underlying the program's theory validated? If not, how did the program work differently from what you had proposed? How can the theory and the logic model representing this theory be changed to reflect how the program worked?

The logic model can be used as a tool to present evaluation findings, as well as to explain the relationships among components of the program. Updating the logic model to include results can be a useful reporting and dissemination tool.

Cautions During Interpretation

Two common errors during results interpretation are **overinterpretation** and **misinterpretation of results**. Unless the evaluation design was a randomized, controlled experiment, results interpretation should not claim causal relationships. Indeed there may be relationships between your program's activities and its outcomes (and hopefully there will be!), but unless all rival explanations can be ruled out, causal associations cannot be claimed. Doing so would be an overinterpretation of your results.

> When interpreting evaluation findings, be careful not to claim the data say more than they actually do!

Additionally, when interpreting results, you should consider possible alternative theories for your results. Considering and recognizing other explanations or contributors to your evaluation results does not diminish the significance of your findings but rather shows an understanding of the environment within which your program was implemented.

Over time, it is a combination of factors, some unrelated to the program itself, that interact to create results. Documenting your program's environment can guard against misinterpretation of results and instead provide a thoughtful description of the circumstances under which the results were obtained. See Appendix C for more information on **Interpreting, Reporting, Communicating, and Using Evaluation Results**.

Although the READ evaluation was a true experimental design, E-Team members knew it would still be worthwhile to consider the possibility that other factors might have influenced the positive findings. The E-Team therefore brainstormed possible competing explanations for the positive results of the READ program.

The E-Team decided that another plausible explanation for the positive results was that the teachers who used READ regularly in the classroom and who used READ assessments as intended may have been more skilled teachers and their students might have had a similar increase in reading scores even without the READ program. The E-Team decided to follow up on fidelity of implementation and its relationship to teacher skills. In addition, while classrooms were randomly assigned to READ to minimize initial differences between READ and non-READ classrooms, it is possible that by chance more skilled teachers were assigned to the READ program group. The E-Team also intends to investigate this issue further in Year 2 of the evaluation.

How Should I Communicate My Results?

As mentioned earlier, evaluation findings should be communicated to program staff on an ongoing and regular basis. These formative findings are critical to program improvement. Setting a schedule for regular meetings between program staff and the evaluation team, as well as building these communications into your time line, will ensure that evaluation findings can truly help the program during its operation. Evaluators can provide quick feedback at any stage of the program to help improve its implementation. For instance, if an evaluator notices from observing professional development sessions that teachers are leaving the training early to attend another faculty meeting, the evaluator should give quick feedback to program staff that the timing of sessions may not be convenient (and for this reason, teachers are not receiving the full benefit of the training).

> Setting up regular times throughout the program's operation to share evaluation findings with program staff and other stakeholders is a key responsibility of the evaluator and critical to a program's success.

Suppose the evaluator finds during the early stages of the program (through interviews or classroom observations) that teachers are struggling with the technology needed to use the program in the classroom. The evaluator can give quick feedback at a monthly meeting or through an email that technology support and technical assistance are needed in the classroom. Remember, however, an evaluator should not report on individual teachers or classrooms unless consent to do so has been obtained. Doing so could violate the ethical

obligation to participants in the evaluation and undermine future data collection efforts. Even quick feedback should maintain confidentiality.

In addition to relaying your findings on an ongoing basis for formative purposes, you will also want to communicate your summative evaluation findings regarding the extent of your program's success to stakeholders, including administrators, school board members, parents, and funders. The first step to communicating your results is to determine your audience. If you have multiple audiences, (e.g., administrators and parents), you may want to consider multiple methods of reporting your findings, including reports, presentations, discussions, and short briefs. Make a list of (a) all people and organizations you intend to communicate your results to; and (b) any others you would like to know about your evaluation findings. For each audience, ask yourself these questions:

- What background do they have regarding the program?
- What will they want to know?
- How much time and interest will they have?
- What do you want the audience to know?

Thinking through these questions will help you tailor your communication. In general, if you are given guidelines on what to report by a funder or by the state or district, try to follow them as closely as you can. If you are not given guidelines, then put yourself in the position of your audience and consider what information you would like to know. Here are some tips to keep in mind:

- If the audience already has background information on the program, try to focus on providing only specific findings from your evaluation. If your audience is not familiar with your program, you can use your program theory and logic model to introduce the program and provide a description of how the program is intended to work.
- Address the goals and objectives that you believe the audience would most want to know about.
- If the audience wants information immediately, write a short summary of major findings and follow up with a longer, more detailed report.

> Don't rely on the typical end-of-year evaluation report to communicate evaluation findings. Communicate to multiple audiences using multiple methods.
>
> In addition to regularly sharing evaluation findings with program staff, let other stakeholders know on an ongoing basis how the program is doing. Think creatively about modes of communication that will reach all stakeholders.

- Don't be afraid to include recommendations or identify possible areas for change. Recommendations are a critical piece to making sure your evaluation findings are used appropriately. If you want to make changes, you are going to have to talk about it sooner or later, and having it in the report is a good way to start the conversation.

Finally, a long report is not the only way to communicate results. It is one way and perhaps the most traditional way, but there are many other methods available. Other options include:

- A memo or letter;
- A special newsletter or policy brief;
- A conference call or individual phone call;
- A presentation before a board or committee, or at a conference;
- A publication in a journal, newspaper, or magazine;
- A workshop;
- A web page or blog; or
- The school district newsletter or website.

Evaluation reports or presentations typically have a common format. First is an executive summary or overview that notes key findings. In fact, some will read only the executive summary, so you want to be sure it has the most important information. Other report sections might include:

- Introduction (including program background and theory);
- Evaluation design (including logic model, evaluation questions, and evaluation methods);
- Results (including all findings from the evaluation, organized by evaluation question);
- Conclusions (including your interpretation of the results);
- Recommendations (including how the program should proceed based on your findings); or
- Limitations (including limitations based on evaluation design, analysis of data, and interpretation of findings).

See Appendix C for more information on **Interpreting, Reporting, Communicating, and Using Evaluation Results**.

The READ oversight team met monthly to discuss program monitoring and improvement. At each meeting, the READ evaluator, Dr. Elm, and the E-Team provided an update to the oversight team. Based on the formative evaluation findings, the oversight team developed recommendations and a plan for the next month.

At the December school board meeting, the oversight team presented a status report, noting important findings from the evaluation. The oversight team asked Dr. Elm to create a full evaluation report for the administration and to present the findings at the August school board meeting. The E-Team also drafted a one-page brief of evaluation findings which was provided to all participants, as well as to the local newspaper.

STEP 5: INFORM and REFINE – How Do I Use the Evaluation Results?

Informing for Program Improvement

One of the most important uses of evaluation findings is for **program improvement**. In fact, for many audiences, your evaluation communication should focus on improvement. In order to do this, evaluation communication and reporting should include not only positive findings but also findings that may not be flattering to your program. These not-so-positive findings are the basis for program improvement.

> Be sure to report both positive and negative findings. Negative findings can be communicated as lessons learned or areas for improvement.

When using evaluation results, ask yourself whether your findings are what you expected. Has the program accomplished what was intended? If yes, do you see areas where it can be made even better? If no, why do you think the program was not as successful as anticipated? Did the program not have enough time to be successful? Was the implementation delayed or flawed? Or perhaps the program theory was not correct. In any case, using evaluation results is vital to improve your program.

Informing for Accountability

Another important use of evaluation findings is for **accountability** purposes. Designing and implementing programs take valuable resources, and your evaluation findings can help you determine whether the expenditure is worth the results.

Accountability pertains to basic questions, such as whether the program was indeed implemented and whether program funding was faithfully spent on the program, and to more involved questions, such as whether the program is a sound investment. For this reason, as with program improvement communications, it is important for your evaluation reporting to include all findings, good and bad, so that informed decisions can be made regarding the program's future. Should the program be continued or expanded? Should it be scaled back? While evaluation reporting can be used for program marketing or for encouraging new funding, evaluation findings should include sufficient information for decisions regarding accountability. A caution, however, is that decisions regarding accountability should be made carefully and be based on evidence from multiple sources derived from a rigorous evaluation.

During her evaluation update at the November oversight team meeting, Dr. Elm shared initial findings from the **evaluation of the implementation** of READ program activities. Indicators showed that many students did not have the technology available at home to access READ. Even within those schools that had high numbers of students with the technology necessary for home access, the classroom variability was large. Only one of the 40 classrooms was able to have 100 percent of students access READ from home. Open-ended survey items revealed that teachers did not feel comfortable offering READ homework assignments to some but not all students in their classroom and therefore chose not to train students in the home use of READ. Only one teacher had trained his students in the home use of READ because all of his students had the technology at home necessary to access READ. This teacher indicated that he would like to continue with the home component of READ.

The oversight team discussed the home-component issue and asked for advice from the E-Team on how to proceed. With the support of the E-Team, the oversight team decided to have a one classroom pilot of the home component but otherwise to remove the home component from the program during Year 1. Based on results from the pilot, implementing a partial home component in Year 2 would be considered.

During the same November update, Dr. Elm provided some findings from the **evaluation of the early/short-term objectives** on the READ logic model. She noted that in October all teachers had reported using READ in their classroom and that over half of teachers reported that they had used READ every week. However, over one-quarter of teachers reported that they had used READ in their classroom only once or twice in the last month. Survey data indicated that some of these teachers felt overwhelmed with the technology and some said they could not fit READ classroom use into their already busy day.

The oversight team discussed this information and decided to make a midcourse adjustment. Before the READ program began, team members had thought that the initial professional development and ongoing technical assistance would be sufficient. However, they now believed that they needed to make one-to-one professional development available to those teachers who would like to have someone come into their classroom and model a lesson using READ. Mrs. Anderson assigned arrangements for this one-on-one professional development to one of the oversight team members.

During her evaluation update at the January oversight team meeting, Dr. Elm shared findings from the **evaluation of the intermediate objectives** on the READ logic model. Dr. Elm explained that on the December teacher survey, slightly less than half the teachers reported that they used the READ assessment data on a weekly basis for planning and differentiating instruction. One in 10 teachers said they had never used the READ assessment data. Dr. Elm further stated that the lack of use of the READ assessment data was likely affecting scores on the READ implementation rubric. From classroom observations, interviews, and surveys, she believed that the quality of teacher use of READ in the classroom was progressing nicely but that the lack of assessment data use was decreasing the overall rubric score.

The oversight team knew that using the READ assessment data to plan and differentiate instruction was critical to the program's success. Mrs. Anderson decided to discuss the issue with the READ faculty at each school in an effort to understand what she could do to facilitate their use of the READ assessment data. Additionally, the E-Team planned to elaborate on the rubric so that subscores could be captured for various components of the rubric. These rubric subscores would be especially useful for analysis when the data are disaggregated by teacher use of READ in the classroom, student interaction in the classroom, and teacher use of READ student assessment data to plan and differentiate instruction. The revised rubric would be developed during the spring, piloted over the summer, and implemented during Year 2.

Finally, at the evaluation update at the end of the school year, Dr. Elm reported on the **preliminary evaluation of long-term goals** of the READ program. Student reading achievement was higher among students of teachers who used READ regularly and as intended, and the difference was statistically significant. Further, students of teachers who used the READ assessment data to tailor classroom instruction had higher reading test scores than students of teachers who did not use the READ assessment data, and again the difference was statistically significant.

Year 1 evaluation findings also indicated that not all teachers had bought into using READ with their students, especially the READ assessment component. The oversight team decided to share the evaluation findings with all teachers at a staff meeting in order to encourage them to use READ in their classroom. Prior to sharing the evaluation findings with teachers, Dr. Elm conducted an anonymous follow-up survey at the staff meeting in an effort to find out why some teachers chose to not use READ.

If your program design and evaluation were inclusive processes that involved stakeholders and participants from the start, it is more likely that your evaluation findings will be used for program improvement and accountability. Involving others in your program's implementation encourages a shared sense of responsibility for the program as well as a shared investment in the program's success. Hearing about a program at its very start and not again until an evaluation report is provided does not foster the ownership among staff, stakeholders, and participants that is needed for a successful program.

Why do we evaluate?

To ensure that the programs we are using in our schools are beneficial to students; to make programs and projects better; and to learn more about what programs work well, for whom, and to what extent.

How do we increase the likelihood that evaluation results will be used?

Create the opportunity for evaluation to have an impact on programmatic decision-making.

How do we create this opportunity?

We can start by:

- Embedding evaluation into our programs from the outset;
- Communicating program findings frequently and regularly; and
- Making the evaluation process participatory from start to finish.

So, **how do you make sure your evaluation report, along with all of your hard work and informative results, is not put on a shelf to gather dust? Make evaluation a participatory process** from understanding and defining the program in Step 1 to informing the program in Step 5.

Refining the Program's Theory

Your evaluation findings should also be used to refine your logic model. As mentioned earlier, the logic model is a living model and its underlying assumptions should be dynamic, changing as new information is learned. If the culture in which your program is implemented is a learning culture, using findings to improve the logic model is a natural. However, in other environments, it may not be as easy to apply your findings to logic model improvement. Regardless, if your program is to continue, you should keep its program logic model up-to-date.

An up-to-date logic model can facilitate future evaluation and serve as the cornerstone of your program. Your program's theory and logic model should be part of the core documentation of your program and can be used to train new program participants, as well as to explain the program to parents, administrative staff, potential funders, and other stakeholders.

Take Action

You have completed a lot of work. You have distributed your evaluation findings through letters, reports, meetings, and informal conversations. You have given presentations. So, what do you do now? How do you make sure that your information is used?

First, think about what changes you would like to see. Before you can attempt to persuade others to use your information, you need to figure out what you would like to happen. What changes would you like to see or what decisions do you think need to be made as a result of your information?

Second, think about what changes others might want. Learning how others would like the information to be used gives you more awareness of where they are coming from and more insight as to how they would best be motivated.

Next, **take action**. You have evidence from your evaluation, you have shared it with others, and you know what you want done. Ask for it! Find out who is in charge of making the changes you want and make sure they hear your findings and your recommendations. Give them a chance to process your suggestions. Then follow up. See Appendix C for more information on **Interpreting, Reporting, Communicating, and Using Evaluation Results**.

The READ oversight team felt that the logic model they created accurately portrayed the program. Yet, since it was clear from November that the home component could not be fully implemented, they wanted to highlight this on the logic model. The team decided to draw a box around the program as it was implemented, excluding the home component. Below the model, a note was provided indicating why the home component was not part of the existing implementation and that it was currently being piloted in one classroom. The oversight team hoped to understand more about the implementation of the home component, as well as the success of the home component, from examining results from the pilot classroom.

The oversight team also wanted to understand more about the strength of the relationship between classroom use of READ and state assessment scores and between use of READ assessment data for instructional planning and state assessment scores. It noted this on the logic model and asked the E-Team to investigate the linkages further in the second year of the evaluation.

Change Takes Time

One final note: **change takes time**. We all want to see the impact of our efforts right away, but in most cases change does not happen quickly. Embedded evaluation allows you to show incremental findings as you strive to achieve your long-term goals, and can help you to set realistic expectations regarding the time it takes to observe change related to your indicators. If you plan to use your evaluation results to advocate program expansion or to secure funding, keep in mind that changing policy based on your findings also will take time. People need to process your evaluation findings, determine for themselves how the findings impact policy and practice, decide how to proceed based on your evidence, and then go through the appropriate process and get the proper approvals before you will see any change in policy from your evaluation findings. As mentioned earlier, including others throughout your program's design and implementation can facilitate the change process. However, even with a participatory evaluation and positive findings, policy change will occur on its own time line.

The READ oversight team recommended that the READ program be offered to all students in the district. It also recommended that the program be incorporated into the regular curriculum. The team felt that the positive findings regarding test scores were strong enough that all students should have access to it.

However, since READ funding was still at the 50% level for the second year, the oversight team planned to work with Dr. Elm and the E-Team for another year in order to continue to refine the implementation of the program in the classroom and to further understand the success of the READ program with students. To do this, the team recommended that the second-year evaluation include student surveys and focus groups as data sources to address objectives related to student interaction and engagement in the classroom.

The oversight team decided to continue to advocate for the program's expansion in the hope that it would be institutionalized soon.

Appendix A: Embedded Evaluation Illustration – READ*

Program Snapshot

The Reading Engagement for Achievement and Differentiation (READ) program is a districtwide initiative focused on improving student reading skills in Grades 3-5. READ uses an experimental evaluation design and theory-based, embedded evaluation methods.

*This example was created solely to illustrate how the principles in this guide could be applied in actual situations. The program, characters, schools, and school districts mentioned in the example are fictitious.

Step 1: Define the Program

Background

For the past 5 years, reading scores in the Grovemont School District have been declining. The curriculum supervisor, Mrs. Anderson, has tried many strategies to improve reading skills. However, scores continue to decline. Mrs. Anderson has been searching for curricular and assessment materials that are better aligned with state reading standards and that provide ongoing standards-based assessment data. Mrs. Anderson found a program called READ (Reading Engagement for Achievement and Differentiation) that looked promising. After reviewing research on the program and documentation from the vendor as well as numerous discussions and interviews with other districts that had implemented the program, Mrs. Anderson and the district superintendent decided to present the READ program to the school board, in order to gain approval for funding the program for Grades 3-5.

At last month's meeting, the school board voted to partially fund the READ program. Due to recent state budget cuts, the school board was only able to fund the program at 50% for 2 years. At the end of the 2 years, the board agreed to revisit its funding decision. The board required an evaluation report and presentation due in September of each year.

Before starting to plan the READ program, Mrs. Anderson invited one teacher from each of the district's six elementary schools, the district reading coach, one of the district's reading specialists, and the district technology coordinator to join the READ oversight team. This 10-member team was charged with planning the READ program and its evaluation. The team asked

an evaluator from the local university to conduct the READ evaluation and to attend oversight team meetings.

The Evaluation

The oversight team asked the external evaluator, Dr. Elm, to help them plan the evaluation. Dr. Elm suggested that the oversight team build evaluation into its program as the team is designing it. By embedding evaluation into the program, information from the evaluation would be available to guide program implementation. Evaluation data would both drive program improvement and be the foundation for future decisions regarding whether the program should be continued, expanded, scaled down, or discontinued.

The oversight team members invited Dr. Elm to lead them through the process of building evaluation into their program planning. Dr. Elm explained that the first step is to gain a thorough understanding of the program. In doing this, Mrs. Anderson shared the materials she had already reviewed with the oversight team. In addition, the oversight team contacted four school districts that had used the READ program successfully in order to learn more about the program. To develop a thorough and shared understanding of the context in which the READ program would be implemented, the team reviewed the state's reading standards, the district's strategic plan, the district's core learning goals and curriculum maps in reading, and the district's technology plan. The team also examined reading grades and state reading assessment scores for the district as a whole, as well as by school, English Language Learner (ELL) status, and special education status for the past 5 years.

The next step, stated Dr. Elm, is to define the program by explaining the program theory. Explaining the program theory will include what the program is intended to accomplish, as well as how and why the program is expected to work. Dr. Elm recommended that the team complete the program theory in three parts: (a) defining the program’s long-term goals, (b) delineating the program’s strategies and activities, and (c) explaining how and why the team believes the program’s activities and strategies will result in the desired outcomes.

Program Goals

Based on their review of research and documentation as well as discussions and interviews with other districts that had implemented the program, and meetings with district administration and school staff, Mrs. Anderson and the oversight team set the following long-term goals for READ:

1. Increased student engagement in reading
2. Improved student reading skills

Program Strategies and Activities

The READ oversight team examined program materials to determine the primary components of the READ program. They determined that the READ program had three strategies: classroom lessons, homework, and assessments. Each of these strategies required certain activities in order to be successful. For instance, teachers would need professional development on how to integrate the READ classroom lessons into their instruction, as well as how to use the READ assessment data. Students would also need training in how to use the READ system in the classroom and at home.

After careful review of the READ program and the district's particular program needs, the oversight team outlined the following primary strategies and activities for the READ program:

1. Interactive, standards-based classroom lessons (using the READ software with interactive classroom technologies and individual handheld mobile devices for each student)
2. Standards-based reading assessments (Internet-based, formative READ assessments of student reading skills administered using the READ software)
3. Standards-based reading homework (Internet-based using READ software)
4. Teacher professional development on integrating READ into classroom instruction (using an interactive wireless pad)
5. Teacher professional development on using READ assessment data for classroom lesson planning
6. Student training on using READ (in the classroom and at home)

Relating Strategies to Goals: Program Theory

During a planning meeting focusing on why READ strategies and activities should result in the desired long-term goals, the oversight team brainstormed the underlying assumptions that were necessary for READ to work. The evaluator, Dr. Elm, facilitated the discussion among the oversight team members, leading them through the process of linking the program's activities and strategies to the long-term goals. Dr. Elm asked each member of the team to record why and how they thought each strategy or activity would lead to increased student engagement and improved student reading skills. Team members shared their reasoning with the group.

Dr. Elm led a discussion with the oversight team in which they examined each team member's ideas regarding why the program should work. Focusing on these ideas but not limited by them, the team members formulated, as a group, the underlying assumptions that were necessary to relate READ strategies and activities to long-term goals. During the discussion, team members were able to build on each other's ideas in order to construct a comprehensive theory that was supported by the group.

As a result of their discussion, the team put forward the following seven assumptions forming the basis of READ's program theory:

1. **Interactive, standards-based classroom lessons using READ software** will increase student interaction during learning, which will lead to increased exposure to standards-based learning opportunities.
2. **Standards-based reading assessments using READ software** will increase the availability of formative, standards-based data on student reading performance, which will lead to increased teacher use of formative standards-based reading assessment data and then improved differentiation of instruction.
3. **Standards-based reading homework using READ software** will increase student exposure to standards-based learning opportunities.
4. **Teacher professional development** on integrating READ into their classroom instruction will increase teacher use of READ, which will lead to improved integration of READ into classroom instruction. Teacher professional development on using READ assessment data for classroom lesson planning will increase teacher use of formative standards-based student reading assessment data. Both will lead to improved differentiation of instruction.
5. **Student training** on using READ in the classroom will increase student interaction during learning. Student training on using READ at home will increase student use of READ at home. Both will lead to increased student exposure to standards-based learning opportunities.
6. Increased student interaction in the classroom and improved differentiation of instruction will result in **increased student engagement**.
7. Increased student exposure to standards-based learning opportunities, improved differentiation of instruction, and increased student engagement will result in **improved reading skills**.

Resources

The oversight team also identified contextual conditions and resources necessary to the success of READ:

1. Program funding for READ, as well as necessary equipment to support infrastructure needs.
2. Program funding for external evaluation assistance.
3. Technology infrastructure at school:
    a. Classroom computer with Internet access
    b. Interactive technologies in each classroom
    c. Interactive, wireless pad for convenient, mobile teacher operation of computer

   d. 25 student handheld mobile devices per classroom for interactive learning

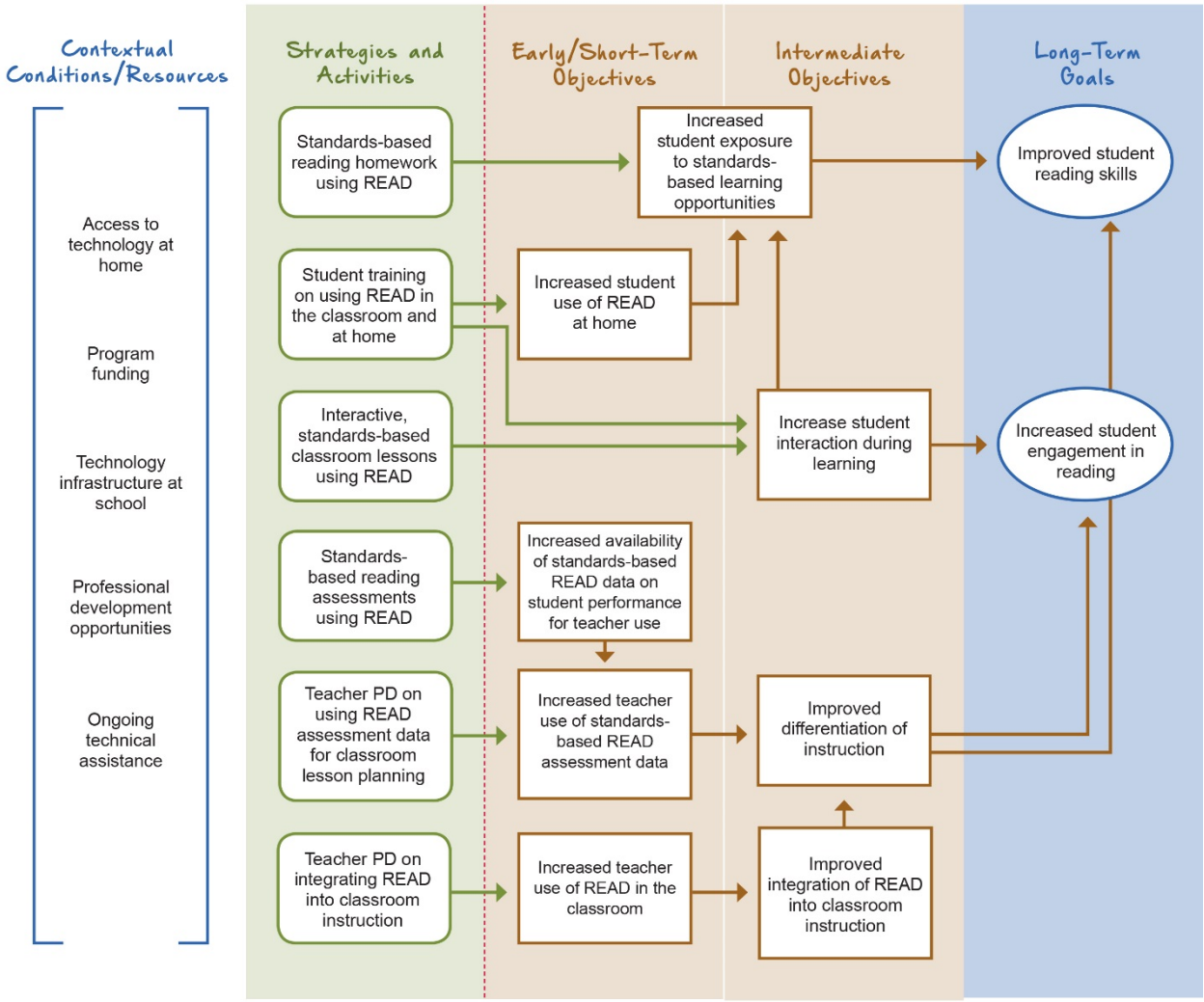
4. Availability of professional development for teachers on:
   a. Using interactive equipment in the classroom with the READ software; ongoing technical assistance from technology coordinator
   b. Integrating the READ software into their instruction
   c. Using READ assessment data for classroom lesson planning and differentiation of instruction
5. Availability of student training on how to use interactive equipment in the classroom, as well as how to use the READ software at home.
6. Student access to technology at home (computer with Internet connection).

Program Logic Model

At this point in the evaluation design, Dr. Elm recommended that the READ oversight team create an evaluation subcommittee, named the E-Team, comprised of 3-5 members. The evaluation subcommittee was formed as a partnership and a liaison between the READ program staff and the external evaluator, and was tasked with helping to design the evaluation and with monitoring the evaluation findings shared by the READ external evaluator. Mrs. Anderson appointed two oversight committee members (the district reading coach and one of the district reading specialists) to the E-Team. She also asked the district supervisor for assessment and evaluation to serve on the E-Team and to be the primary internal contact for the READ external evaluator. Finally, she invited Dr. Elm to serve as the chair of the E-Team and to serve as the lead, external evaluator of the READ program. As the external evaluator, Dr. Elm would conduct the evaluation and share findings with the E-Team and oversight team. The four-member E-Team's first task was to create the READ logic model.

Using the program definition developed by the oversight team, the E-Team worked to create a logic model. The E-Team started with the long-term goals on the right side of the model. The E-Team listed the contextual conditions and resources on the left. Just to the right of the context, the E-Team listed the strategies and activities. Next, the E-Team used the oversight team's assumptions to work through the early/short-term and intermediate objectives. The resulting logic model is provided in **Figure 3: READ Logic Model**.

Figure 3: READ Logic Model



Now that it had a draft logic model, the E-Team planned to share it with the oversight team in order to fine-tune, clarify, and finalize. Next, the oversight team and the E-Team would work together to develop evaluation questions.

Note: **Figure 6: Logic Model Template** is provided in Appendix E. Logic models can be created using the drawing template in a simple word processing application. There are also several applications available that are specifically tailored for creating logic models, and there are others that enable you to create a diagram using different shapes.

Step 2: Plan the Evaluation

At the next READ planning meeting, the E-Team shared the draft logic model with the full oversight team. Oversight team members reviewed the model and felt comfortable that it represented the assumptions and logic as they had agreed on at their last meeting. No changes were needed to the logic model at this time. Next, the E-Team and the oversight team used the logic model to develop evaluation questions for the READ program.

Evaluation Questions – Strategies and Activities

Using each of the strategies and activities listed on the left-hand side of the logic model, the E-Team worked with the READ oversight team to develop evaluation questions. For each strategy or activity, they developed questions addressing whether the strategy or activity had been carried out, as well as questions addressing some contextual conditions and resources necessary for program implementation. The READ E-Team and oversight team created six evaluation questions to assess READ strategies and activities.

Table 4: Evaluation Questions for Strategies and Activities

Strategies and Activities	Evaluation Questions
Interactive, standards-based classroom lessons (using the READ software with interactive classroom technologies and individual handheld mobile devices for each student)	To what extent did teachers have access to the necessary technology in the classroom to use READ in their instruction?
Standards-based reading assessments (Internet-based, formative assessments of student reading skills administered within the READ software)	To what extent were READ assessments made available to students and teachers? Examine overall, by school, and by grade level.

Strategies and Activities	Evaluation Questions
Standards-based reading homework (Internet-based using READ software)	To what extent did students have access to READ at home? Examine overall and by grade level, race, gender, and socioeconomic status.
Teacher professional development on integrating READ into classroom instruction (using an interactive wireless pad)	To what extent did teachers receive professional development on how to integrate READ into their classroom instruction?
Teacher professional development on using READ assessment data for classroom lesson planning	To what extent did teachers receive professional development on how to incorporate READ assessment data into their classroom lesson planning?
Student training on using READ (in the classroom and at home)	To what extent were students trained in how to use READ?

Note: These questions are intended to evaluate the degree to which the program had the opportunity to be successful, as well as to determine if additional program supports are needed for successful implementation.

Evaluation Questions – Early/Short-Term and Intermediate Objectives

Next, the E-Team worked with the READ oversight team to create several evaluation questions addressing READ early/short-term and intermediate objectives:

Table 5: Evaluation Questions for Early/Short-Term and Intermediate Objectives

Early/Short-Term and Intermediate Objectives	Evaluation Questions
Increased student use of READ at home (early/short-term)	How often did students receive READ homework assignments? To what extent did students complete READ homework assignments? **Note frequency and duration of use.
Increased teacher use of READ in the classroom (early/short-term)	In what ways and how often did teachers use READ in the classroom with students? **Note frequency, duration, and nature of use.

Early/Short-Term and Intermediate Objectives	Evaluation Questions
Increased student exposure to standards-based learning opportunities (early/short-term)	To what extent did students complete READ homework assignments? How often did teachers use READ in the classroom with students?
Increased availability of standards-based, formative READ assessment data on student reading performance (early/short-term)	How often did teachers access READ student assessment data? **Note frequency and type of access.
Increased teacher use of standards-based READ assessment data (early/short-term)	In what ways did teachers use READ student assessment data?
Increased student interaction during learning (intermediate)	To what extent and how did students interact during classroom instruction when READ was used? **Note frequency and type of interaction.
Improved integration of READ into classroom instruction (intermediate)	In what ways and to what extent did teachers integrate READ into their classroom instruction? **Note the quality with which READ was integrated into classroom instruction by teachers.
Improved differentiation of instruction (intermediate)	In what ways and to what extent did teachers use READ assessment data to plan and differentiate instruction? **Note what data were used and how data were used in instructional planning.

Evaluation Questions – Long-Term Goals

Finally, the E-Team and the READ oversight team created evaluation questions addressing READ long-term goals:

Table 6: Evaluation Questions for Long-Term Goals

Long-Term Goals	Evaluation Questions
Increased student engagement in reading	To what extent and in what ways did READ foster student engagement during reading lessons?
Improved student reading skills	To what extent did READ improve student learning in reading? • To what extent did student learning improve after READ was implemented? • To what extent did learning outcomes vary with teacher use of READ in the classroom? • To what extent did learning outcomes vary with teacher use of READ assessment data to plan and differentiate instruction? • How did student performance on the READ assessments correlate with student performance on state assessments? • In what ways did learning outcomes vary by initial reading performance on state assessments? • In what ways did learning outcomes vary by grade level? • In what ways did learning outcomes vary by special education status and English language proficiency? • In what ways did learning outcomes vary with the frequency of READ use at home?

Data Collection – Indicators and Targets

With the evaluation questions that the READ oversight team and E-Team had created, the E-Team was ready to expand on each question with indicators and accompanying targets. Using the logic model as its guide, the E-Team created the evaluation matrix below detailing the logic model components, associated evaluation questions, indicators, and accompanying targets.

For example, as the E-Team members began developing indicators for logic model components, they realized that student exposure to standards-based learning opportunities was an important construct composed of multiple components. Student exposure was assumed to

occur early on through READ homework assignments and teacher use of READ in the classroom, both of which the E-Team identified as indicators of exposure. Then at the intermediate stage, student interaction during classroom lessons using READ (measured by the rubric described below) was assumed to further increase student exposure.

The evaluation matrix is presented in three tables:

1. Strategies and Activities (**Table 7: Evaluation Matrix Addressing Strategies and Activities During the Initial Implementation—Indicators and Targets**)
2. Early/Short-Term and Intermediate Objectives (**Table 8: Evaluation Matrix Addressing Early/Short-Term and Intermediate Objectives—Indicators and Targets**)
3. Long-term Goals (**Table 9: Evaluation Matrix Addressing Long-Term Goals—Indicators and Targets**)

As you read through the tables, you will see that the evaluation will collect much of the data through the four instruments described below: (1) the READ implementation rubric, (2) the teacher survey, (3) the READ student assessment, and (4) the annual state student reading assessment.

1. *READ implementation rubric:* The E-Team created the READ implementation rubric to examine the quality of teacher practice when using READ during classroom instruction, student interaction during learning, teacher integration of READ into classroom instruction, and teacher use of READ student assessment data to plan and differentiate instruction. Dr. Elm will administer the READ implementation rubric on a monthly basis, alternating between classroom observations one month and interviews with teachers the following month.
2. *Teacher survey:* Dr. Elm will conduct the teacher survey in October as a baseline and again in December, February, April, and June. The teacher survey has multiple sections including some open-ended questions. Most items will be included every time the survey is administered. Others (such as items on the initial account setup and access) will be administered only when appropriate.
3. *READ student assessment:* The READ software itself includes an embedded, formative, standards-based READ assessment to measure student learning before and after each lesson. The data from the embedded READ assessment are stored in the READ system for teachers to use in assessing student learning and planning their instruction. The evaluation also will use these data.
4. *State student reading assessment:* The evaluation also will use state reading assessment scores to measure student learning in reading. The state reading assessment is administered in April of each academic year. Reading scores from the spring prior to READ program implementation will be used as a baseline against which to measure student reading improvement.

The evaluation will use data collected using each of these four instruments—the READ implementation rubric, teacher survey, READ student assessment, and reading scores from the state assessment. READ student assessment data and state student reading assessment data will be disaggregated and examined by quality of teacher use (using the READ implementation rubric), frequency of home use, initial reading performance on state assessments, grade level, gender, ethnicity, special education status, and English language proficiency. In addition, reading scores on the READ student assessment will be analyzed in relation to the state assessment reading scores to determine the degree to which the READ assessments correlate with the state reading assessment.

Tables 7, 8, and 9 show the evaluation questions, indicators, and targets developed in **Step 2: Plan the Evaluation**.

Table 7: Evaluation Matrix Addressing Strategies and Activities During the Initial Implementation—Indicators and Targets

Logic Model Components	Evaluation Questions	Indicators	Targets
Interactive, standards-based classroom lessons (using the READ software with interactive classroom technologies and individual handheld mobile devices for each student)	To what extent did teachers have access to the necessary technology in the classroom to use READ in their instruction?	Increased number of teachers with access to the necessary technology in their classroom to use READ	By the start of the school year, all teachers will have the necessary technology in their classroom to use READ.
Standards-based reading assessments (Internet-based, formative assessments of student reading skills administered within the READ software)	To what extent were READ assessments made available to students and teachers? Examine overall, by school, and by grade level.	Increased number of teachers with access to READ assessments Increased number of students with access to READ assessments	By the start of the school year, all teacher accounts will have been set up in READ. By the end of September, all student accounts will have been set up in READ.
Standards-based reading homework (Internet-based using READ software)	To what extent did students have access to READ at home? Examine overall and by grade level, race, gender, and socioeconomic status.	Increased number of students with access to READ at home	By the end of September, all teachers will have determined how many students have the technology necessary to access READ from home.

Logic Model Components	Evaluation Questions	Indicators	Targets
Teacher professional development on integrating READ into classroom instruction (using an interactive wireless pad)	To what extent did teachers receive professional development on how to integrate READ into their classroom instruction?	Increased number of teachers trained in how to effectively use READ in their classroom instruction	By the start of the school year, all teachers will have received professional development on how to integrate READ into their classroom instruction.
Teacher professional development on using READ assessment data for classroom lesson planning	To what extent did teachers receive professional development on how to incorporate READ assessment data into their classroom lesson planning?	Increased number of teachers trained in how to use READ assessment data in their lesson planning	By the start of the school year, all teachers will have received professional development on how to use READ assessment data in their lesson planning.
Student training on using READ (in the classroom and at home)	To what extent were students trained in how to use READ?	Increased number of students trained in how to use READ	By the end of September, all teachers will have trained their students in the use of READ (for use in the classroom and at home).

Table 8: Evaluation Matrix Addressing Early/Short-Term and Intermediate Objectives—Indicators and Targets

Logic Model Components	Evaluation Questions	Indicators	Targets
Increased student use of READ at home (early/short-term)	How often did students receive READ homework assignments? To what extent did students complete READ homework assignments? **Note frequency and duration of use.	Increased number of teachers assigning READ homework Increased number of students completing READ homework, within a reasonable time	By November, over 50% of teachers will be assigning weekly READ homework. By December, over 50% of students will be completing weekly READ homework assignments. Students will spend no more than 20 minutes to complete READ homework. (Note: Completion rates and duration of use are available through the READ online system.)

Logic Model Components	Evaluation Questions	Indicators	Targets
Increased teacher use of READ in the classroom (early/short-term)	In what ways and how often did teachers use READ in the classroom with students? **Note frequency, duration, and nature of use.	Increased number of teachers using READ in the classroom with students Improved teacher use of READ in the classroom with students	By October, all teachers will be using READ in the classroom with students. By November, 25% of teachers will score a 2 or above (out of 4) on the READ implementation rubric. By December, 50% of teachers will score a 2 or above on the READ implementation rubric.
Increased student exposure to standards-based learning opportunities (early/short-term)	To what extent did students complete READ homework assignments?	Increased number of students completing READ homework	By December, over 50% of students will be completing weekly READ homework assignments.
	How often did teachers use READ in the classroom with students?	Increased number of teachers using READ in the classroom with students	By October, all teachers will be using READ in the classroom with students.
Increased availability of standards-based, formative READ assessment data on student reading performance (early/short-term)	How often did teachers access READ student assessment data? **Note frequency and type of access.	Increased number of teachers accessing READ student assessment data	By October, 50% of teachers will have accessed READ student assessment data. By November, all teachers will have accessed READ student assessment data.
Increased teacher use of standards-based READ assessment data (early/short-term)	In what ways did teachers use READ student assessment data?	Improved teacher use of READ student assessment data	By February, 25% of teachers will score a 3 or above on the READ implementation rubric.
Increased student interaction during learning (intermediate)	To what extent and how did students interact during classroom instruction when READ was used? **Note frequency and type of interaction.	Increased student interaction during learning, as measured by the student component of the READ implementation rubric (rubric completed through classroom observations and teacher interviews)	By February, 25% of classrooms will have a score of 3 or above (out of 4) on the READ implementation rubric. By April, 50% of teachers will score a 3 or above on the READ implementation rubric.

Logic Model Components	Evaluation Questions	Indicators	Targets
Improved integration of READ into classroom instruction (intermediate)	In what ways and to what extent did teachers integrate READ into their classroom instruction? **Note the quality with which READ was integrated into classroom instruction by teachers.	Improved integration of READ lessons into classroom instruction, as measured by teacher scores on the READ implementation rubric (rubric completed through classroom observations and teacher interviews)	By April, 50% of teachers will score a 3 or above on the READ implementation rubric. By June, 75% of teachers will score a 3 or above and 25% of teachers will score a 4 on the READ implementation rubric.
Improved differentiation of instruction (intermediate)	In what ways and to what extent did teachers use READ assessment data to plan and differentiate instruction? **Note what data were used and how data were used in instructional planning.	Increased number of teachers using READ assessment data to plan instruction Improved use of READ assessment data to differentiate instruction	By December, all teachers will be using READ assessment data on a weekly basis to plan instruction. By April, 50% of teachers will score a 3 or above on the READ implementation rubric. By June, 75% of teachers will score a 3 or above and 25% of teachers will score a 4 on the READ implementation rubric.

Table 9: Evaluation Matrix Addressing Long-Term Goals—Indicators and Targets

Logic Model Components	Evaluation Questions	Indicators	Targets
Increased student engagement in reading	In what ways did READ foster student engagement during reading lessons?	Increased frequency and improved quality of student engagement in the classroom, as measured by the READ implementation rubric	By February, 25% of classrooms will have a score of 3 or above (out of 4) on the READ implementation rubric. By April, 50% of classrooms will score a 3 or above on the READ implementation rubric. By June, 75% of classrooms will score a 3 or above and 25% of teachers will score a 4 on the READ implementation rubric.

Logic Model Components	Evaluation Questions	Indicators	Targets
Improved student reading skills	To what extent did READ improve student learning in reading? → To what extent did student learning improve after READ was implemented? → To what extent did learning outcomes vary with teacher use of READ in the classroom? → To what extent did learning outcomes vary with teacher use of READ assessment data to plan and differentiate instruction? → How did student performance on the READ assessments correlate with student performance on state assessments? → In what ways did learning outcomes vary by initial reading performance on state assessments? → In what ways did learning outcomes vary by grade level? → In what ways did learning outcomes vary by special education status and English language proficiency? → In what ways did learning outcomes vary with the frequency of READ use at home?	Increased scores on tests assessing students' reading ability (including both state assessments and the formative assessments provided within the READ software)	Within 2 years, the increase in student scores on the state standards-based reading assessment will be statistically significant for those students who participated in READ versus those students who did not participate in READ. State reading scores and READ assessment data will be disaggregated and examined by quality of READ teacher use (using the READ implementation rubric), frequency of READ home use, initial reading performance on state assessments, grade level, gender, ethnicity, special education status, and English language proficiency. Reading scores on the state assessment will be analyzed in relation to scores on the READ assessment data, in order to determine the degree to which READ assessments correlate with the state assessments.

Evaluation Design

Grovemont School District had 80 third- through fifth-grade classrooms across six elementary schools (28 third-grade classrooms, 28 fourth-grade classrooms, and 24 fifth-grade classrooms). District class size for grades 3 through 5 ranged from 22 to 25 students per classroom. Because of state budget cuts and reduced funding for the program, the E-Team knew that Mrs. Anderson and the READ oversight team would have to make some difficult choices about how to structure and evaluate their program.

Some members of the oversight team wanted to implement the program in fifth grade only for the first year, and then reexamine funds to see if they might be able to expand down to fourth grade in Year 2. Others voted to start the program at two of the six elementary schools and then try to include an additional school in Year 2. Dr. Elm and the E-Team recommended that they consider partially implementing the program at all six schools and across all three grades.

Dr. Elm explained that they would receive much better information about how their program was working and, more importantly, how it could be improved, if they were able to compare results from those classrooms that were using the program with those that were not. Dr. Elm knew that students at all of the schools in Grovemont School District were randomly assigned to teachers during the summer before each school year. However, Dr. Elm explained that in order to minimize initial differences between those classrooms that participate in READ and those that do not, they should consider randomly assigning half of the classrooms to continue with the existing district curriculum while the other half would supplement their existing curriculum with the READ program. Dr. Elm also recommended that they first divide the classrooms by school and grade level so that each school and grade would have one half of the classrooms assigned to the program. Teachers whose classrooms were not assigned to the program would be assured that if the program proved successful, they would be on board by Year 3. However, if the program did not have sufficient benefits for the students, it would be discontinued in all classrooms after Year 2. Dr. Elm concluded that building a strong evaluation into their program would provide them with credible information as to how their program was working and that having data to direct their program adjustments and improvements would give the program the best opportunity to be successful. The READ oversight team agreed to think about this idea and reconvene in 1 week to make a decision.

The E-Team also distributed the evaluation matrix it had created based on the READ logic model. The E-Team asked the oversight team to review the matrix and provide any feedback or comments.

The following week, the E-Team and READ oversight team reconvened to decide how to structure the program and to work on the evaluation design. Mrs. Anderson had spoken with the district superintendent about the evaluator's suggestion of implementing READ in half the district's third- through fifth-grade classrooms, with the promise that it would be expanded to all classrooms in Year 3 if the program was successful. Although logistically it would be easier to implement the program in two or three schools or one or two grades than to implement it in half the classrooms in all schools and at all grades, the superintendent understood the benefit of the added effort. The evaluation would provide higher quality data to inform decisions for program improvement and decisions regarding the program's future.

Mrs. Anderson shared the superintendent's comments with the oversight team and evaluation subcommittee. Like the superintendent, team members felt conflicted by the choice between

simpler logistics or a stronger evaluation design. Dr. Elm understood the dilemma all too well, but as an evaluator and an educator, she believed that a strong evaluation would result in improved program implementation and improved program outcomes.

Dr. Elm recognized that implementing the program in all classrooms in one grade level across the district would offer the weakest evaluation design and the least useful information but would likely be the simplest option logistically. Another option would be to start the program in all classrooms at two or three schools. In such a case, the other schools could be used as comparisons. For this reason, Dr. Elm explored the comparability of the six elementary schools in case the team decided to go that route. Five of the elementary schools had somewhat comparable state test scores in reading, while the sixth school had lower state test scores, and the difference was statistically significant. In addition, Schools 1 through 5 had similar (and fairly homogenous) populations, while School 6 had a much lower socioeconomic student population and a much higher percentage of ELL students. Because the district was interested in how the program worked with ELL students, the team knew that the evaluation needed to include School 6. However, if School 6 were used in a three-school implementation, the team would not have a comparable school against which to benchmark its results.

While not the simplest option, the oversight team decided that its best option would be to structure the program in such a way as to maximize the quality of the information from the evaluation. The team chose to build a strong evaluation into the READ program design to provide the formative information needed for program improvement and valid summative information for accountability.

Based on the READ oversight team's decision about how to structure the program, Dr. Elm and the E-Team drafted the following evaluation design. They presented the design at the next oversight team meeting. The oversight team voted to approve the design as follows:

> **Design:** Multiple-group, experimental design (students randomly assigned to classrooms by the school prior to the start of the school year and classrooms randomly assigned to the READ program group or a non-READ comparison group).
>
> Program group (READ): 40 classrooms (22 to 25 students per classroom).
>
> Comparison group (non-READ): 40 classrooms (22 to 25 students per classroom).
>
> Classrooms will be stratified by grade level within a school and randomly assigned to either the READ program group or a comparison group. The READ and non-READ groups will each include 14 third-grade classrooms, 14 fourth-grade classrooms, and 12 fifth-grade classrooms.

Enriching the evaluation design: Program theory and logic modeling will be used to examine program implementation as well as short-term, intermediate, and long-term outcomes.

Data Collection Methods

The E-Team decided on data collection methods, including the data sources, for each evaluation question and associated indicators. Two examples are provided below.

1. In what ways and to what extent did teachers integrate READ into their classroom instruction?
   - → A READ rubric will be used to measure teacher implementation of READ in the classroom.
   - → The rubric will be completed through classroom observations and teacher interviews.
   - → The READ implementation rubric will be on a 4-point scale, with a 4 representing the best implementation.
   - → Data will be collected monthly, alternating between classroom observations one month and interviews the following month.
2. To what extent did READ improve student learning in reading?
   - → The state reading assessment will be used to measure student learning in reading. It is administered in April of each academic year, beginning in second grade.
   - → READ assessment data will be used as a formative measure to examine student reading performance.
   - → State reading scores and READ assessment data will be disaggregated and examined by quality of teacher use (using the READ implementation rubric), frequency of home use, initial reading performance, grade level, gender, ethnicity, special education status, and English language proficiency.
   - → Previous year state reading assessment scores will be used as a baseline against which to measure student reading improvement.
   - → Reading scores on the state assessment will be analyzed in relation to scores on the READ assessments in order to determine the degree to which READ assessments correlate with the state reading assessment.

For a full list of evaluation questions, data sources, and data collection methods, see the READ Evaluation Matrix tables 10, 11, and 12 in Step 3.

Step 3: Implement the Evaluation

The READ external evaluator collected a mix of quantitative and qualitative data to address evaluation questions. Qualitative data collected through observations and interviews were coded using the READ implementation rubric and analyzed using descriptive statistics, including means and frequency distributions. Student reading assessment data were analyzed by testing for statistical significance, comparing mean test scores between groups of students and over time.

The following is an example using one of the READ intermediate objectives:

1. **Logic Model Component:** Improved integration of READ into classroom instruction (intermediate objective).
2. **Evaluation Question:** In what ways and to what extent did teachers integrate READ into their classroom instruction?
3. **Indicator:** Improved integration of READ lessons into classroom instruction.
4. **Targets:** By April, 50% of teachers will score a 3 or above (out of 4) on the READ implementation rubric. By June, 75% of teachers will score a 3 or above on the READ implementation rubric.
5. **Data Source:** READ implementation rubric (developed by the E-Team and administered by Dr. Elm)
6. **Data Collection:** Rubric completed through alternating, monthly classroom observations and teacher interviews.
7. **Data Analysis:** Rubric scores aggregated into frequency distributions and means; change over time to be analyzed.

The full READ Evaluation Matrix is included in tables 10, 11, and 12. Note that the evaluation matrix was completed in steps. The logic model components are taken directly from the READ logic model created in **Step 1: Define the Program**. The logic model components consist of strategies and activities, early/short-term and intermediate objectives, and long-term goals. The evaluation questions were created in **Step 2: Plan the Evaluation**, guided by the READ logic model. Indicators and targets were derived in Step 2 using the READ logic model and evaluation questions. At the end of Step 2, data collection sources and methods were chosen for each READ indicator. Data analysis methods were determined in **Step 3: Implement the Evaluation**. (See Appendix E for **Table 26: Evaluation Matrix Template**.)

Table 10: READ Evaluation Matrix—Strategies and Activities/Initial Implementation

Logic Model Components	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
Interactive, standards-based classroom lessons (using the READ software with interactive classroom technologies and individual hand-held mobile devices for each student)	To what extent did teachers have access to the necessary technology in the classroom to use READ in their instruction?	Increased number of teachers with access to the necessary technology in their classroom to use READ	By the start of the school year, all teachers will have the necessary technology in their classroom to use READ.	Technology installation records; teacher survey	Technology installation records examined in September for evidence of necessary classroom technology Teacher survey administered in October, including items on technology in the classroom	Records analyzed with basic descriptive statistics (counts and percentages) of classrooms with necessary technology Teacher survey analyzed with basic descriptive statistics including means and frequency distributions; open-ended items on the survey summarized, and if warranted, analyzed for themes

Logic Model	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
Standards-based reading assessments (Internet-based, formative assessments of student reading skills administered within the READ software)	To what extent were READ assessments made available to students and teachers?	Increased number of teachers with access to READ assessments Increased number of students with access to READ assessments	By the start of the school year, all teacher accounts will have been set up in READ. By the end of September, all student accounts will have been set up in READ.	Technology records (teacher accounts); teacher survey Technology records (student accounts)	Technology records examined in September for evidence of teacher and student account setup/activation Teacher survey administered in October (includes items on teacher account setup/activation)	Records analyzed with basic descriptive statistics (counts and percentages) on setup/activated teacher and student accounts Teacher survey analyzed with basic descriptive statistics on setup/activated accounts
Standards-based reading homework (Internet-based using READ software)	To what extent did students have access to READ at home?	Increased number of students with access to READ at home	By the end of September, all teachers will have determined how many students have the technology necessary to access READ from home.	Teacher survey	Teacher survey to be administered in October (includes items on student technology availability)	Teacher survey analyzed with basic descriptive statistics including means and frequency distributions; open-ended items summarized, and if warranted, analyzed for themes

Logic Model Components	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
Teacher professional development on integrating READ into classroom instruction (using an interactive wireless pad)	To what extent did teachers receive professional development on how to integrate READ into their classroom instruction?	Increased number of teachers trained in how to effectively use READ in their classroom instruction	By the start of the school year, all teachers will have received professional development on how to integrate READ into their classroom instruction.	Professional development records Teacher survey	Professional development examined in September for evidence it was offered, as well as attendance Teacher survey administered in October (includes items on professional development)	Records summarized and analyzed with basic descriptive statistics, where appropriate Teacher survey analyzed as described above
Teacher professional development on using READ assessment data for classroom lesson planning	To what extent did teachers receive professional development on how to incorporate READ assessment data into their classroom lesson planning?	Increased number of teachers trained in how to use READ assessment data in their lesson planning	By the start of the school year, all teachers will have received professional development on how to use READ assessment data in their lesson planning.	Professional development records Teacher survey	Professional development examined for offerings, as well as attendance Teacher survey administered in October (includes items on professional development)	Records summarized and analyzed with basic descriptive statistics, where appropriate Teacher survey analyzed as described above
Student training on using READ (in the classroom and at home)	To what extent were students trained in how to use READ?	Increased number of students trained in how to use READ.	By the end of September, all teachers will have trained their students in the use of READ.	Teacher survey	Teacher survey administered in October (includes items on the training of students)	Teacher survey analyzed as described above

Table 11: Evaluation Matrix—Early/Short-Term and Intermediate Objectives

Logic Model Components	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
Increased student use of READ at home (early/short-term)	How often did students receive READ homework assignments?	Increased number of teachers assigning READ homework	By November, over 50% of teachers will be assigning weekly READ homework.	Teacher survey	Teacher survey, including items on teacher practice regarding homework	Teacher survey analyzed with basic descriptive statistics including means and frequency distributions; open-ended items summarized, and if warranted, analyzed for themes
	To what extent did students complete READ homework assignments? **Note frequency and duration of use.	Increased number of students completing READ homework within a reasonable time	By December, over 50% of students will be completing weekly READ homework assignments. Students will spend no more than 20 minutes to complete weekly READ homework assignments. (Note: Completion rates and duration of use are available through the READ online system.)	READ online records	READ online records examined monthly for evidence of student use	READ online records analyzed with basic descriptive statistics

Logic Model Components	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
Increased teacher use of READ in the classroom (early/short-term)	In what ways and how often did teachers use READ in the classroom with students? **Note frequency, duration, and nature of use.	Improved teacher use of READ in the classroom with students	By November, 25% of teachers will score a 2 or above (out of 4) on the READ implementation rubric. By December, 50% of teachers will score a 2 or above on the READ implementation rubric.	READ implementation rubric	Rubric data collected monthly (for each teacher), alternating between classroom observation and teacher interviews	Rubric data analyzed by means and frequency distributions of rubric scores; change over time analyzed by testing for statistical significance
Increased student exposure to standards-based learning opportunities (early/short-term)	In what ways and how often did teachers use READ in the classroom with students? **Note frequency, duration, and nature of use	Increased number of teachers using READ in the classroom with students	By October, all teachers will be using READ in the classroom with students.	Teacher survey	Teacher survey, including items on teacher practice regarding classroom use	Teacher survey analyzed with basic descriptive statistics including means and frequency distributions; open-ended items summarized, and if warranted, analyzed for themes

Logic Model Components	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
	To what extent did students complete READ homework assignments? **Note frequency and duration of use.	Increased number of students completing READ homework, within a reasonable time	By December, over 50% of students will be completing weekly READ homework assignments. Students will spend no more than 20 minutes to complete weekly READ homework assignments.	READ online records	READ online records examined monthly for evidence of student use	READ online records analyzed with basic descriptive statistics
Increased availability of standards-based, formative READ assessment data on student reading performance (early/short-term)	How often did teachers access READ student assessment data? **Note frequency and type of access.	Increased number of teachers accessing READ student assessment data	By October, 50% of teachers will have accessed READ student assessment data. By November, all teachers will have accessed READ student assessment data.	READ online records	READ online records examined monthly to determine access patterns	READ online records analyzed with basic descriptive statistics

Logic Model Components	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
Increased teacher use of standards-based READ assessment data (early/short-term)	In what ways did teachers use READ student assessment data?	Improved teacher use of READ student assessment data	By February, 25% of teachers will score a 3 or above on the READ implementation rubric.	READ implementation rubric	Rubric data collected monthly (for each teacher), alternating between classroom observation and teacher interviews	Rubric data analyzed by means and frequency distributions of scores; change over time analyzed by testing for statistical significance
Increased student interaction during learning (intermediate)	To what extent and how did students interact during classroom instruction when READ was used? **Note frequency and type of interaction.	Increased student interaction during learning	By February, 25% of classrooms will score of 3 or above on the READ implementation rubric. By April, 50% of teachers will score a 3 or above on the rubric.	READ implementation rubric	Rubric data collected monthly (for each teacher), alternating between classroom observation and teacher interviews	Rubric data analyzed by means and frequency distributions of rubric scores; change over time analyzed by testing for statistical significance
Improved integration of READ into classroom instruction (intermediate)	In what ways and to what extent did teachers integrate READ into their classroom instruction? **Note the quality with which READ was integrated into classroom	Improved integration of READ lessons into classroom instruction	By April, 50% of teachers will score a 3 or above on the READ implementation rubric. By June, 75% of teachers will score a 3 or above and 25% of teachers	READ implementation rubric	Rubric data collected monthly (for each teacher), alternating between classroom observation and teacher interviews	Rubric data analyzed by means and frequency distributions of rubric scores; change over time analyzed by testing for statistical significance

Logic Model Components	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
	instruction by teachers.		will score a 4 on the rubric.			
Improved differentiation of instruction (intermediate)	In what ways and to what extent did teachers use READ assessment data to plan and differentiate instruction? **Note what data were used and how data were used in instructional planning.	Increased number of teachers using READ assessment data to plan instruction	By December, all teachers will be using READ assessment data on a weekly basis to plan instruction. By April, 50% of teachers will score a 3 or above on the READ implementation rubric. By June, 75% of teachers will score a 3 or above and 25% of teachers will score a 4 on the READ implementation rubric.	Teacher survey	Teacher survey, including items on teacher practice regarding use of READ assessment data	Teacher survey analyzed with basic descriptive statistics including means and frequency distributions; open-ended items summarized, and if warranted, analyzed for themes

Logic Model Components	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
		Improved use of READ assessment data to differentiate instruction	By December, all teachers will be using READ assessment data on a weekly basis to plan instruction. By April, 50% of teachers will score a 3 or above on the READ implementation rubric. By June, 75% of teachers will score a 3 or above and 25% of teachers will score a 4 on the READ implementation rubric.	READ implementation rubric	Rubric data collected monthly (for each teacher), alternating between classroom observation and teacher interviews	Rubric data analyzed by means and frequency distributions of rubric scores; change over time analyzed by testing for statistical significance

Table 12: Evaluation Matrix—Long-Term Goals

Logic Model Components	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
Increased student engagement in reading	To what extent and in what ways did READ foster student engagement during reading lessons?	Increased frequency and improved quality of student engagement in the classroom, as measured by the READ implementation rubric	By February, 25% of classrooms will have a score of 3 or above (out of 4) on the READ implementation rubric. By April, 50% of classrooms will score a 3 or above on the READ implementation rubric. By June, 75% of classrooms will score a 3 or above and 25% of teachers will score a 4 on the READ implementation rubric.	READ implementation rubric	Monthly, alternating between classroom observation and teacher interviews	Means and frequency distributions of READ rubric scores determined; change over time analyzed using significance testing

Improved student reading skills	To what extent did READ improve student learning in reading? To what extent did student learning improve after READ was implemented? To what extent did learning outcomes vary with teacher use of READ in the classroom? To what extent did learning outcomes vary with teacher use of READ assessment data to plan and differentiate instruction? How did student performance on the READ assessments correlate with student performance on state assessments? In what ways did learning	Increased scores on tests assessing students' reading ability (including both state assessments and formative assessments provided within the READ software)	Within 2 years, the increase in student scores on the state standards-based reading assessment will be statistically significant for those students who participated in READ versus those students who did not participate in READ. State and READ assessment data will be disaggregated and examined by quality of teacher use (using the READ implementation rubric), frequency of home use, initial reading performance, grade level, gender, ethnicity, special education status, and English language proficiency. Reading scores on the state assessment will be	State reading assessment data READ assessment data READ implementa-tion rubric Teacher survey Demographic data from school records READ online records	April of each academic year	T-test of mean test scores (on state reading assessment and READ assessments) between READ and non-READ students, taking into account prior reading performance on the state reading assessment; results disaggregated by teacher use, grade, gender, race, English language proficiency; correlational testing of state reading test scores and READ assessment scores

Logic Model Components	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
	outcomes vary by initial reading performance, grade level, special education status, and English language proficiency? In what ways did learning outcomes vary with the frequency of READ use at home?		analyzed in relation to scores on the READ assessment data, in order to determine the degree to which READ assessments correlate with the state reading assessments.			

All data collected through the evaluation were managed and stored by Dr. Elm, the external evaluator. The computer used for storage and analysis was located in a locked office. Only the external evaluator had access to the raw data. Data were backed up weekly to an external drive, which was kept in a locked drawer. To protect teacher and student privacy, identification numbers were assigned to all participants. Teacher and student names were not recorded with the data.

READ online records regarding student and teacher use, rubric data, and survey data were only accessible by the external evaluator. Results that were released were only in aggregation and had no identifying information. All evaluation data were secured and kept confidential to protect individual privacy.

Step 4: Interpret the Results

The READ evaluation subcommittee (the E-Team) examined the evaluation results. Some highlights from the findings are provided below. Use of these findings will be discussed in **Step 5: Inform and Refine – Using the Results**.

Summative

1. First-year results indicated that READ and state reading scores for READ students were higher than those for non-READ students. The gains were especially compelling for students in classrooms in which READ was used regularly and with fidelity, where increases in reading scores were over three times that of non-READ students.
2. Students in classrooms where READ was used regularly and with fidelity increased their reading scores by twice that of students in READ classrooms where READ was used minimally.
3. Students of teachers who used READ assessment data as intended to differentiate instruction increased their reading scores on the state assessment by twice as much as students of teachers who did not use READ assessment data as intended.
4. Student scores on READ assessments had a statistically significant and strong positive correlation with student scores on the state reading assessment, indicating that these two assessments are likely well aligned and that READ assessment data are likely a good indicator of performance on the state reading assessment.

Formative

5. Student assessment data could not be analyzed by home use of the READ program because only one classroom implemented the home component.
6. At the start of the year, teacher use of READ was promising, and the program met its targets. However, as the program progressed and as more teachers were pressed to

improve their use of READ, several targets were not met. READ student assessment data were not used as regularly by teachers as the classroom component of READ.

A full accounting of evaluation results by logic model component and evaluation question is provided in tables 13, 14, and 15.

Table 13: READ Evaluation Results—Strategies and Activities/Initial Implementation

Logic Model Components	Evaluation Questions	Indicators	Targets	Evaluation Findings
Interactive, standards-based classroom lessons (using the READ software with interactive classroom technologies and individual handheld mobile devices for each student)	To what extent did teachers have access to the necessary technology in the classroom to use READ in their instruction?	Increased number of teachers with access to the necessary technology in their classroom to use READ	By the start of the school year, all teachers will have the necessary technology in their classroom to use READ.	☑ By September, all READ teachers (100%) had the necessary technology in their classroom to use READ.
Standards-based reading assessments (Internet-based, formative assessments of student reading skills administered within the READ software)	To what extent were READ assessments made available to students and teachers?	Increased number of teachers with access to READ assessments Increased number of students with access to READ assessments	By the start of the school year, all teacher accounts will have been set up in READ. By the end of September, all student accounts will have been set up in READ.	☑ By September, all READ teacher accounts (100%) had been set up. ☑ By the end of September, all student accounts (100%) had been set up.

Logic Model Components	Evaluation Questions	Indicators	Targets	Evaluation Findings
Standards-based reading homework (Internet-based using READ software)	To what extent did students have access to READ at home?	Increased number of students with access to READ at home	By the end of September, all teachers will have determined how many students have the technology necessary to access READ from home.	☑ By the end of September, all teachers (100%) had determined how many students had home access to READ. → At three of the schools, most students (90%) had home access to READ. → At two of the schools, about half the students (54%) had home access to READ. → At one school, less than 20% of students had the technology necessary to access READ from home. → Even within those schools that had high numbers of students with home access, the classroom variability was large. Only one of the 40 classrooms had 100% of students with access to READ from home.
Teacher professional development on integrating READ into classroom instruction (using an interactive wireless pad)	To what extent did teachers receive professional development on how to integrate READ into their classroom instruction?	Increased number of teachers trained in how to effectively use READ in their classroom instruction	By the start of the school year, all teachers will have received professional development on how to integrate READ into their classroom instruction.	☑ By September, all teachers (100%) had received professional development on how to integrate READ into their classroom instruction.

Logic Model Components	Evaluation Questions	Indicators	Targets	Evaluation Findings
Teacher professional development on using READ assessment data for classroom lesson planning	To what extent did teachers receive professional development on how to incorporate READ assessment data into their classroom lesson planning?	Increased number of teachers trained in how to use READ assessment data in their lesson planning	By the start of the school year, all teachers will have received professional development on how to use READ assessment data in their lesson planning.	☑ By September, all teachers (100%) had received professional development on how to use READ assessment data in their classroom lesson planning.
Student training on using READ (in the classroom and at home)	To what extent were students trained in how to use READ?	Increased number of students trained in how to use READ	By the end of September, all teachers will have trained their students in the use of READ (for use in the classroom and at home).	☑ By October, all teachers (100%) had trained their students in the classroom use of READ. ☒ By the end of October, only one teacher (2%) had trained his students in the home use of READ. → Open-ended survey items revealed that teachers chose not to train students in the home use of READ unless every student in the classroom was able to take advantage of the home component. Since only one classroom had 100% participation, only one teacher trained students on home use.

Table 14: READ Evaluation Results—Early/Short-Term and Intermediate Objectives

Logic Model Components	Evaluation Questions	Indicators	Targets	Evaluation Findings
Increased student use of READ at home (early/short-term)	How often did students receive READ homework assignments?	Increased number of teachers assigning READ homework	By November, over 50% of teachers will be assigning weekly READ homework.	☑ By November, only one teacher (2%) was assigning weekly READ homework.
	To what extent did students complete READ homework assignments? **Note frequency and duration of use.	Increased number of students completing READ homework within a reasonable time	By December, over 50% of students will be completing weekly READ homework. Students will spend no more than 20 minutes to complete weekly READ homework. (Note: Completion rates and duration of use are available through the READ online system.)	☑ By December, most students (70%) in the classroom where the READ homework component was used were completing the assignment. These students spent, on average, 15 minutes to complete the weekly READ homework.
Increased teacher use of READ in the classroom (early/short-term)	In what ways and how often did teachers use READ in the classroom with students? **Note frequency, duration, and nature of use.	Improved teacher use of READ in the classroom with students	By November, 25% of teachers will score a 2 or above (out of 4) on the READ implementation rubric. By January, 50% of teachers will score a 2 or above on the READ implementation rubric.	☑ By November, one-third of teachers (33%) scored a 2 or above on the READ implementation rubric. By January, over half the teachers (58%) scored a 2 or above on the READ implementation rubric.

Logic Model Components	Evaluation Questions	Indicators	Targets	Evaluation Findings
Increased student exposure to standards-based learning opportunities (early/short-term)	In what ways and how often did teachers use READ in the classroom with students? **Note frequency, duration, and nature of use.	Increased number of teachers using READ in the classroom with students	By October, all teachers will be using READ in the classroom with students.	☑ By October, all teachers (100%) reported some classroom use of READ, while a little over half (58%) reported regular (at least weekly) use of READ in the classroom. By October, over one-quarter of teachers (30%) reported that they had only used READ in the classroom once or twice in the last month.
	To what extent did students complete READ homework assignments? **Note frequency and duration of use.	Increased number of students completing READ homework within a reasonable time	By December, over 50% of students will be completing weekly READ homework. Students will spend no more than 20 minutes to complete weekly READ homework. (Note: Completion rates and duration of use are available through the READ online system.)	→ By December, most students (70%) in the classroom where the READ homework component was used were completing the assignment. These students spent, on average, 15 minutes to complete the weekly READ homework.
Increased availability of READ standards-based, formative assessment data on student reading performance (early/short-term)	How often did teachers access READ student assessment data? **Note frequency and type of access.	Increased number of teachers accessing READ student assessment data	By October, 50% of teachers will have accessed READ student assessment data. By November, all teachers will have accessed READ student assessment data.	☑ By October, over half the teachers (58%) had accessed the READ student assessment data. By November, one-fifth of teachers (20%) had not accessed the READ student assessment data. Thirty-two teachers (80%) had accessed the READ assessment data, while over half (58%) accessed the READ assessment data on a regular (at least weekly) basis.

Logic Model Components	Evaluation Questions	Indicators	Targets	Evaluation Findings
Increased teacher use of standards-based READ assessment data (early/short-term)	In what ways did teachers use READ student assessment data?	Improved teacher use of READ student assessment data	By January, 25% of teachers will score a 3 or above (out of 4) on the READ implementation rubric.	☑ In January, 11 teachers (28%) scored a 3 or above on the READ implementation rubric. → Eight teachers (20%) scored a 1 on the READ implementation rubric.
Increased student interaction during learning (intermediate)	To what extent and how did students interact during classroom instruction when READ was used? **Note frequency and type of interaction.	Increased student interaction during learning	By February, 25% of classrooms will have a score of 3 or above (out of 4) on the READ implementation rubric. By April, 50% of classrooms will score a 3 or above on the READ rubric.	☑ In February, 11 classrooms (28%) scored a 3 or above on the READ implementation rubric. ☑ In April, 21 classrooms (52%) scored a 3 or above on the READ implementation rubric.
Improved integration of READ into classroom instruction (intermediate)	In what ways and to what extent did teachers integrate READ into their classroom instruction? ** Note the quality with which READ was integrated into classroom instruction by teachers.	Improved integration of READ lessons into classroom instruction	By March, 50% of teachers will score a 3 or above on the READ implementation rubric. By May, 75% of teachers will score a 3 or above and 25% of teachers will score a 4 on the READ implementation rubric.	☑ By March, 21 teachers (52%) scored a 3 or above on the READ implementation rubric. → Twelve teachers (30%) scored a 2 on the READ implementation rubric, while seven (18%) scored a 1 on the rubric. ☒ By May, 27 teachers (68%) scored a 3 or above on the READ implementation rubric. → Nine teachers (22%) scored a 2 on the rubric, while four (10%) scored a 1 on the rubric. ☑ By May, 15 teachers (38%) scored a 4 on the READ implementation rubric.

Logic Model Components	Evaluation Questions	Indicators	Targets	Evaluation Findings
Improved differentiation of instruction (intermediate)	In what ways and to what extent did teachers use READ assessment data to plan and differentiate instruction? ** Note what data were used and how data were used in instructional planning.	Increased number of teachers using READ assessment data to plan instruction Improved use of READ assessment data to differentiate instruction	By December, all teachers will be using READ assessment data on a weekly basis to plan and differentiate instruction. By March, 50% of teachers will score a 3 or above on the READ implementation rubric. By May, 75% of teachers will score a 3 or above and 25% of teachers will score a 4 on the READ implementation rubric.	☒ On the December teacher survey, less than half the teachers (48%) reported using READ assessment data on a weekly basis to plan and differentiate instruction. → Also on the December survey, most teachers (88%) said they had used the READ assessment data at least some to plan instruction, while a little over 10% said they had never used the READ assessment data in their classroom lesson planning. ☑ By March, 21 teachers (52%) scored a 3 or above on the READ implementation rubric. → Twelve teachers (30%) scored a 2 on the READ implementation rubric, while seven (18%) scored a 1 on the rubric. ☒ By May, 27 teachers (68%) scored a 3 or above on the READ implementation rubric. → Nine teachers (22%) scored a 2 on the rubric, while four (10%) scored a 1 on the rubric. ☑ By May, 15 teachers (38%) scored a 4 on the READ implementation rubric.

Table 15: READ Evaluation Results—Long-Term Goals

Logic Model Components	Evaluation Questions	Indicators	Targets	Evaluation Findings
Increased student engagement in reading	To what extent and in what ways did READ foster student engagement during reading lessons?	Increased frequency and improved quality of student engagement in the classroom, as measured by the READ implementation rubric	By February, 25% of classrooms will have a score of 3 or above (out of 4) on the READ implementation rubric. By April, 50% of classrooms will score a 3 or above on the READ implementation rubric. By June, 75% of classrooms will score a 3 or above and 25% of teachers will score a 4 on the READ implementation rubric.	☑ By February, 28% of classrooms scored a 3 or above on the READ implementation rubric. ☑ By April, 52% of classrooms scored a 3 or above on the rubric. ☒ By June, 68% of classrooms scored a 3 or above on the rubric. ☑ By June, 38% of classrooms scored a 4 on the rubric.
Improved student reading skills	To what extent did READ improve student learning in reading? To what extent did student learning improve after READ was implemented?	Increased scores on tests assessing students' reading ability (state and READ assessments)	Within 2 years, the increase in student scores on the state standards-based reading assessment will be statistically significant for those students who participated in READ versus those students who did not participate in READ.	☑ First year results indicate that state reading scores for READ students were higher than those for non-READ students, and the difference was statistically significant.

Logic Model Components	Evaluation Questions	Indicators	Targets	Evaluation Findings
	To what extent did READ improve student learning in reading? How did student performance on the READ assessments correlate with student performance on state assessments?		Reading scores on the state assessment will be analyzed in relation to scores on the READ assessment data, in order to determine the degree to which READ assessments correlate with the state assessments.	☑ Student scores on READ assessments had a statistically significant and strong positive correlation with student scores on the state reading assessment, indicating that the assessments are likely well aligned.

Logic Model Components	Evaluation Questions	Indicators	Targets	Evaluation Findings
	To what extent did READ improve student learning in reading? To what extent did learning outcomes vary with teacher use of READ in the classroom? To what extent did learning outcomes vary with teacher use of READ assessment data to plan and differentiate instruction? In what ways did learning outcomes vary by initial reading performance on the state reading test, grade level, special education status, and English language proficiency?		Reading assessment data will be disaggregated and examined by quality of teacher use (using the READ implementation rubric), initial reading performance, grade level, gender, ethnicity, special education status, and English language proficiency.	☑ Students in classrooms where READ was used regularly and with fidelity: → Demonstrated a statistically significant increase in their reading scores from last year to this year. → Increased their reading scores by twice as much as students in READ classrooms where READ was used minimally. Among READ students: → Third-grade students showed greater reading gains on the state assessment than did fourth- and fifth-grade students. → There were no differences in reading gains by gender or ethnicity. → Regular education students showed greater reading gains than did special education students. → Scores for English Language Learner students were mixed. In third grade, ELL students showed statistically significant reading gains, while the differences in reading scores on the state assessment were not statistically significant for ELL students in either the fourth or fifth grade.

Logic Model Components	Evaluation Questions	Indicators	Targets	Evaluation Findings
	To what extent did READ improve student learning in reading? In what ways did learning outcomes vary with the frequency of READ use at home?		Reading assessment data will be disaggregated and examined by frequency of home use.	☒ Assessment data could not be analyzed by home use because only one classroom implemented the home component.

Although the READ evaluation was a true experimental design, E-Team members knew it would still be worthwhile to consider the possibility that other factors might have influenced the positive findings. The E-Team therefore brainstormed possible competing explanations for the positive results of the READ program.

The E-Team decided that another plausible explanation for the positive results was that the teachers who used READ regularly in the classroom and who used READ assessments as intended may have been more skilled teachers and their students might have had a similar increase in reading scores even without the READ program. The E-Team decided to follow up on fidelity of implementation and its relationship to teacher skills. In addition, while classrooms were randomly assigned to READ to minimize initial differences between READ and non-READ classrooms, it is possible that by chance more skilled teachers were assigned to the READ program group. The E-Team also intends to investigate this issue further in Year 2 of the evaluation.

Communicating Results

The READ oversight team met monthly to discuss program monitoring and improvement. At each meeting, the READ evaluator, Dr. Elm, and the E-Team provided an update to the oversight team. Based on the formative evaluation findings, the oversight team developed recommendations and a plan for the next month.

At the December school board meeting, the oversight team presented a status report, noting important findings from the evaluation. The oversight team asked Dr. Elm to create a full evaluation report for the administration and to present the findings at the August school board meeting. The E-Team also drafted a one-page brief of evaluation findings which was provided to all participants, as well as to the local newspaper.

Step 5: Inform and Refine – Using the Results

Informing the Program

During her evaluation update at the November oversight team meeting, Dr. Elm shared initial findings from the **evaluation of the implementation** of READ program activities. Indicators showed that many students did not have the technology available at home to access READ. Even within those schools that had high numbers of students with the technology necessary for home access, the classroom variability was large. Only one of the 40 classrooms was able to have 100 percent of students access READ from home. Open-ended survey items revealed that teachers did not feel comfortable offering READ homework assignments to some but not all students in their classroom and therefore chose not to train students in the home use of READ. Only one teacher had trained his students in the home use of READ because all of his students

had the technology at home necessary to access READ. This teacher indicated that he would like to continue with the home component of READ.

The oversight team discussed the home-component issue and asked for advice from the E-Team on how to proceed. With the support of the E-Team, the oversight team decided to have a one classroom pilot of the home component but otherwise to remove the home component from the program during Year 1. Based on results from the pilot, implementing a partial home component in Year 2 would be considered.

During the same November update, Dr. Elm provided some findings from the **evaluation of the early/short-term objectives** on the READ logic model. She noted that in October all teachers had reported using READ in their classroom and that over half of teachers reported that they had used READ every week. However, over one-quarter of teachers reported that they had used READ in their classroom only once or twice in the last month. Survey data indicated that some of these teachers felt overwhelmed with the technology and some said they could not fit READ classroom use into their already busy day.

The oversight team discussed this information and decided to make a midcourse adjustment. Before the READ program began, team members had thought that the initial professional development and ongoing technical assistance would be sufficient. However, they now believed that they needed to make one-to-one professional development available to those teachers who would like to have someone come into their classroom and model a lesson using READ. Mrs. Anderson assigned arrangements for this one-on-one professional development to one of the oversight team members.

During her evaluation update at the January oversight team meeting, Dr. Elm shared findings from the **evaluation of the intermediate objectives** on the READ logic model. Dr. Elm explained that on the December teacher survey, slightly less than half the teachers reported that they used the READ assessment data on a weekly basis for planning and differentiating instruction. One in 10 teachers said they had never used the READ assessment data. Dr. Elm further stated that the lack of use of the READ assessment data was likely affecting scores on the READ implementation rubric. From classroom observations, interviews, and surveys, she believed that the quality of teacher use of READ in the classroom was progressing nicely but that the lack of assessment data use was decreasing the overall rubric score.

The oversight team knew that using the READ assessment data to plan and differentiate instruction was critical to the program's success. Mrs. Anderson decided to discuss the issue with the READ faculty at each school in an effort to understand what she could do to facilitate their use of the READ assessment data. Additionally, the E-Team planned to elaborate on the rubric so that subscores could be captured for various components of the rubric. These rubric subscores would be especially useful for analysis when the data are disaggregated by teacher

use of READ in the classroom, student interaction in the classroom, and teacher use of READ student assessment data to plan and differentiate instruction. The revised rubric would be developed during the spring, piloted over the summer, and implemented during Year 2.

Finally, at the evaluation update at the end of the school year, Dr. Elm reported on the **preliminary evaluation of long-term goals** of the READ program. Student reading achievement was higher among students of teachers who used READ regularly and as intended, and the difference was statistically significant. Further, students of teachers who used the READ assessment data to tailor classroom instruction had higher reading test scores than students of teachers who did not use the READ assessment data, and again the difference was statistically significant.

Year 1 evaluation findings also indicated that not all teachers had bought into using READ with their students, especially the READ assessment component. The oversight team decided to share the evaluation findings with all teachers at a staff meeting in order to encourage them to use READ in their classroom. Prior to sharing the evaluation findings with teachers, Dr. Elm conducted an anonymous follow-up survey at the staff meeting in an effort to find out why some teachers chose not use READ.

Refining the Program Logic

The READ oversight team felt that the logic model they created accurately portrayed the program. Yet, since it was clear from November that the home component could not be fully implemented, they wanted to highlight this on the logic model. The team decided to draw a box around the program as it was implemented, excluding the home component. Below the model, a note was provided indicating why the home component was not part of the existing implementation and that it was currently being piloted in one classroom. The oversight team hoped to understand more about the implementation of the home component, as well as the success of the home component, from examining results from the pilot classroom.

The oversight team also wanted to understand more about the strength of the relationship between classroom use of READ and state assessment scores and between use of READ assessment data for instructional planning and state assessment scores. It noted this on the logic model and asked the E-Team to investigate the linkages further in the second year of the evaluation.

Making Recommendations

The READ oversight team recommended that the READ program be offered to all students in the district. It also recommended that the program be incorporated into the regular curriculum. The team felt that the positive findings regarding test scores were strong enough that all students should have access to it.

However, since READ funding was still at the 50 percent level for the second year, the oversight team planned to work with Dr. Elm and the E-Team for another year in order to continue to refine the implementation of the program in the classroom and to further understand the success of the READ program with students. To do this, the team recommended that the second-year evaluation include student surveys and focus groups as data sources to address objectives related to student interaction and engagement in the classroom.

The oversight team decided to continue to advocate for the program's expansion in the hope that it would be institutionalized soon.

Appendix B: Embedded Evaluation Illustration – NowPLAN*

Program Snapshot

Strategic Planning for Learning and Achievement in Nowgarden (NowPLAN) is part of a statewide, strategic planning initiative. The example in Appendix B focuses on the building-level evaluation of a state's strategic technology plan (NowPLAN-T). The NowPLAN-T evaluation uses theory-based, embedded evaluation within a mixed-method design.

*The example set out in this appendix is provided solely for the purpose of illustrating how the principles in this guide can be applied in actual situations. The programs, characters, schools, and school district mentioned in the example are fictitious.

Step 1: Define the Program

Background

Every six years, Nowgarden School District's State Department of Education requires each district to create a new strategic plan. This strategic plan is intended to drive the district's initiatives and strategies over the next six years, as well as to provide a means with which to monitor and evaluate the implementation of these initiatives and strategies.

Nowgarden School District recently completed its strategic planning process. With the help of teachers, administrators, board members, and the community, Nowgarden developed a 6-year plan that includes action plans for student learning, professional development, additional learning opportunities, safety and security, technology, and communication. The district's technology plan, in particular, was modeled after the state technology plan and is intended as the foundation to build a districtwide technology program that will support the 21st century learner.

The Evaluation

The Nowgarden School District administration asked a local evaluation organization to help them plan and conduct an evaluation of their strategic plan. The organization assigned an evaluator to work with the district to create an evaluation design. The external evaluator and the district created an evaluation of the district's 6-year strategic plan that includes both quantitative and qualitative measures and has two foci:

1. Formative evaluation to help shape and improve the implementation of the plan strategies.

2. Summative evaluation of the plan to determine its overall success with students.

The main focus of this multiyear evaluation is to monitor how well the district has achieved its primary learning-related long-term goal: to improve the achievement of all students.

As part of its strategic planning, Nowgarden identified the six key strategies below that it will employ to meet the primary long-term goal. The strategic plan evaluation also will examine and monitor the extent to which and how well the district's primary learning goal is being met as the district implements these strategies.

While Nowgarden School District and its external evaluator will design a comprehensive evaluation of all six strategies, they chose to first design the evaluation of the district's technology component of the strategic plan.

Program Goals

Nowgarden School District's strategic plan and its related technology plan have one primary long-term goal: to improve student achievement. However, a longer-term goal of the district is to ultimately improve postsecondary success for all students.

Program Strategies and Activities – NowPLAN

Nowgarden identified six key strategies to meet the district's long-term goals:

1. Provide an engaging and challenging education program.
2. Provide the necessary resources and professional development for teachers.
3. Provide additional opportunities and supports for students.
4. Provide a safe and healthy educational environment.
5. Provide the technological tools necessary for the 21st century learner.
6. Effectively communicate with and engage the community.

Strategy #5, provide the technological tools necessary for the 21st century learner, will be the focus of this example.

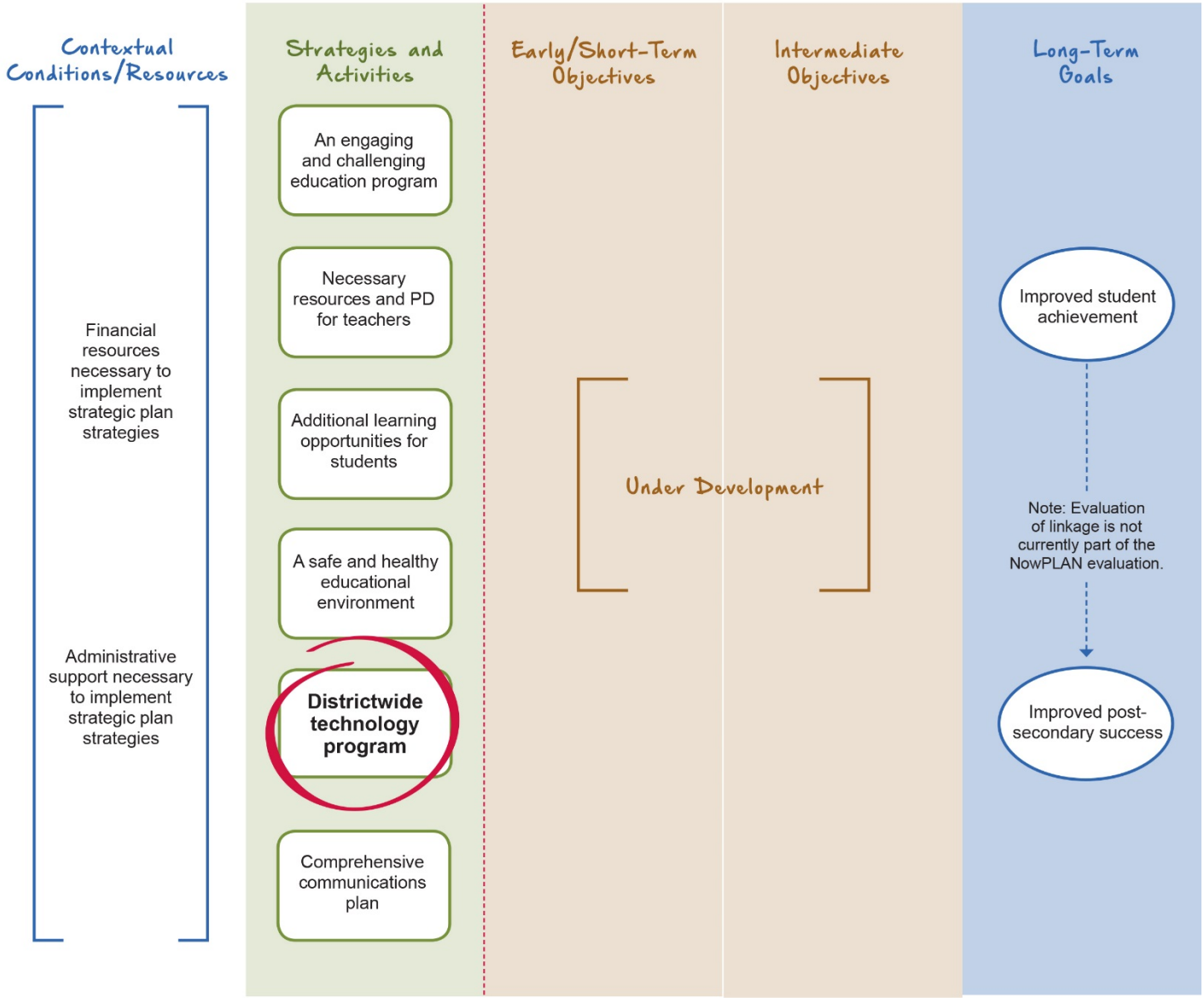
A logic model shell of the entire Nowgarden strategic plan, titled NowPLAN Logic Model, is provided in **Figure 4: NowPLAN Logic Model**.

Note that the "improved postsecondary success" goal is not currently addressed through the NowPLAN evaluation. However, the district does plan to examine progress toward this goal in the future through graduate follow-up surveys and focus groups.

Also note that the strategy related to the districtwide technology plan is highlighted. The relationship between the district's technology plan and its long-term goals will be explored throughout the remainder of this appendix.

When referring to the technology component of the overall strategic plan, the acronym NowPLAN-T will be used.

Figure 4: NowPLAN Logic Model



Program Strategies and Activities – NowPLAN-T

The NowPLAN-T evaluation examines Nowgarden's strategy to create a technology plan that provides the technological tools necessary for the 21st century learner. The district identified seven activities that comprise the technology plan strategy.

1. **Districtwide Technology Curriculum:** to create a districtwide technology curriculum K-12. The technology curriculum should integrate technology into the core curriculum.
2. **Student Technology Orientation:** to create a technology orientation plan for students (initially for all students, and then for new students transferring into the district).
3. **Technology Professional Development Model:** to create a districtwide professional development model. The model will focus on improving student learning, best instructional practice, and administrative functions such as data management and assessment.
4. **Teacher Technology Orientation:** to create a technology orientation plan for teachers (initially for all teachers, and then for new teachers transitioning into the district).
5. **Technology-based Communications:** to fully develop the use of technology resources for communication throughout the school community.
6. **Technology-based Additional Learning Opportunities:** to provide technology-based additional learning opportunities (ALOs) for students, including but not limited to distance learning and extended school day opportunities.
7. **Hardware and Software Acquisition Plan:** to create a districtwide software and hardware acquisition plan.

Relating Strategies to Goals: Program Theory

The overall theory behind the NowPLAN-T program is that by providing

- A districtwide technology curriculum,
- A technology orientation plan for students,
- A technology professional development model,
- A technology orientation plan for teachers,
- Technology resources for communication,
- Technology-based additional learning opportunities, and
- A districtwide software and hardware acquisition plan,

↳ Student achievement will be improved, AND

↳ Student postsecondary success ultimately will be improved.

The external evaluator helped the district people develop a set of assumptions as to why they believed NowPLAN-T strategies would result in improved student achievement. They based these assumptions on research and developed them in consultation with educational technology professionals.

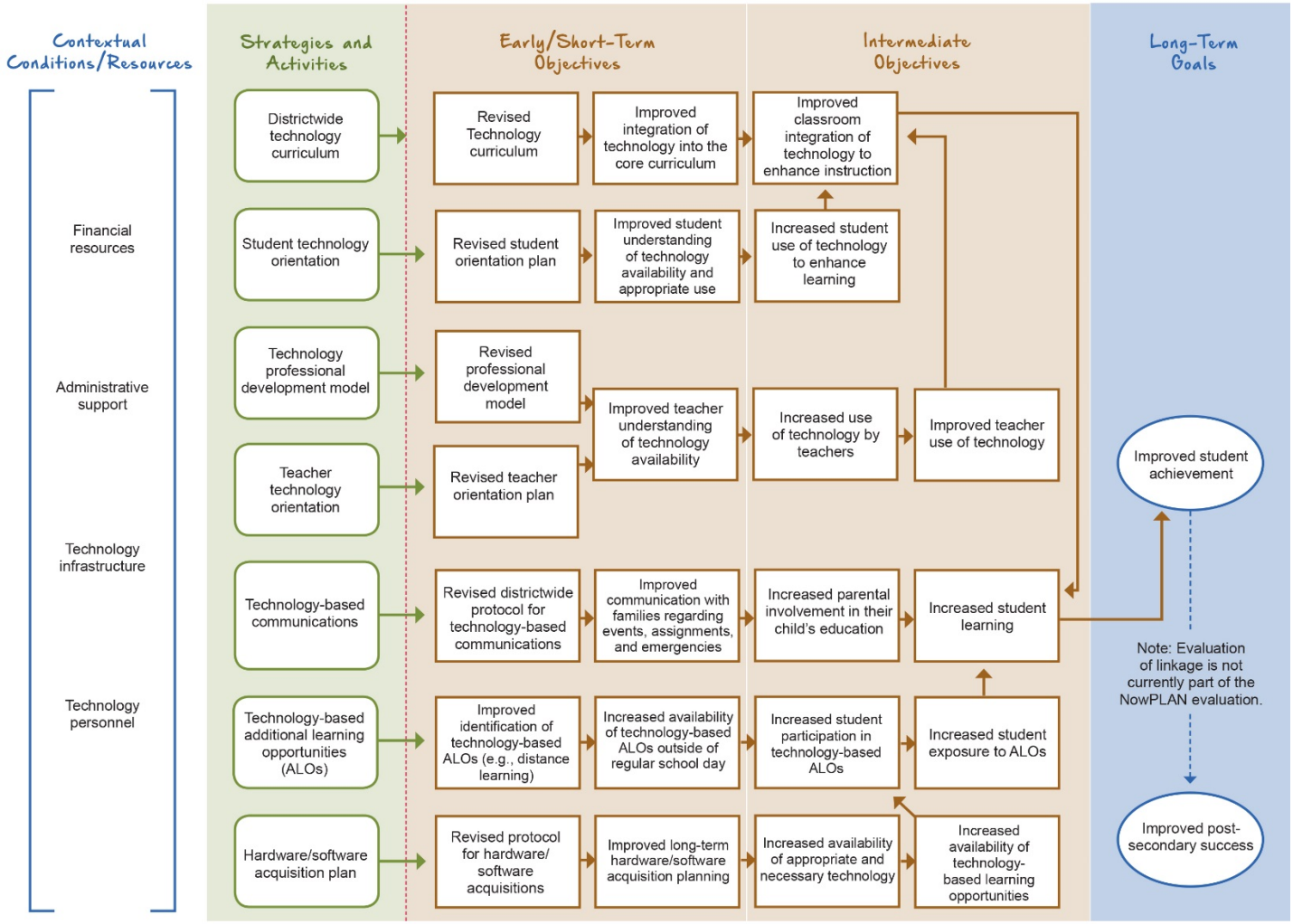
1. A **districtwide technology curriculum** will lead to a revised technology curriculum, which will lead to improved integration of technology into the core curriculum. This integration will lead to improved integration of technology in the classroom to enhance instruction.
2. A **technology orientation plan for students** will lead to a revised student orientation plan, which will lead to improved student understanding of technology availability and appropriate use. Improved understanding will lead to increased student use of technology to enhance learning and then improved integration of technology to enhance instruction.
3. A **technology professional development model** will lead to a revised professional development model, which will lead to improved teacher understanding of technology availability and then an increased use of technology by teachers. Increased use of technology will lead to improved teacher use of technology, which will in turn improve the integration of technology to enhance instruction.
4. A **technology orientation plan for teachers** will lead to a revised teacher orientation plan, which will lead to improved teacher understanding of technology availability and then to increased use of technology by teachers. Increased use of technology will lead to improved teacher use of technology. Improved teacher use will in turn improve the integration of technology to enhance instruction.
5. **Technology-based communications** will lead to a revised districtwide protocol for technology-based communications, which will lead to improved communication with families regarding events, assignments, and emergencies. Improved communication will lead to increased parental involvement in their child's education.
6. **Technology-based additional learning opportunities** (ALOs) will improve the identification of technology-based ALOs. Improved identification will lead to increased availability of ALOs outside of the regular school day, which will lead to increased student participation in technology-based learning opportunities. Increased participation will lead to increased student exposure to learning opportunities.
7. A **districtwide software and hardware acquisition plan** will lead to a revised protocol for hardware and software acquisitions, which will lead to improved long-term acquisition planning. Improved acquisition planning will lead to increased availability of appropriate and necessary technology, then to increased availability of technology-based learning opportunities, and then to increased student participation, which will lead to increased student exposure to learning opportunities.

8. Improved use and integration of technology to enhance instruction, increased parental involvement in their child's education, and increased student exposure to learning opportunities will lead to increased student learning and **improved student achievement** and ultimately **improved postsecondary success**.

Program Logic Model

Figure 5 illustrates these assumptions in the NowPLAN-T logic model.

Figure 5: NowPLAN-T Logic Model



Resources

Nowgarden School District worked with its external evaluator to identify several contextual conditions and resources that are necessary for the success of the district's technology plan. These are listed in the first column of the NowPLAN-T logic model and include financial resources to implement the technology plan, administrative support throughout the NowPLAN-T implementation, technology infrastructure (or resources to build this infrastructure), and the necessary technology personnel.

Step 2: Plan the Evaluation

Evaluation Design

Nowgarden School District is a public school district educating over 56,000 students. The district has 38 elementary schools (grades K-5), 13 middle schools (grades 6-8), and 11 high schools (grades 9-12). Nowgarden employs over 4,000 teachers across its 62 schools.

The district's technology plan is implemented districtwide. The NowPLAN-T program and its evaluation will employ a single group evaluation design and will include both quantitative and qualitative data sources.

Design: Mixed-method, nonexperimental design

To improve the quality of the NowPLAN-T evaluation, Nowgarden's external evaluator and the district worked to embed the evaluation into the technology plan. The evaluation is theory based and uses logic modeling to relate NowPLAN-T strategies and activities to long-term district goals. The evaluation design is longitudinal (6 years) and uses repeated measures. (Some data will be collected annually, while other data will be collected quarterly.)

Enriching the Evaluation Design: Logic modeling; longitudinal data with some repeated measures. In-depth case study of NowPLAN-T schools and sampling of classrooms within schools will be used to collect observational data.

Data Collection Methods

Data will be collected through a variety of methods, including interviews, documents, surveys, rubrics, and assessments. All schools in the district will participate in certain components of the evaluation, while a purposefully selected group of schools will be chosen for a comprehensive case study, based on the quality of their implementation of the NowPLAN-T activities.

In the districtwide portion of the evaluation, all teachers and students will participate in surveys examining the use and integration of technology into the classroom. Teacher and student

surveys will be based on the NowPLAN-T rubrics. That is, along with questions regarding the program's implementation, teachers and students will self-report on rubric components.

To provide a rich understanding of the classroom context of NowPLAN-T, the external evaluator will conduct a case study in a purposefully selected group of schools. Schools will be selected for the case study based upon the quality with which they have implemented NowPLAN-T activities. The case study will focus on four schools and provide linkages between classroom-level implementation and student learning that may not be fully understood by examining survey data only. A random sample of classrooms will be chosen at each case study school to participate in observations using the NowPLAN-T rubrics. The external evaluator and trained members of an evaluation team assembled by the external evaluator will observe these classrooms periodically throughout the school year. During observations, the NowPLAN-T rubrics will be completed and classroom context recorded. Case study teachers will also participate in in-depth interviews.

Student surveys (delivered electronically) will address classroom and home use of technology for primary learning activities, as well as additional learning opportunities offered as part of NowPLAN. The evaluation also will use participation logs from additional learning opportunities to determine opportunities that were offered and attendance. Additionally, a parent survey will be electronically administered to all parents in the district. The survey will be voluntary and anonymous. Information from the parent survey will examine communication and parent involvement related to NowPLAN components. The district would like but does not currently have the funds to conduct a graduate follow-up survey to address the second long-term goal of improved postsecondary success. The proposed survey is included in the evaluation design as a placeholder for future evaluation possibilities.

Step 3: Implement the Evaluation

The NowPLAN-T evaluation matrix is provided in tables 16, 17, and 18. Evaluation questions, indicators, and targets, as well as data sources, collection, and analysis are addressed by logic model component.

Note that some very early (short-term) objectives are included in **Table 16: NowPLAN-T Evaluation Matrix—Strategies and Activities/Initial Implementation**. The remaining short-term and intermediate objectives are covered in **Table 17: NowPLAN-T Evaluation Matrix—Early/Short-Term and Intermediate Objectives**. **Table 18: NowPLAN-T Evaluation Matrix—Long-Term Goals** addresses long-term goals on improving student achievement and postsecondary success.

Table 16: NowPLAN-T Evaluation Matrix—Strategies and Activities/Initial Implementation

Logic Model Components	Evaluation Questions*	Indicators	Targets	Data Sources	Data Collection	Data Analysis
Districtwide technology curriculum/revised technology curriculum Student technology orientation/revised student orientation plan Technology professional development model/revised professional development model Teacher technology orientation/revised teacher orientation plan Technology-based communications/revised districtwide protocol for technology-based communication Hardware/software acquisition plan/revised protocol for hardware/ software acquisitions	In what ways were the districtwide technology curriculum, student technology orientation plan, technology professional development model, teacher technology orientation plan, districtwide protocol for technology-based communication, and protocol for hardware/soft-ware acquisitions revised?	Creation of a revised districtwide technology curriculum, student technology orientation plan, technology professional development model, teacher technology orientation plan, districtwide protocol for technology-based communication, and protocol for hardware/software acquisitions	By the end of Year 1, a revised districtwide hardware/software acquisition plan and districtwide protocol for technology-based communication will have been created. By the end of Year 2, a revised districtwide technology curriculum and technology professional development model will have been created. By the end of Year 3, revised student and teacher technology orientation plans will have been developed.	Technology records Document analysis Meeting minutes Interviews with technology personnel	Technology records and documents, as well as meeting minutes, reviewed monthly Interviews with technology personnel conducted quarterly	Documents/minutes summarized for evidence of implementation Interview data summarized, and if warranted, analyzed for themes

Logic Model Components	Evaluation Questions*	Indicators	Targets	Data Sources	Data Collection	Data Analysis
Technology-based additional learning opportunities (ALOs)/improved identification of ALOs	To what extent were technology-based ALOs identified?	Increasing number of technology-based ALOs identified	By the end of Year 2, a process to identify technology-based ALOs will be operational.	Interviews with technology personnel	Interviews with technology personnel conducted quarterly	Interview data summarized, and if warranted, analyzed for themes

*Note: Logic model components are combined in the evaluation questions but will be disaggregated in the data analysis.

Table 17: NowPLAN-T Evaluation Matrix—Early/Short-Term and Intermediate Objectives

Logic Model Components	Evaluation Questions	Indicators	Targets**	Data Sources	Data Collection	Data Analysis
Improved integration of technology into the core curriculum	How was technology integrated into the core curriculum?	Increased number of schools with improved classroom integration of technology	By the end of Year 4, at least <>% of schools will score a 3 or better on the NowPLAN-T rubrics. By the end of Year 6, <>% will score a 4 out of 4 on the NowPLAN-T rubrics.	NowPLAN-T rubrics Teacher surveys	Baseline rubric data collected at start of Year 1 Rubric data collected quarterly (for each school), through teacher surveys (all classrooms) and classroom observations (case study classrooms)	Rubric data analyzed by frequency distributions of rubric scores Changes over time analyzed using significance testing

Logic Model Components	Evaluation Questions	Indicators	Targets**	Data Sources	Data Collection	Data Analysis
Improved classroom integration of technology to enhance instruction	In what ways and to what extent was technology integrated into classroom instruction?	Increased number of schools with improved curricular integration of technology	By the end of Year 4, at least <>% of schools will score a 3 or better on the NowPLAN-T rubrics. By the end of Year 6, <>% will score a 4 out of 4 on the NowPLAN-T rubrics.	NowPLAN-T rubrics Teacher surveys	Baseline rubric data collected at start of Year 1 Rubric data collected quarterly (for each school), through teacher surveys (all classrooms) and classroom observations (case study classrooms)	Rubric data analyzed by frequency distributions of rubric scores Changes over time analyzed using significance testing
Improved student understanding of technology availability and appropriate use	To what extent do students understand the technology available to them, as well as its appropriate use?	Increased number of students who have an understanding of available technology	Using Year 1 survey data as a baseline --by the end of Year 4 at least <>%, Year 5 at least <>%, and Year 6 at least <>% of students will appropriately understand and use available technology to enhance learning.	Student surveys	Baseline student survey administered during Year 1 Student survey administered annually (and electronically) to all students	Survey data analyzed using frequency distributions and basic descriptive statistics Changes over time analyzed using significance testing

Logic Model Components	Evaluation Questions	Indicators	Targets**	Data Sources	Data Collection	Data Analysis
Increased student use of technology to enhance learning	In what ways and how often do students use technology for learning (in the classroom and at home)?	Increased number of students who use technology in their learning activities	Using Year 1 survey data as a baseline --by the end of Year 4 at least <>%, Year 5 at least <>%, and Year 6 at least <>% of students will appropriately understand and use available technology to enhance learning.	Student surveys	Baseline student survey administered during Year 1 Student survey administered annually (and electronically) to all students	Survey data analyzed using frequency distributions and basic descriptive statistics Changes over time analyzed using significance testing
Improved teacher understanding of technology availability	To what extent do teachers understand the technology available to them?	Increased number of schools with improved teacher understanding of the technology available to them	By the end of Year 4, at least <>% of schools will score a 3 or better on the NowPLAN-T rubrics, and will demonstrate an understanding of technology (as measured by the teacher survey). By the end of Year 6, <>% will score a 4 out of 4 on the NowPLAN-T rubrics, and will demonstrate an understanding of technology (as measured by the teacher survey).	Teacher surveys NowPLAN-T rubrics	Baseline teacher survey data and rubric data collected at start of Year 1 Rubric data collected quarterly (for each school), through teacher surveys (all classrooms) and classroom observations (case study classrooms)	Survey data analyzed through basic descriptive statistics and frequency distributions Rubric data analyzed by frequency distributions of rubric scores Changes over time analyzed using significance testing

Logic Model Components	Evaluation Questions	Indicators	Targets**	Data Sources	Data Collection	Data Analysis
Increased use of technology by teachers	To what extent do teachers use technology to improve student learning?	Increased number of schools with increased teacher use of technology to improve student learning				
Improved teacher use of technology	In what ways do teachers use technology to improve student learning?	Increased number of schools with improved teacher use of technology to improve student learning				
Improved communication with families	To what extent has communication with families improved?	Increased number of families who report improved communication	Using Year 1 survey data as a baseline --by the end of Year 4 at least <>%, Year 5 at least <>%, and Year 6 at least <>% of parents will report improved communication.	Parent survey	Baseline parent survey administered during Year 1 Parent survey administered annually (and electronically) to all families	Survey data analyzed using frequency distributions and basic descriptive statistics Changes over time analyzed using significance testing

Logic Model Components	Evaluation Questions	Indicators	Targets**	Data Sources	Data Collection	Data Analysis
Increased parental involvement	In what ways has technology contributed to parental involvement?	Increased number of parents for whom communication has contributed to increased parental involvement	Using Year 1 survey data as a baseline --by the end of Year 4 at least <>%, Year 5 at least <>%, and Year 6 at least <>% of parents will report that communication has contributed to their increased parent involvement.			
Increased availability of technology-based additional learning opportunities (ALOs)	To what extent are technology-based ALOs available to students?	Increased number of technology-based ALOs offered to students	Each year, the number of technology-based ALOs offered to students will increase by 20% (e.g., online courses, supplemental programs).	Technology records	Technology records and participation logs reviewed quarterly	Technology records and participation logs reviewed for evidence of ALOs availability

Logic Model Components	Evaluation Questions	Indicators	Targets**	Data Sources	Data Collection	Data Analysis
Increased student participation in technology-based ALOs	How often do students participate in technology-based ALOs?	Increased number and percent of students participating in technology-based ALOs (within and outside of the regular school day)	Using Year 1 survey data as a baseline --by the end of Year 2 at least <>%, Year 3 at least <>%, Year 4 at least <>%, Year 5 at least <>%, and Year 6 at least <>% of students will participate in technology-based ALOs.	Participation logs Student surveys	Baseline student survey administered during Year 1 Student survey administered annually (and electronically) to all students	Survey data analyzed using frequency distributions and basic descriptive statistics Changes over time analyzed using significance testing Results disaggregated by type of ALO (e.g., home-based, outside of school day)
Increased student exposure to technology-based ALOs	To what extent and in what ways do students participate in technology-based ALOs?	Increased number and percentage of students who have increased their overall learning time through technology-based ALOs (Note: investigate the nature of use, i.e., replacing an ALO or adding a new ALO)	Using Year 1 survey data as a baseline – students will have increased their learning time through technology-based ALOs by the end of Year 2 at least <>%, Year 3 at least <>%, Year 4 at least <>%, Year 5 at least <>%, and Year 6 at least <>%.	Participation logs Student surveys	Baseline student survey administered during Year 1 Student survey administered annually (and electronically) to all students	Survey data analyzed using frequency distributions and basic descriptive statistics Changes over time analyzed using significance testing Results disaggregated by type of ALO (e.g., home-based, outside of school day) and nature of the ALO (e.g., if the student replaced an ALO with a technology-based ALO)

Logic Model Components	Evaluation Questions	Indicators	Targets**	Data Sources	Data Collection	Data Analysis
Improved long-term hardware/ software acquisition planning	In what ways has hardware/software acquisition planning improved?	Increased number of teachers and technology staff who report improved acquisition planning	By the end of Year 2 at least <>%, Year 3 at least <>%, Year 4 at least <>%, Year 5 at least <>%, and Year 6 at least <>% of teachers/staff will report improved planning.	Interviews with technology staff Teacher surveys	Interviews conducted annually Teacher surveys administered annually	Interviews summarized and analyzed by theme Survey data analyzed using frequency distributions and basic descriptive statistics
Increased availability of appropriate and necessary technology	To what extent has the availability of appropriate and necessary technology improved?	Increased number of teachers and technology staff who report improved availability of necessary technology	By the end of Year 2 at least <>%, Year 3 at least <>%, Year 4 at least <>%, Year 5 at least <>%, and Year 6 at least <>% of teachers/staff will report improved availability.	Interviews with technology staff Teacher surveys	Interviews conducted annually Teacher surveys administered annually	Interviews summarized and analyzed by theme Survey data analyzed using frequency distributions and basic descriptive statistics

Increased student learning	**To what extent has technology contributed to student learning as measured by local assessments?** **To what extent did learning outcomes vary by school and classroom technology use?** **(Not currently evaluated: In what ways have technology-based ALOs contributed to student learning?)**	**Increased scores on local assessments** **Increased correlation between local assessment scores and technology implementation**	**Students in schools with high rubric scores will have higher gains on local assessments than students in schools with lower rubric scores, and the difference will be statistically significant.**	**Local assessments** **NowPLAN-T rubric scores**	**Local assessment data collected quarterly** **Baseline rubric data collected at start of Year 1** **Rubric data collected quarterly (for each school) through teacher surveys (all classrooms) and classroom observations (case study classrooms)**	**Correlational analyses between local assessment scores and NowPLAN-T rubric scores** **T-test of mean test scores pre-NowPLAN-T and each academic year post-NowPLAN-T** **Results disaggregated by school, grade level, gender, race/ethnicity, special education status, and English language proficiency**

**Targets will be updated once baselines are measured.

Table 18: NowPLAN-T Evaluation Matrix—Long-Term Goals

Logic Model Component	Evaluation Questions	Indicators	Targets	Data Source	Data Collection	Data Analysis
Improved student achievement	To what extent did the district's technology plan contribute to student achievement? → To what extent did student learning improve after NowPLAN-T was implemented? → To what extent did learning outcomes vary by school and classroom technology use?	Increased scores on statewide standards-based achievement assessments Increased correlation between achievement scores and NowPLAN-T implementation	Within 2 years, the correlation between improvement in student scores on the statewide standards-based achievement tests and scores on the NowPLAN-T technology rubrics will be statistically significant. Students in schools (and classrooms) with high rubric scores will have higher achievement gains than students in schools (and classrooms) with lower rubric scores, and the difference will be statistically significant.	State test scores in reading and math (as well as science and writing) NowPLAN-T rubric scores	State tests conducted in April of each academic year Rubric data collected quarterly, through teacher surveys (all classrooms) and classroom observations (case study classrooms)	Correlational analyses between state achievement test scores and NowPLAN-T rubric scores T-test of mean test scores pre-NowPLAN-T and each academic year post-NowPLAN-T Results disaggregated by school, classroom, grade level, gender, race/ethnicity, special education status, and English language proficiency
Improved postsecondary success	To what extent was postsecondary success related to implementation of NowPLAN-T?	Increased correlation between achievement scores and postsecondary success	Note: This component is not currently evaluated. Indicators will be refined and targets will be determined at a later date.	Graduate follow-up surveys and focus groups	Note: This component is not currently evaluated. Data collection will be determined at a later date.	Descriptive statistics Qualitative analysis of focus group data

Logic Model Component	Evaluation Questions	Indicators	Targets	Data Source	Data Collection	Data Analysis
	To what extent was postsecondary success related to student achievement?	Increased correlation between NowPLAN-T implementation and postsecondary success	Note: This component is not currently evaluated. Indicators will be refined and targets will be determined at a later date.	State test scores Graduate follow-up surveys		Descriptive statistics Correlational studies and significance testing

Step 4: Interpret the Results

The NowPLAN-T external evaluator will meet with district staff quarterly to provide interim findings for program improvement and midcourse adjustments. The external evaluator will also provide an annual evaluation report focusing on the district's progress towards meeting the targets set for NowPLAN-T objectives. Biannual newsletters (traditional and electronic), as well as periodic presentations and press releases, will be used to communicate NowPLAN-T progress and findings to staff, parents, and students.

Step 5: Inform and Refine – USING the Results

Evaluation results will be used to inform and improve NowPLAN-T, refine the NowPLAN-T logic model (as necessary), and make recommendations and decisions regarding the future direction of NowPLAN-T.

Additional Notes

Embedding Evaluation in the Strategic Plan

Nowgarden School District had been through strategic planning before. Past strategic plans were completed only because they were required and then filed away and rarely consulted. The superintendent, who was hired 3 years ago, saw strategic planning as an opportunity for the district to reflect and grow. The superintendent did not want to ask teachers, administrators, and community members to spend their valuable time participating in a strategic planning process that was not going to be used to its fullest potential. The superintendent knew that a good strategic plan could be used for positive change and growth, and that embedding evaluation within the plan itself would provide information for continuous improvement. District administration agreed that a powerful strategic plan with an embedded evaluation provides the ingredients for success.

The strategic planning team chose to use the NowPLAN rubrics as the cornerstone of its strategic plan. It developed rubrics for each strategy included in the strategic plan. The NowPLAN rubrics were to be used as a guide and benchmarking tool. All teachers received professional development on using the NowPLAN rubrics for self-assessment. Administrators believed that familiarity with the rubrics would provide teachers with an understanding of the district's expectations (i.e., what the district has determined good practice to "look like"), and that use of the rubrics would encourage self-reflection and ultimately improvement. Embedding the NowPLAN rubrics into everyday practice, from the classroom teacher's use to the curriculum supervisor's reviews, was Nowgarden's way to translate the district's strategic plan into practice. The Nowgarden superintendent knew that the district's strategic plan,

including its technology plan, could drive change if it was a living, breathing plan that was incorporated as the foundation and into every aspect of the district's operation.

The superintendent also knew that an important aspect of using rubrics to understand expectations and drive change is consistency. Understanding of the rubrics must be uniform, and application of the rubric must be consistent. For this reason, the external evaluator was asked to compare externally completed observation-based rubric ratings (from case study classrooms) with self-report rubric ratings by classroom teachers. By doing this, the evaluator could uncover discrepancies and inconsistencies in understanding and application. These findings were to be provided to program staff to be used to plan professional development activities that aid in rubric use and to improve the reliability of rubric data.

The evaluator assured the district staff that the reporting of such data would in no way violate teacher confidentiality and privacy. The evaluation team planned to collect and manage data such that individual privacy was maintained. Data were to be stored with "dummy" keys that would allow linkages between data sets, but that would not relate to any internal, district identifier. Only the external evaluator would have access to key coding. Linkages between data sources, such as observational data and survey data, would be performed by the evaluation team and findings would be reviewed prior to release to ensure that individual identities could not be directly or deductively determined.

While summarizations of NowPLAN (including NowPLAN-T) rubric data would be provided to program staff for formative program changes, the evaluator planned to also use rubric data to relate implementation to long-term outcomes. The program theory laid out by district administration assumed that higher rubric scores would be positively related to higher student achievement scores. Similarly, they hypothesized that lower rubric scores (that is, less sophisticated levels of implementation) would be associated with lower student achievement scores.

NowPLAN-T Rubrics

Tables 19 through 25 represent the 1:1 Implementation Rubric. The William & Ida Friday Institute for Educational Innovation at North Carolina State University kindly granted permission to reproduce the 1:1 Implementation Rubric in Appendix B. The rubric was developed by research staff at the Friday Institute. The 1:1 Implementation Rubric is based on the International Society for Technology in Education's National Educational Technology Standards (ISTE's NETS) framework. It was also developed using the North Carolina IMPACT Guidelines, the Texas Star Chart, and the North Carolina Learning Technology Initiatives (NCLTI) Planning Framework.

The rubric provides an assessment of the daily impact and use of technology programs and services on the teaching and learning process. It can be used to examine technology programs

at the district level, as well as the school and classroom levels. The 1:1 Implementation Rubric is intended to aid in reflecting on your technology implementation. For more information, visit https://eval.fi.ncsu.edu/11-implementation-rubric/.

Although the programs, characters, schools, and school district mentioned in Appendix B are fictitious examples provided to illustrate how the principles in this guide can be applied, the 1:1 Implementation Rubric is a real instrument. The 1:1 Implementation Rubric serves as the NowPLAN-T evaluation rubric for the fictitious Nowgarden School District. The NowPLAN-T evaluation incorporates the rubric as a powerful indicator of classroom and school performance. The rubric will be used by teachers for self-assessment and by evaluators during observations of case study classrooms.

The NowPLAN-T evaluation will use the Friday Institute 1:1 Implementation Rubric to examine four dimensions of classroom technology use and teacher experience with technology. The four implementation areas are shown below:

1. Curriculum and Instruction
2. Infrastructure and Technical Support
3. Leadership, Administration, and Instructional Support
4. Professional Development

Each of the four implementation areas has six elements of reflection. The NowPLAN-T evaluator created the chart below, using the elements from the Friday Institute 1:1 Implementation Rubric.

Table 19: 1:1 Implementation Rubric: Implementation Areas and Elements of Reflection

Implementation Area	Element of Reflection
Curriculum & Instruction (CI)	Classroom Use Access to Digital Content Content Area Connections Technology Applications Student Mastery of Technology Applications Web-based Lessons
Infrastructure & Technical Support (IA)	Students: Computer Access/Connectivity Classroom Technology Technical Support LAN/WAN Student Access to Distance Learning
Leadership, Administration, & Instructional Support (LA)	Leadership and Vision Planning Instructional Support Communication and Collaboration Sustainability Policy
Professional Development (PD)	Professional Development Experiences Models of Professional Development Educator Capability Participation in Technology-Driven Professional Development Levels of Understanding Student Training

Scores on each element will range from 1 to 4. Rubric comparison points are awarded as follows for each level of technology implementation:

1 point = Early (Starting) Technology

2 points = Developing Technology

3 points = Advanced (Prepared) Technology

4 points = Target Computing (i.e., exemplary implementation)

A total classification score is calculated for each of the four implementation areas by adding the scores across the six elements. Thus, scores in each implementation area of the rubric can be a maximum of 24 points. These scores are then classified into one of four categories. A score of 21 to 24 points is considered "target," while a score of 15 to 20 points is "advanced/prepared." A classification of "developing" is assigned to a score of 9 to 14 points, and fewer than 9 points is considered "early/starting."

Using the Score Chart to Complete the Rubric

Using the elements from the Friday Institute 1:1 Implementation Rubric, the NowPLAN-T evaluator created the score chart below in Table B.5. The evaluator explained that to complete the rubric, you needed to consider each of the four implementation areas separately. For each implementation area, you would decide where the classroom's technology use and experience falls on each of the six elements comprising that dimension.

Each element has one or two characteristics that describe each level of technology implementation. These bulleted characteristics were developed by research staff at the Friday Institute and are shown in 1:1 Implementation Rubric in tables 22 through 25. Note that all characteristics describing a level of technology implementation must be achieved for points to be awarded for that level. For example, if both characteristics for the developing level accurately describe a classroom but characteristics for higher levels would not be accurate descriptions, then you grade that classroom as Developing. Use the score chart to record your scores, calculate the total score, and identify the classification level of implementation for each implementation area.

Table 20: 1:1 Implementation Rubric: NowPLAN-T Score Chart

Implementation Area	Element of Reflection	Early	Developing	Advanced	Target	TOTAL (6-24)	Classification (Circle One)
Curriculum & Instruction (CI)	CI1: Classroom Use	1	2	3	4		Target Advanced
	CI2: Access to Digital Content	1	2	3	4		

Implementation Area	Element of Reflection	Early	Developing	Advanced	Target	TOTAL (6-24)	Classification (Circle One)
	CI3: Content Area Connections	1	2	3	4		Developing Early
	CI4: Technology Applications	1	2	3	4		
	CI5: Student Mastery of Technology Applications	1	2	3	4		
	CI6: Web-based Lessons	1	2	3	4		
Infrastructure & Technical Support (IA)	IA1: Students: Computer	1	2	3	4		Target Advanced Developing Early
	IA2: Access/Connectivity	1	2	3	4		
	IA3: Classroom Technology	1	2	3	4		
	IA4: Technical Support	1	2	3	4		
	IA5: LAN/WAN	1	2	3	4		
	IA6: Student Access to Distance Learning	1	2	3	4		
Leadership, Administration, & Instructional Support (LA)	LA1: Leadership and Vision	1	2	3	4		Target Advanced Developing Early
	LA2: Planning	1	2	3	4		
	LA3: Instructional Support	1	2	3	4		
	LA4: Communication and Collaboration	1	2	3	4		
	LA5: Sustainability	1	2	3	4		
	LA6: Policy	1	2	3	4		
Professional Development (PD)	PD1: Professional Development Experiences	1	2	3	4		Target Advanced Developing Early
	PD2: Models of Professional Development	1	2	3	4		
	PD3: Educator Capability	1	2	3	4		
	PD4: Participation in Technology-Driven Professional Development	1	2	3	4		

Implementation Area	Element of Reflection	Early	Developing	Advanced	Target	TOTAL (6-24)	Classification (Circle One)
	PD5: Levels of Understanding	1	2	3	4		
	PD6: Student Training	1	2	3	4		

In the following pages, Table 21 shows the score chart developed by research staff at the Friday Institute. The chart developed by the Friday Institute with bulleted characteristics for curriculum and instruction is presented in Table 22; for infrastructure and technical support in Table 23; for leadership, administration, and instructional support in Table 24; and for professional development in Table 25.

These tables have been reformatted from their original versions to fit the pages.

These evaluation instruments were identified, modified, or developed through support provided by The Friday Institute. The Friday Institute grants you permission to use these instruments for educational, non-commercial purposes only. You may use an instrument as is, or modify it to suit your needs, but in either case you must credit its original source. By using these instruments, you agree to allow the Friday Institute to use the data collected for additional validity and reliability analysis. You also agree to share with The Friday Institute publications, presentations, evaluation reports, etc. that include data collected and/or results from your use of these instruments. The Friday Institute will take appropriate measures to maintain the confidentiality of all data. For information about additional permissions, or if you have any questions or need further information about these instruments, please contact Dr. Jeni Corn, Director of Evaluation of the Friday Institute, jeni_corn@ncsu.edu.

Table 21: The Friday Institute 1:1 Implementation Rubric 2: Score Chart

Curriculum & Instruction						**Total**
CI1 Classroom Use	CI2 Access to Digital Content	CI3 Content Area Connection	CI4 Technology Applications	CI5 Student Mastery of Technology Applications	CI6 Web-Based Lessons	
Infrastructure & Technology Support						
IA1 Students: Computer	IA2 Access/ Connectivity	IA3 Classroom Technology	IA4 Technical Support	IA5 LAN/WAN	IA6 Student Access to Distance Learning	
Leadership, Administration & Instructional Support						
LAI1 Leadership & Vision	LAI2 Planning	LAI3 Instructional Support	LAI4 Communication & Collaboration	LAI5 Sustainability	LAI6 Policy	
Professional Development						
PD1 Professional Development Experiences	PD2 Model of Professional Development	PD3 Educator Capability	PD4 Participation in Technology-Driven Professional Development	PD5 Levels of Understanding	PD6 Student Training	

1:1: Implementation Summary:

Implementation Area	**Total**	**Classification***
Curriculum & Instruction		
Infrastructure & Technology Support		
Leadership, Administration & Instructional Support		
Professional Development		

***Classification:** Early (Starting) Technology (6-8 pts.), Developing Technology (9-14 pts.), Advanced Technology (15-20 pts.), Target Computing (21-24 pts.)

If you have any questions or need further information about these instruments, please contact Dr. Jeni Corn, Director of Evaluation of the Friday Institute, jeni_corn@ncsu.edu.

THE WILLIAM & IDA FRIDAY INSTITUTE FOR EDUCATIONAL INNOVATION

THE WILLIAM & IDA FRIDAY INSTITUTE FOR EDUCATIONAL INNOVATION

If you have any questions or need further information about these instruments, please contact Dr. Jeni Corn, Director of Evaluation of the Friday Institute, jeni_corn@ncsu.edu.

Table 22: The Friday Institute 1:1 Implementation Rubric 3: Curriculum and Instruction

Curriculum & Instruction	Early (Starting) Technology	Developing Technology	Advanced (Prepared) Technology	Target Computing
CI1 **Classroom Use**	Teachers occasionally use technology to support instruction and present teacher-centered lectures. Students use technology for skill reinforcement.	Teachers use technology to drive instruction, improve productivity, and model technology skills. Students use technology to communicate and present information.	Teachers use technology as a collaborative tool in teacher-led and some student-centered learning experiences to facilitate the development of students' higher order thinking skills and to interact with content experts, peers, parents, and community. Students use technology to evaluate information and analyze data to solve problems.	Teachers and students are immersed in a student-centered learning environment where technology is seamlessly integrated into the learning process and used to solve real world problems. Students use technology to develop, assess, and implement solutions to real world problems.
CI2 **Access to Digital Content**	Teachers have occasional access to digital resources for instruction.	Teachers have regular access to digital resources in the classroom	Teachers have regular access to digital resources in various instructional settings (e.g., school, home, community).	Teachers have on demand access to digital resources anytime/anywhere.
CI3 **Content Area Connections**	Teachers use technology for basic skills practice with little or no connection with content objectives.	Teachers use technology to support content objectives.	Teachers integrate technology in subject areas.	Teachers seamlessly apply technology across all subject areas to provide learning opportunities beyond the classroom.

Curriculum & Instruction	Early (Starting) Technology	Developing Technology	Advanced (Prepared) Technology	Target Computing
CI4 **Technology Applications**	Teachers are aware of technology applications for grades K-12.	Teachers have a general understanding of appropriate technology applications for their content areas.	Teachers are knowledgeable of and consistently use appropriate technology applications for their content areas and grade levels.	Teachers seamlessly integrate technology applications in collaborative, cross-curricular units of instruction.
CI5 **Student Mastery of Technology Applications**	Up to 25% of students have mastered technology applications.	Between 26-50% or students have mastered technology applications.	Between 51-85% of students have mastered technology applications.	Between 86-100% of students have mastered technology applications.
CI6 **Web-Based Lessons**	Teachers use a few web-based activities with students.	Teachers have customized several web-based lessons, which include online standards-based content, resources, and learning activities that support learning objectives.	Teachers have created many web-based lessons, which include online standards-based content, resources, and learning activities that support learning objectives.	Teachers have created and integrate web-based lessons which include online standards-based content, resources, and learning activities that support learning objectives throughout the curriculum.

THE WILLIAM & IDA FRIDAY INSTITUTE FOR EDUCATIONAL INNOVATION

If you have any questions or need further information about these instruments, please contact Dr. Jeni Corn, Director of Evaluation of the Friday Institute, jeni_corn@ncsu.edu.

Table 23: The Friday Institute 1:1 Implementation Rubric 4: Infrastructure and Technical Support

Infrastructure & Technical Support	Early (Starting) Technology	Developing Technology	Advanced (Prepared) Technology	Target Computing
IA1 **Student:Computer**	Less than two (2) student computers available per classroom.	Two (2) to five (5) connected multimedia student computers available per classroom. At least one connected multimedia student lab or mobile cart is available.	Six (6) or more connected multimedia student computers available per classroom.	1 to 1 access to multimedia computers for all students in the classroom when needed. Ability to take computers home.
IA2 **Access/Connectivity**	No need to the Internet in the classroom.	Internet access to at least one computer in the classroom.	Direct Internet access with reasonable response time in the classroom.	Direct Internet connectivity in the classroom with adequate bandwidth to access e-learning technologies and resources for all students. Consistent access at home and school.

Infrastructure & Technical Support	Early (Starting) Technology	Developing Technology	Advanced (Prepared) Technology	Target Computing
IA3 **Classroom Technology**	Teachers have shared access to resources such as, but not limited to, digital cameras, PDAs, MP3 players, probes, interactive white boards, projection systems, scanners classroom sets of graphing calculators.	Teachers have access to a designated computer and shared resources such as, but not limited to, digital cameras, PDAs, MP3 players, probes, interactive white boards, projection systems, scanners classroom sets of graphing calculators.	Teachers have access to a designated computer, and dedicated and assigned use of commonly used technology such as, but not limited to, digital cameras, PDAs, MP3 players, probes, interactive white boards, projection systems, scanners classroom sets of graphing calculators.	Teachers have ready access to designated computer and a fully equipped classroom to enhance student instruction. Technologies include those earlier, as well as the use of new and emerging technologies.
IA4 **Technical Support**	When needed, the response time for technical support is greater than twenty-four (24) hours.	When needed, the response time for technical support is less than twenty-four (24) hours.	When needed, the response time for technical support is less than eight (8) hours.	When needed, the response time for technical support is less than four (4) hours.
IA5 **LAN/WAN**	Students and teachers have access to technologies such as print/file sharing and some shared resources outside the classroom.	Students and teachers have access to technologies such as print/file sharing, multiple applications, and district servers.	Students and teachers have access to technologies such as print/file sharing, multiple applications, and district-wide resources on the campus network.	All classrooms are connected to a robust LAN/WAN that allows easy access to multiple district-wide resources for students and teachers, including, but not limited to, video streaming and desktop videoconferencing.
IA6 **Student Access to Distance Learning**	Students have no or limited access to online learning with rich media such as streaming video, podcasts, applets, animation, etc.	Students have scheduled access to online learning with rich media such as streaming video, podcasts, applets, animations, etc.	Students have anytime access to online learning with rich media such as streaming video, podcasts, applets, animation, etc.	Students have anytime access to online learning with rich media such as streaming video, podcasts, applets, and animation, and sufficient bandwidth storage to customize online instruction.

THE WILLIAM & IDA FRIDAY INSTITUTE FOR EDUCATIONAL INNOVATION

If you have any questions or need further information about these instruments, please contact Dr. Jeni Corn, Director of Evaluation of the Friday Institute, jeni_corn@ncsu.edu.

Table 24: The Friday Institute 1:1 Implementation Rubric 5: Leadership, Administration, and Instructional Support

Leadership, Administration & Instructional Support	Early (Starting) Technology	Developing Technology	Advanced (Prepared) Technology	Target Computing
LAI1 **Leadership & Vision**	Leadership has the basic awareness of the potential of technology in education to lead to student achievement.	Leadership develops a shared vision and begins to build buy-in for comprehensive integration of technology leading to increased student achievement.	Leadership communicates and implements a shared vision and obtains buy0in for comprehensive integration of technology leading to increased student achievement. Distributive leadership facilitates sustainability of the initiative. A student leader is included in the planning team.	Leadership promotes a shared vision with policies that encourage continuous innovation with technology leading to increased student achievement. Teams of instructional, curriculum, technology, and administrative personnel to work together to goals and strategies for an effective 1:1 initiative.
LAI2 **Planning**	Few technology goals and objectives are incorporated in the school/district improvement plan.	Several technology goals and objectives are incorporated in the school/district improvement plan.	Technology-rich school district plan sets annual technology benchmarks based on the technology applications standards.	Leadership team has a collaborative, technology-rich school/district improvement plan grounded in research and aligned with district strategic plan focused on student achievement.

Leadership, Administration & Instructional Support	Early (Starting) Technology	Developing Technology	Advanced (Prepared) Technology	Target Computing
LAI3 **Instructional Support**	Teachers have limited opportunity for technology integration and planning or professional development.	Teachers have time for professional development on the integration of technology.	Teacher teams are provided time to create and participate in learning communities to stimulate, nurture, and support the use of technology to maximize teaching and learning.	Education leaders and teacher teams facilitate and support the use of technology to enhance instructional methods. On-demand, up-to-date student data is available to administrators and teachers to drive instructional decision-making.
LAI4 **Communication & Collaboration**	School leaders use technology for limited written communication with teachers and parents.	Technology is used for communication and collaboration among colleagues, staff, parents, students, and the community.	Current information tool and systems are used for communication, management of schedules and resources, performance assessment, and professional development. Technology is used to engage leaders from the business community.	Variety of media and formats, including telecommunications and the school website used to communicate, interact, and collaborate with all education stakeholders. Marketing strategies are used to engage the business community and seek volunteers to assist with promoting the initiative.

Leadership, Administration & Instructional Support	Early (Starting) Technology	Developing Technology	Advanced (Prepared) Technology	Target Computing
LAI5 **Sustainability**	Limited discretionary funds for implementation of technology strategies to meet goals and objectives outline in the school/district improvement plan.	Discretionary funds and other resources are allocated to advance implementation of some technology strategies to meet goals and objectives outlines in the school/district improvement plan.	Discretionary funds and other resources are allocated to advance implementation of most of the technology strategies to meet the goals and objectives outlined in the school/district improvement plan.	Discretionary funds and other resources are allocated to advance implementation of all the technology strategies to meet the goals and objectives outlined in the school/district improvement plan. A team of stakeholders is assembled to create a long-term funding plan for the initiative. These individuals include the district leadership team, local business partners, and outside business individuals).
IA7 **Policy**	Planning team is in a place to develop policies for ensuring student safety and appropriate use of computers.	Policies for ensuring student safety and appropriate use of computers are in place.	Policies are enforced for ensuring student safety and appropriate use of computers are in place.	Policies for ensuring student safety and appropriate use of computers in accord with the Children's Internet Protection Act (CIPA), while still enabling teachers and students access to a wide range of information and communication resources (AUP, plans for parent, teacher, student information, filtering, virus/spyware protection)

THE WILLIAM & IDA FRIDAY INSTITUTE FOR EDUCATIONAL INNOVATION

If you have any questions or need further information about these instruments, please contact Dr. Jeni Corn, Director of Evaluation of the Friday Institute, jeni_corn@ncsu.edu.

Table 25: The Friday Institute 1:1 Implementation Rubric 6: Professional Development

Professional Development	Early (Starting) Technology	Developing Technology	Advanced (Prepared) Technology	Target Computing
PD1 **Professional Development Experiences**	Teachers participate in professional development on basic technology literacy skills and district information systems.	Teachers have participated in professional development on integrating technology into content area activities for students as well as to streamline productivity and management tasks.	Teachers have participated in professional development on technology integration into the curriculum through the creation of new lessons and activities that promote higher order thinking skills and collaboration with experts, peers, and parents.	Teachers collaborate with other professionals in developing new learning environments to empower students to think critically to solve real-world problems and communicate with experts across business, industry and higher education.
PD2 **Models of Professional Development**	Teachers participate in large group professional development sessions to acquire basic technology skills.	Teachers participate in large group professional development sessions focusing on increasing teacher productivity and building capacity to integrate technology effectively into content areas with follow-up that facilitates implementation.	Teachers participate in on-going professional development, including training, observation/assessment, study groups, and mentoring.	Teachers participate in multiple professional development opportunities that support anytime, anywhere learning available through delivery systems including individually guided activities, inquiry/action research, and involvement in a development/improvement process.
PD3 **Educator Capability**	Educators are aware of the certification for technology applications.	Most educators meet two (2) to three (3) technology application standards.	Most educators meet four (4) to five (5) of the technology application standards.	Most educators meet all six (6) of the technology application standards.

Professional Development	Early (Starting) Technology	Developing Technology	Advanced (Prepared) Technology	Target Computing
PD4 **Participation in Technology-Driven Professional Development**	Teachers participate in less than nine (9) hours of technology professional development per year.	Teachers participate in nine (9) to eighteen (18) hours of technology professional development per year.	Teachers participate in nineteen (19) to twenty-nine (29) hours of technology professional development per year.	Teachers participate in thirty (30) or more hours of technology professional development per year.
PD5 **Levels of Understanding**	Teachers understand technology basics and how to use teacher productivity tools.	Teachers adapt technology knowledge and skills for content area instruction.	Teachers use technology as a tool in and across content areas to enhance higher order thinking skills.	Teachers create new, interactive, collaborative, and customized learning environments.
PD6 **Student Training**	Training on school technology policies and software is not provided to students.	Training on school technology policies and software is being planned for students.	Training on school technology policies and software is provided to students once a year.	Training on school technology policies and software is provided to students multiple times a year.

Appendix C: Evaluation Resources

The evaluation resources in this appendix can help you find more information on a topic. Note that many of these resources address multiple evaluation subjects, so the inclusion of a resource under one topic should not at all imply that it does not also pertain to other areas. There are many good evaluation texts, so you will undoubtedly find additional resources. This list is not exhaustive by any means. However, these resources will get you started if you are interested in a more in-depth look on a subject of interest.

Evaluation Approaches

Alkin, M. (2004). *Evaluation roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage Publications.

Fetterman, D., and Wandersman, A. (2005). *Empowerment evaluation principles in practice*. New York, NY: Guilford Press.

Patton, M. (2008). *Utilization focused evaluation*. (4th ed.) Thousand Oaks, CA: Sage Publications.

Paul, J. (2005). *Introduction to the philosophies of research and criticism of the education and the social sciences*. Upper Saddle River, NJ: Prentice-Hall.

Preskill, H., and Jones, N. (2009). *A practical guide for engaging stakeholders in developing evaluation questions*. Princeton, NJ: Robert Wood Johnson Foundation.

Rossi, P., Lipsey, M., and Freeman, H. (2004). *Evaluation: A systematic approach*. (7th ed.) Thousand Oaks, CA: Sage Publications.

Shadish, William R. Jr., Cook, Thomas D., and Leviton, Laura C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage Publications.

Stake, R. (2004). *Standards-based and responsive evaluation*. Thousand Oaks, CA: Sage Publications.

Stufflebeam, D. (2001). *Evaluation models: New directions in evaluation, No. 89*. San Francisco, CA: Jossey-Bass.

Program Theory and Logic Modeling

Frechtling, J. (2007). *Logic modeling methods in program evaluation*. San Francisco, CA: Jossey-Bass.

Knowlton, L., and Phillips, C. (2009). *The logic model guidebook*. Thousand Oaks, CA: Sage Publications.

Weiss, C. (1998). *Evaluation*. Upper Saddle River, NJ: Prentice Hall.

Research and Evaluation Design, Including Reliability and Validity

Haertel, G., and Means, B. (2003). *Evaluating educational technology: Effective research designs for improving learning*. New York, NY: Teachers College Press.

Lauer, P. (2006). *An education research primer: How to understand, evaluate, and use it*. San Francisco, CA: Jossey-Bass.

Means, B., and Haertel, G. (2004). *Using technology evaluation to enhance student learning*. New York, NY: Teachers College Press.

Mosteller, F., and Boruch, R. (2002). *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.

Shadish, W., Cook, T., and Campbell, D. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

Shavelson, R., and Towne, L. (2002). *Scientific research in education*. Washington, DC: National Academy Press.

Stake, R. (1995). *The art of case study research*. Thousand Oaks, CA: Sage Publications.

Stake, R. (2010). *Qualitative research: Studying how things work*. New York, NY: Guilford Press.

Trochim, W., and Donnelly, J. (2006). *The research methods knowledge base* (3rd ed.). Cincinnati, OH: Atomic Doc Publishing.

What Works Clearinghouse. (2011). *Procedures and standards handbook* (version 2.1). Retrieved from http://ies.ed.gov/ncee/wwc/references/idocviewer/doc.aspx?docid=19&tocid=1.

Wholey, J., Hatry, H., and Newcomer, K. (Eds.). (2010). *Handbook of practical program evaluation* (3rd ed.). San Francisco, CA: Jossey-Bass.

Threats to Validity

Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin Co.

Budgeting Time and Money

Bamberger, M., Rugh, J., and Mabry, L. (2011). *RealWorld evaluation*. Thousand Oaks, CA: Sage Publications.

Posavac, E., and Carey, R. (2003). *Program evaluation methods and case studies*. Upper Saddle River, NJ: Prentice Hall.

W. K. Kellogg Foundation. (2004). *Evaluation handbook*. Battle Creek, MI: W. K. Kellogg Foundation.

Ethical Issues

APA. (1982). *Ethical principles in the conduct of research with human participants*. Washington, DC: American Psychological Association.

National Center for Education Statistics. SLDS Technical Brief (NCES 2011-601), *Basic concepts and definitions for privacy and confidentiality in student education records*. Retrieved from http://nces.ed.gov/pubs2011/2011601.pdf.

National Center for Education Statistics. SLDS Technical Brief (NCES 2011-602), *Data stewardship: Managing personally identifiable information in electronic student education records*. Retrieved from http://nces.ed.gov/pubs2011/2011602.pdf.

U.S. Department of Education. (2007). *Mobilizing for evidence-based character education*. Washington, D.C.: Office of Safe and Drug-Free Schools. Retrieved from http://www2.ed.gov/programs/charactered/mobilizing.pdf.

U.S. Department of Education. *Privacy technical assistance center*. Retrieved from http://ptac.ed.gov/.

U.S. Department of Education. *Safeguarding student privacy*. Retrieved from http://www2.ed.gov/policy/gen/guid/fpco/ferpa/safeguarding-student-privacy.pdf.

Wilder Research. (2007). *Evaluation tip sheets: Ethical issues*. St. Paul, MN: Wilder Research. Retrieved from http://www.wilderresearch.org.

Yarbrough, D., Shulha, L., Hopson, R., and Caruthers, F. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage Publications.

See also general information, regulations, and guidance on the protection of human subjects at the U.S. Department of Education at http://www2.ed.gov/about/offices/list/ocfo/humansub.html.

Data Collection, Preparation, and Analysis

Bradburn, N., and Sudman, S. (2004). Asking questions: *The definitive guide to questionnaire development*. San Francisco, CA: John Wiley and Sons.

Cox, J., and Cox, K. (2008). Your *opinion, please!: How to build the best questionnaires in the field of education* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Fowler, F. (2008). *Survey research methods* (4th ed.). Thousand Oaks, CA: Sage Publications.

Friesen, B. (2010). *Designing and conducting your first interview*. San Francisco, CA: John Wiley and Sons.

Krueger, R., and Casey, M. (2009). *Focus groups: A practical guide for applied research* (4th ed.). Thousand Oaks, CA: Sage Publications.

Wilder Research. (2010). *Evaluation tip sheets: Making sense of your data*. St. Paul, MN: Wilder Research. Retrieved from http://www.wilderresearch.org.

See also evaluation publications, research syntheses, and technical assistance resources at the U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE) at http://ies.ed.gov/ncee.

Evaluation Pitfalls

Hall, G., and Hord, S. (2010). *Implementing change: Patterns, principles, and potholes* (3rd ed.). Boston, MA: Allyn and Bacon.

Wholey, J., Hatry, H., and Newcomer, K. (2010). *Handbook of practical program evaluation* (3rd ed.). San Francisco, CA: John Wiley and Sons.

Wilder Research. (2007). *Evaluation tip sheets: Does it measure Up?*. St. Paul, MN: Wilder Research.

Interpreting, Reporting, Communicating, and Using Evaluation Results

Hittleman, D., and Simon, A. (2005). *Interpreting educational research* (4th ed.). Upper Saddle River, NJ: Pearson.

Hord, S., Rutherford, W., Huling-Austin, L., and Hall, G. (1998). *Taking charge of change*. Austin, TX: Southwest Educational Development Laboratory.

McMillan, J. (2011). *Educational research: Fundamentals for the consumer* (6th ed.). Boston, MA: Allyn and Bacon.

McMillan, J., and Wergin, J. (2009). *Understanding and evaluating educational research* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Torres, R., Preskill, H., and Piontek, M. (1996). *Evaluation strategies for communicating and reporting: Enhancing learning in organizations*. Thousand Oaks, CA: Sage Publications.

Appendix D: Evaluation Instruments for Educational Technology Initiatives

Links to instruments that may be helpful to you for evaluating educational technology initiatives are provided below. As with **Appendix C: Evaluation Resources**, this list of evaluation resources is not meant to be exhaustive. The links are current as of publication date. However, should they no longer work at some point in the future, information on the author may help you to track them down. There are many useful evaluation instruments available, some in the public domain and some through private organizations. Hopefully these will provide you with a starting point as you build your instrument library.

Technology Policy Implementation Rubric (NCREL)

Location: http://www.air.org/focus-area/education/index.cfm?fa=viewContent&content_id=3006&id=10

Author: Learning Point Associates, North Central Regional Educational Laboratory (2004)

Description: The *Technology Policy Implementation Rubric* "can be used to assess a state's implementation of educational technology policies in 19 areas. The rating scale provides indicators for four levels of implementation: outstanding, high, medium, and low."

NETS for Students: Achievement Rubric (NCREL)

Location: http:// www.comfsm.fm/national/administration/VPA/researchdocs/techPlan/NCREL%20p-12rubric.pdf

Author: Learning Point Associates, North Central Regional Educational Laboratory (2005)

Description: The *NETS for Students: Achievement Rubric* defines "four achievement levels in relation to the NETS. The rubric is being developed to assist state and school district leaders in their efforts to measure and monitor the development of student technology literacy throughout the elementary and secondary grades."

NETS for Teachers: Achievement Rubric (NCREL)

Location: http://www.air.org

Author: Learning Point Associates, North Central Regional Educational Laboratory (2005)

Description: The *NETS for Teachers: Achievement Rubric* is the teacher counterpart rubric to the *NETS for Students: Achievement Rubric*.

1:1 Technology Implementation Rubric (NCSU-FI)

Location: https://eval.fi.ncsu.edu/11-implementation-rubric/

Author: The William and Ida Friday Institute for Educational Innovation, North Carolina State University (2010)

Description: A technology implementation rubric that "is based on Technology Standards & Performance Indicators for Students (ISTE NETS-S), the NC IMPACT Guidelines, Texas Star Chart, and NC Learning Technology Initiatives (NCLTI) Planning Framework. This rubric provides a global perspective of school media and technology programs at both the building and system levels."

Integration of Technology Observation Instrument (ASU-West)

Location: http://www.west.asu.edu/pt3/assessment/observation.htm

Author: University of Arizona West PT3 (2002)

Description: The Integration of Technology Observation Instrument includes a preobservation form to be completed by a classroom teacher, and a timed observation form and postobservation form to be completed by the observer. It is a component of the University of Arizona West *Preparing Tomorrow's Teachers to use Technology* (PT3) project. More information can be found at http://www.west.asu.edu/pt3. Please note that this instrument is copyrighted and written permission of the authors must be obtained prior to use.

The Technology Integration Matrix (USF)

Location: http://fcit.usf.edu/matrix/matrix.php

Author: University of South Florida, College of Education, Florida Center for Instructional Technology (2011)

Description: The Technology Integration Matrix (TIM) is an online, multimedia tool that examines classroom technology use in order to develop a common vocabulary to describe levels of technology integration. The matrix assesses five characteristics of effective learning environments: active, constructive, goal directed, authentic, and collaborative. These are measured against five levels of technology integration: entry, adoption, adaptation, infusion, and transformation. Each of the resulting 25 boxes includes a written explanation and a classroom video example for math, science, social studies, and language arts. Text-only versions of the online matrix are also available. Two additional tools are also described, though online versions are not available. The first is the Technology Integration Matrix Observation Tool (TIM-O), which contains yes/no questions designed to evaluate classroom technology integration using the

terms of the Technology Integration Matrix. The second, the Technology Comfort Measure (TCM), is a teacher self-assessment containing 35 questions and photographs showing classroom technology use. Results provide teachers with a profile and professional development suggestions. More information can be found at http://fcit.usf.edu/matrix/index.php.

The Looking for Technology Integration Protocol (UNC-Greensboro)

Location: http://www.serve.org/uploads/docs/LoFTI_1.1.pdf

Author: University of North Carolina at Greensboro SERVE Center, North Carolina Department of Public Instruction Educational Technology Division (2005)

Description: The Looking for Technology Integration (LoFTI) protocol is an observation tool that profiles school educational technology implementation. Observation Record Forms may be used to assess the learning environment, teaching and learning styles, student engagement, use of technology, and hardware and software use. A Data Tally Tool is available for presenting data (http://www.serve.org/uploads/docs/LoFTIpaperpencilAnalysis.pdf). More information can be found at http://www.serve.org/lofti.aspx.

School Technology Needs Assessment (UNC-Greensboro)

Location: http://www.serve.org/uploads/docs/STNA3.0.0.pdf

Author: University of North Carolina at Greensboro SERVE Center, North Carolina Department of Public Instruction Educational Technology Division (2006)

Description: The School Technology Needs Assessment (STNA) is designed for planning and formative evaluation. It is available as an online or paper-pencil survey tool. Questions address conditions for technology use, professional development, and classroom practices. A checklist for STNA preparation is available at http://www.serve.org/uploads/docs/STNAchecklist.pdf. The STNA website is located at http://www.serve.org/stna.aspx. A guide for interpreting STNA findings, as well as additional resources (including instructions for using the online survey) can be found on the STNA website.

Additional resources: Corn, J. (2007, November). *Investigating the validity and reliability of the School Technology Needs Assessment (STNA)*: http://www.serve.org/uploads/docs/STNA_paper.pdf

Profiling Educational Technology Integration (SETDA)

Location: http://www.setda.org/web/guest/petitools

Author: State Educational Technology Directors Association (SETDA), developed in partnership with Metiri Group (2004)

Description: Profiling Educational Technology Integration (PETI) Evaluating Educational Technology Effectiveness includes a set of tools designed to examine school, district, and state use of educational technology over time. Assessment is focused on both technology readiness and effective use of technology, and is aligned with No Child Left Behind, Title II, Part D. Tools in the PETI suite include a teacher survey, building-level survey, district survey, artifact review forms, principal interview protocol, classroom observation protocol, school walk-through protocol, school range-of-use observation tool, and teacher focus group protocol. An overview of and introduction to the SETDA/Metiri Group's *PETI - Evaluating Educational Technology Effectiveness*, as well as the PETI framework and information regarding reliability and validity, can be found at http://www.setda.org/web/guest/peti.

Levels of Technology Implementation (DCET)

Location: http://www.dcet.k12.de.us/instructional/loti/index.shtml

Author: Delaware Center for Educational Technology (1994/2003)

Description: The Levels of Technology Implementation (LoTi) instrument is a 37-question online teacher self-assessment. Questions address various topics, including teacher personal technology use, current instructional practice, and level of technology implementation. Information is reported individually to teachers and in aggregate to districts and the state. LoTi is funded by the Delaware Center for Educational Technology (DCET) and is available to Delaware Public and Charter schools. Other schools may contact DCET for information about the LoTi instrument.

Insight (TCET)

Location: http://www.tcet.unt.edu/insight/instruments/

Author: Texas Center for Educational Technology, University of North Texas (2005)

Description: Insight, the South Central Instrument Library and Data Repository, links to several useful tools for examining the use and integration of technology in classrooms and schools. Instruments include a variety of self-assessments, checklists, affective questionnaires, and technology implementation surveys. Information regarding Insight materials can be found at http://www.tcet.unt.edu/insight/about.php.

Observation Protocol for Technology Integration in the Classroom (NETC)

Location: http://www.netc.org/assessing/home/integration.php

Author: Northwest Educational Technology Consortium (2004)

Description: The Observation Protocol for Technology Integration in the Classroom (OPTIC) was developed by the Northwest Educational Technology Consortium (NETC) in 2004. OPTIC is an observation protocol that relies on checklists and rubrics to assess the degree of technology integration in classrooms and schools. Federal funding for the regional technology consortia program ended in September 2005. However, the Northwest Regional Educational Laboratory (NWREL), now Education Northwest, continues to make the OPTIC resources available to educators.

Capacity Building Instruments (NCSU-FI)

Location: https://eval.fi.ncsu.edu/instruments-2/

Author: The William and Ida Friday Institute for Educational Innovation, North Carolina State University

Description: The Friday Institute for Educational Innovation (FI) at North Carolina State University has compiled a list of instruments for evaluation capacity building. Instruments include inventories, checklists, surveys, and rubrics, including versions of LoFTI and STNA modified for special use. The site includes links to instruments, as well as instructions for using the instrument. A form is also provided for requesting permission to use instruments included on the site.

IMPACT Surveys (Sun Associates/Alabama State Department of Education)

Location: http://www.sun-associates.com/index.html

Author: Sun Associates

Description: The Alabama State Department of Education, Technology Initiatives Section, uses two sets of surveys to assess and monitor the impact of its state technology plan. The first set of surveys, Indicators for Measuring Progress in Advancing Classroom Technology (IMPACT), examines growth in technology use and perceptions towards technology. IMPACT surveys are self-report and are conducted with teachers, administrators, and technology coordinators. The second set, Speak Up surveys, are part of a national research project conducted by Project Tomorrow (see below). For more information on Alabama educational technology initiatives, see http://www.alsde.edu/html/sections/section_detail.asp?section=61&footer=sections.

Speak Up Surveys (Project Tomorrow)

Location: http://www.tomorrow.org/speakup/index.html

Author: Project Tomorrow (2011)

Description: Speak Up is a group of surveys available to schools and districts. Speak Up surveys are conducted online and are voluntary. Nationally aggregated data are available for comparison purposes. Speak Up includes self-report surveys for teachers, students, administrators, and parents. The Speak Up website includes information about the project, directions on how to participate, and sample survey questions.

School 2.0 ETOOLKIT (CTL)

Location: http://etoolkit.org/etoolkit/

Author: Center for Technology in Learning at SRI International

Description: The School 2.0 eToolkit was created by the Center for Technology in Learning (CTL) at SRI International. It includes an online reflection tool comprised of teacher, principal, and technology coordinator questionnaires. The reflection tool focuses on skills in technology integration and identifies areas for growth. Resources relating to several categories, including planning and implementation—technology evaluation, are also provided on the website. The tool kit is currently maintained by the Central Susquehanna Intermediate Unit (CSIU) and content is provided by the International Society for Technology in Education (ISTE).

School Technology and Readiness Chart (Texas Education Agency)

Location: http://starchart.epsilen.com/docs/TxTSC.pdf

Author: Texas Education Agency (2006)

Description: The School Technology and Readiness (STaR) Chart was created by Texas Education Agency to help teachers in Texas to self-assess their progress toward meeting state technology goals. The STaR Chart measures technology integration in four areas: teaching and learning; educator preparation and development; leadership, administration, and instructional support; and infrastructure for technology. More information on the Texas STaR Chart can be found at http://starchart.epsilen.com/.

Appendix E: Evaluation Templates

Figure 6: Logic Model Template

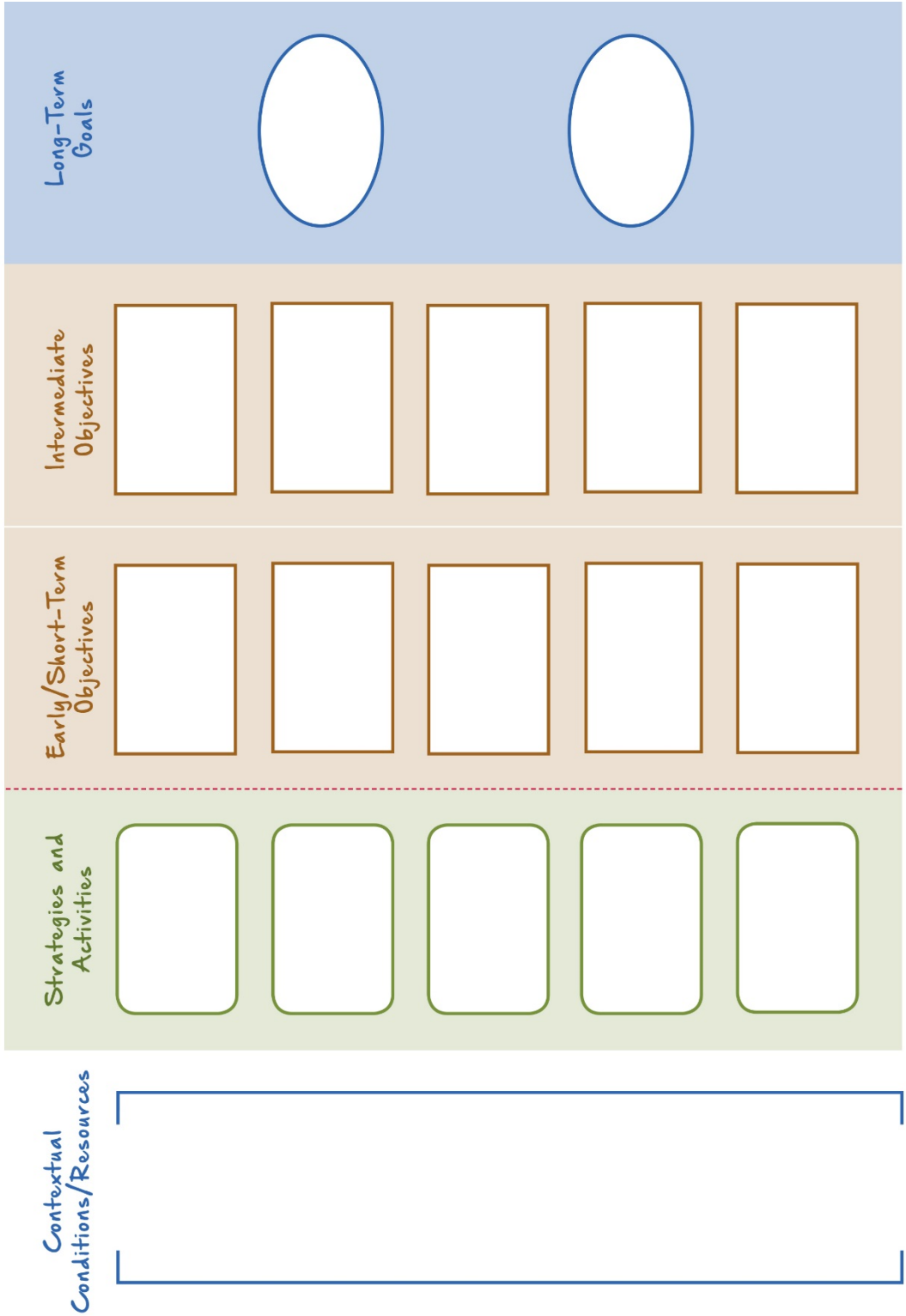


Table 26: Evaluation Matrix Template

	Logic Model Components	Evaluation Questions	Indicators	Targets	Data Sources	Data Collection	Data Analysis
Strategies and Activities/ Initial Implementation							
Early/Short-term and Intermediate Objectives							
Long-term Goals							

Appendix F: Lists of Tables and Figures

List of Tables

Table 1: Evaluation Matrix Example Shell 33
Table 2: Evaluation Methods and Tools: Overview 48
Table 3: Evaluation Methods and Tools: Procedures 54
Table 4: Evaluation Questions for Strategies and Activities 80
Table 5: Evaluation Questions for Early/Short-Term and Intermediate Objectives 81
Table 6: Evaluation Questions for Long-Term Goals 83
Table 7: Evaluation Matrix Addressing Strategies and Activities During the Initial Implementation—Indicators and Targets 85
Table 8: Evaluation Matrix Addressing Early/Short-Term and Intermediate Objectives—Indicators and Targets 86
Table 9: Evaluation Matrix Addressing Long-Term Goals—Indicators and Targets 89
Table 10: READ Evaluation Matrix—Strategies and Activities/Initial Implementation 95
Table 11: Evaluation Matrix—Early/Short-Term and Intermediate Objectives 98
Table 12: Evaluation Matrix—Long-Term Goals 104
Table 13: READ Evaluation Results—Strategies and Activities/Initial Implementation 109
Table 14: READ Evaluation Results—Early/Short-Term and Intermediate Objectives 112
Table 15: READ Evaluation Results—Long-Term Goals 116
Table 16: NowPLAN-T Evaluation Matrix—Strategies and Activities/Initial Implementation ... 134
Table 17: NowPLAN-T Evaluation Matrix—Early/Short-Term and Intermediate Objectives 135
Table 18: NowPLAN-T Evaluation Matrix—Long-Term Goals 143
Table 19: 1:1 Implementation Rubric: Implementation Areas and Elements of Reflection 148
Table 20: 1:1 Implementation Rubric: NowPLAN-T Score Chart 149
Table 21: The Friday Institute 1:1 Implementation Rubric 2: Score Chart 152
Table 22: The Friday Institute 1:1 Implementation Rubric 3: Curriculum and Instruction 153
Table 23: The Friday Institute 1:1 Implementation Rubric 4: Infrastructure and Technical Support 155
Table 24: The Friday Institute 1:1 Implementation Rubric 5: Leadership, Administration, and Instructional Support 157
Table 25: The Friday Institute 1:1 Implementation Rubric 6: Professional Development 160
Table 26: Evaluation Matrix Template 174

List of Figures

Figure 1: Embedded Evaluation Model..... 8
Figure 2: Possible Logic Model Headings..... 20
Figure 3: READ Logic Model..... 79
Figure 4: NowPLAN Logic Model..... 127
Figure 5: NowPLAN-T Logic Model..... 131
Figure 6: Logic Model Template 173