PROGRAM EVALUATION TOOLKIT

HANOVER RESEARCH

TABLE OF CONTENTS

TOOLKIT PURPOSE AND AUDIENCE

PURPOSE

Program evaluation requires a comprehensive approach to planning, implementation, and efficacy review to ensure fidelity of the program and to yield positive outcomes. Conducting an effective program evaluation helps answer two core questions:

- Has the program been successful in attaining the anticipated implementation objectives?
- Has the program been successful in attaining the anticipated participant outcome objectives?

AUDIENCE

This toolkit can be used by:

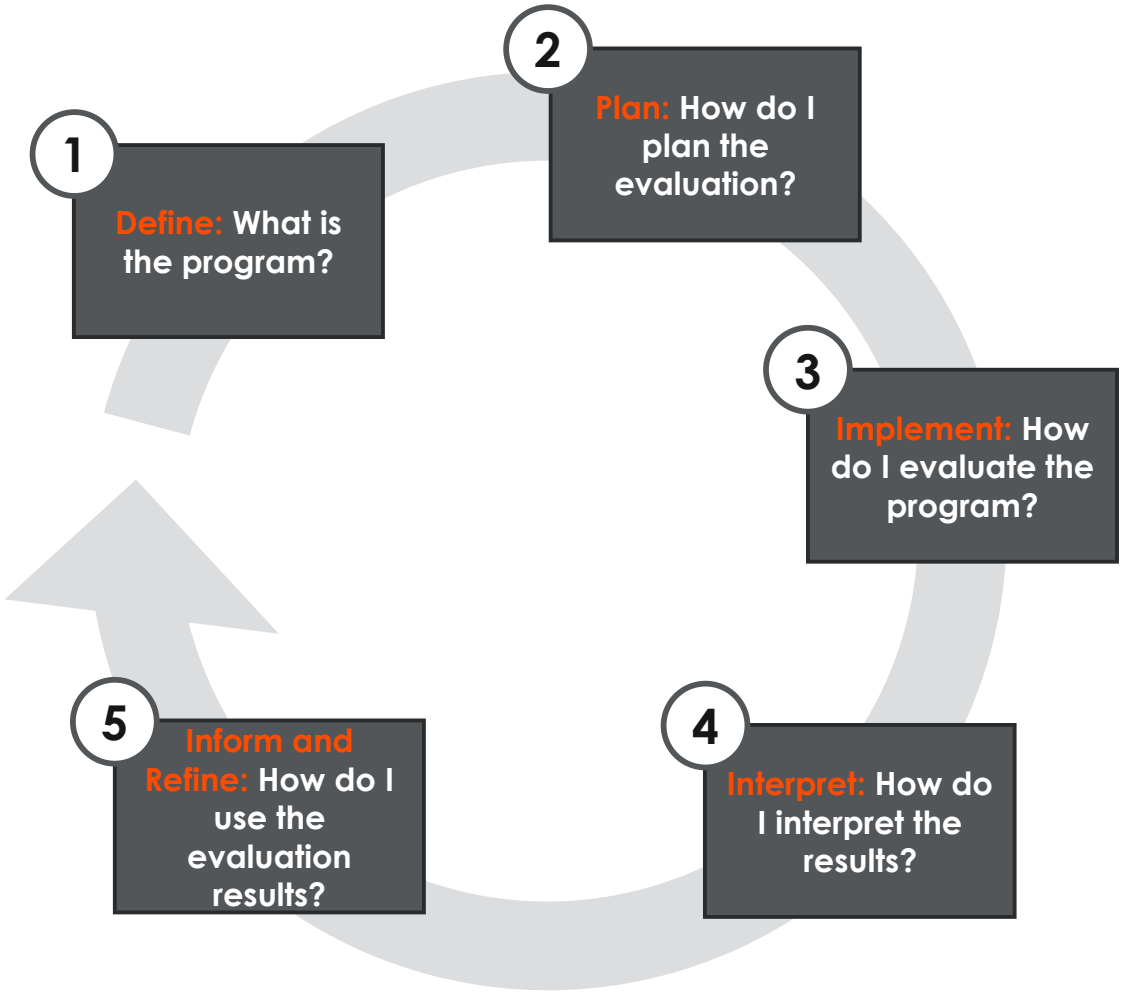
- Regional education service agencies
- School district leaders
- School or district teams tasked with program planning and implementation
- Program evaluation directors

Source: Goldstein, 2010, pp. 6-7.

THE PROGRAM EVALUATION PROCESS

The program evaluation process consists of five core steps:

1. **DEFINE**
2. **PLAN**
3. **IMPLEMENT**
4. **INTERPRET**
5. **INFORM AND REFINE**



Source: U.S. Department of Education, 2014, p. 9.

OVERVIEW

This toolkit provides strategies and resources for:

1. **Planning Program Evaluations**
2. **Implementing Program Evaluations**
3. **Determining Program Efficacy**


PLANNING FOR EVALUATION

OVERVIEW

Planning for a program evaluation requires the consideration of both logistical and strategic factors. The planning process should result in the identification of the following key factors, which will inform the evaluation goals and design:

Logistical Planning

IDENTIFY THE PROGRAM OR INITIATIVE Providing a brief description of the program or initiative being evaluated.

DEFINE THE PURPOSE Explaining why the program is being evaluated and the overarching goals of the evaluation.

DETERMINE THE END USERS Considering who will use the evaluation and how they will use it.

Strategic Planning

OUTLINE THE RESEARCH QUESTIONS Listing each of the research questions to address in the evaluation, in order of priority.

DEVELOP A TIMELINE Listing a general timeframe and/or dates of any key meetings at presentations, if known.

IDENTIFY THE REQUIRED RESOURCES Determining each resource, data point, assigned staff member, goals, outcomes, required analysis, potential challenges, and timeframe for each research question.

LOGISTICAL PLANNING

SELECTING RESEARCH-BASED PROGRAMS

Best Evidence Encyclopedia (BEE): A database of evidence-based programs reviewed by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education.

Doing What Works: A WestEd resource library including multimedia, interviews, case studies, and other educator-focused tools.

Center on Instruction: A resource collection including research studies, teacher tools, practitioner guides, professional development events/training, field examples, standards and assessments, and practitioner guides for various content areas and programs.

What Works Clearinghouse: A hub of evidence-based programs across content areas, education levels, and specialized foci.

Promising Practices Network: A collection of programs that are either proven to work, promising to work, or have otherwise been reviewed according to outcome-focused criteria.

These platforms include research-based and evidence-backed programs, initiatives, and activities to review from a starting point of program selection.

PROGRAM IDENTIFICATION

At the beginning of a program evaluation planning process, use the ***Program, Strategy, and Effort Organizer Tool*** to identify all major programs in the district that may be eligible for evaluation. Those included in the tool can be:

- ✓ Academic/behavioral/social-emotional improvement programs;
- ✓ Activities geared toward improving learning;
- ✓ Academic/behavioral/social-emotional initiatives; and
- ✓ Districtwide strategic initiatives.

Program, Strategy, and Effort Organizer Tool

Instructions: List all major programs currently in use in your school or department as well as any programs that you plan on launching in the near future. Major programs can comprise a variety of items, including, but not limited to, programs aligned with district priorities, programs funded by Title I, programs serving the district's neediest students, and programs that require significant investment of staff time or dollars.

Curriculum Initiatives	Instructional Strategies	Support for Teachers	Non-Academic Programs
E.g., reading programs, world language initiatives	E.g., interventions for struggling readers; class size reductions; alternative schools	E.g., teacher mentorship programs and focused PD initiatives	E.g., social worker programs, autism inclusion programs; parent engagement initiatives

Source: Strategic Data Project, p. 20.

PROGRAM PRIORITIZATION

In a collaborative Strategic Data Project, Knox County Schools, Fayette County Public Schools, Minneapolis Public Schools, and the Michigan Department of Education developed a ***Program Selection Rubric*** in their **Fellowship Capstone Project**.

After identifying all programs that could be eligible for evaluation, use the ***Program Selection Rubric*** to prioritize which programs or strategies may have the highest potential for evaluation.

Program Selection Rubric

Instructions: List the high-potential programs, strategies, or efforts, then score each of them.
Key: 0 = Not at all; 5 = Definitely

Program or Strategy	Primary Considerations (1-5 Scale)				Secondary Considerations (1-3 Scale)		Other Considerations (0-1 Scale)		Total
	Aligned to Strategy	Large Reach or Plans for Expansion	Significant Investment of Staff Time	Significant Investment of Funds	Direct Impact on Learning	Politically Feasible to Change	Data Unavailable	Uncertain Effectiveness	

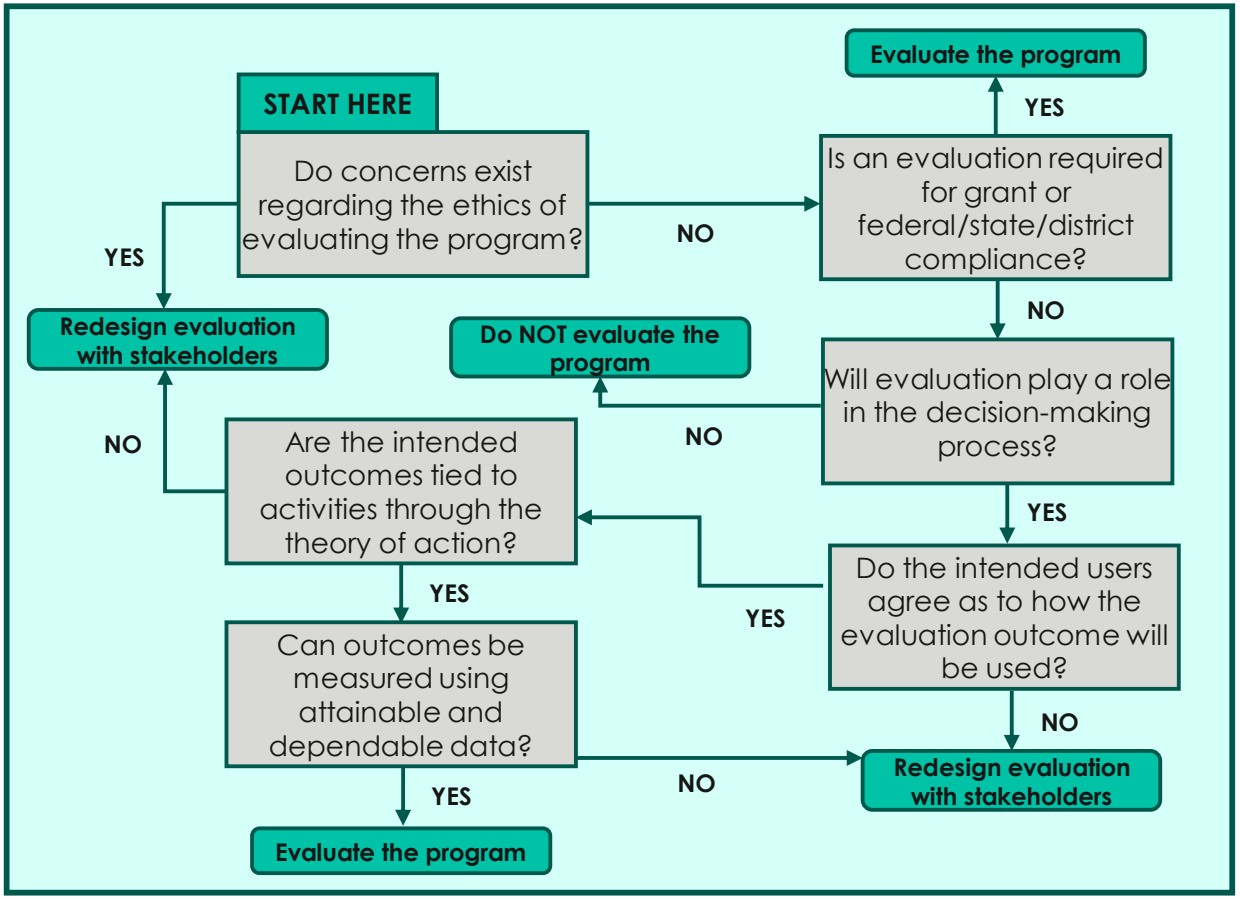
Source: Strategic Data Project, p. 43.

EVALUATION DECISION-MAKING

As an element of the **Fellowship Capstone Project**, Knox County Schools also includes a ***Program Evaluation Decision-Making Tree*** to guide evaluation directors to determine whether a program is a good candidate for evaluation.

After identifying potential programs for evaluation and selecting high-potential programs, use the ***Program Evaluation Decision-Making Tree*** to determine next steps in the process.

Program Evaluation Decision-Making Tree



Source: Strategic Data Project, p. 25.

DETERMINING PURPOSE

Evaluation directors should determine the purpose of evaluating the program. Determining the purpose directs subsequent steps in the planning process by aligning the purpose with the appropriate questions, objectives, and resources for evaluation.

NEEDS ASSESSMENT

PROCESS EVALUATION

ASSESSMENT OF PROGRAM THEORY

IMPACT OR OUTCOME EVALUATION

EFFICIENCY ASSESSMENT

In this stage of the planning process, evaluation directors can narrow their focus on what and how they evaluate the program in order to yield final outcomes from the process that answer the correct questions.

Source: Colorado Department of Education, 2016, p. 11.

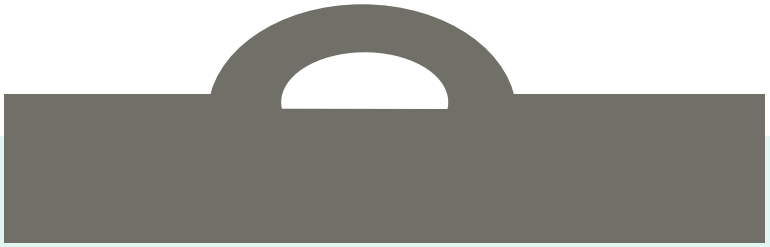
DEFINING PURPOSE AND SCOPE

After determining the purpose of the program evaluation, directors then define the scope, according to the chosen purpose.

NEEDS ASSESSMENT	Addresses whether conditions need to be resolved or improved or whether there is a need for the program.
ASSESSMENT OF PROGRAM THEORY	Determines whether a program is reasonable, feasible, ethical, or appropriate in concept or design.
PROCESS EVALUATION	Assesses program implementation, operation, or fidelity.
IMPACT OR OUTCOME EVALUATION	Evaluates whether programs achieve their intended or desired outcomes.
EFFICIENCY ASSESSMENT	Verifies the cost effectiveness of a program or assesses program costs.

Source: Colorado Department of Education, 2016, p. 11.

CONSIDERING AND ENGAGING STAKEHOLDERS



Consider the stakeholders and end-users of the program evaluation and which aspects of the evaluation align with their needs.

DISTRICT AND SCHOOL STAKEHOLDERS

- ❑ District administrators
- ❑ School officials
- ❑ Program staff
- ❑ Department of Education
- ❑ Families and parents
- ❑ Students

COMMUNITY AND OUTSIDE STAKEHOLDERS

- ❑ Program funders
- ❑ Potential funders
- ❑ Advocacy organizations
- ❑ Agency personnel
- ❑ Community program partners

Target the stakeholders who can impact program funding, credibility, and daily operation. Stakeholders can also provide guidance during the evaluation process and influence how to use evaluation results.

Sources: Administration for Children and Families, 2010; The Pell Institute; Ontario Centre of Excellence for Child and Youth Mental Health.

OTHER LOGISTICAL PLANNING TOOLS

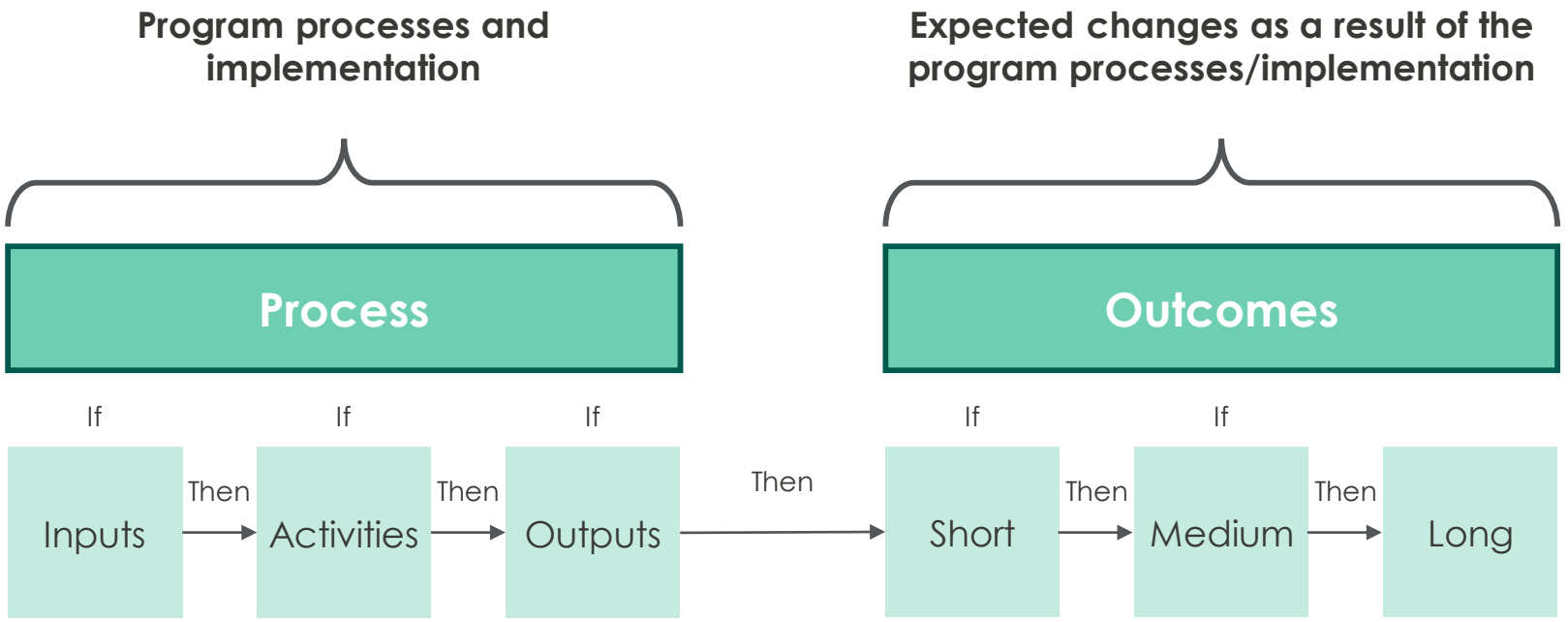
- **Scaling Barriers to Ensure Success in Program Evaluation** – Includes tools from school districts used in the logistical planning process.
- **Program Evaluation Toolkit** – *Phase I: Planning Evaluation* focuses on the logistical planning process.
- **Jump-Start Your School's Program Evaluation: Part 1** – Focuses on logistical planning at the school-level.
- **The Program Manager's Guide to Evaluation** – Designed to guide program managers through the planning process of program evaluation.
- **Best Practices in Program Evaluation** – Includes best practices in planning and implementing a program evaluation.

STRATEGIC PLANNING

PROGRAM LOGIC MODELS

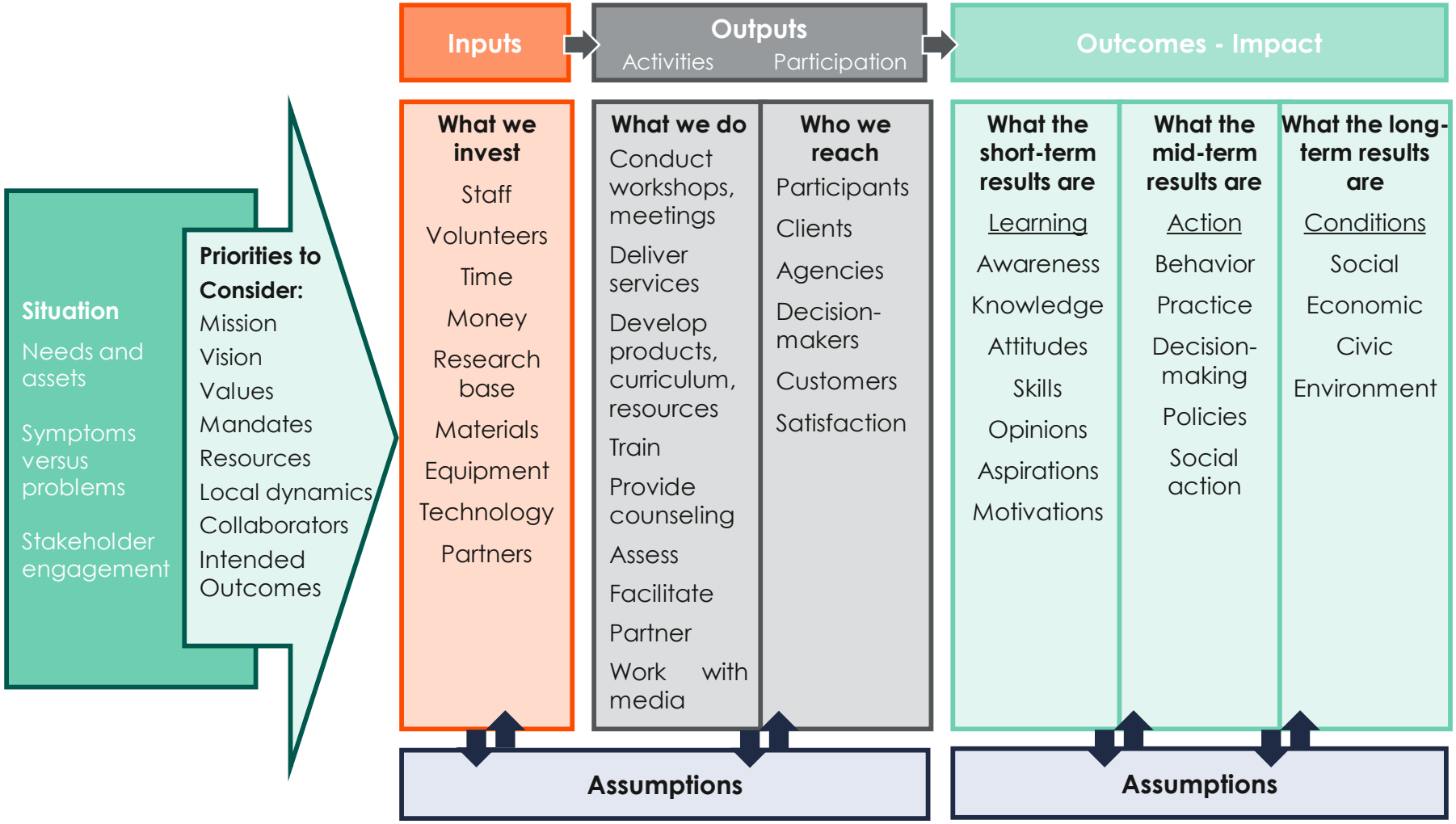
Use logic models to visually display and clarify the design of the program being evaluated to understand its full theory of change. Understanding all processes, activities, inputs, and outputs of the program and how they intend to impact short-, mid-, and long-term goals informs research questions.

*Use the **Logic Model Template** on the following slide to construct the logic model for the program chosen for evaluation.*



Source: Corporation for National & Community Service, p. 5.

LOGIC MODEL TEMPLATE



Source: Colorado Department of Education, 2016, p. 14.

LOGIC MODULE SAMPLES

Samples

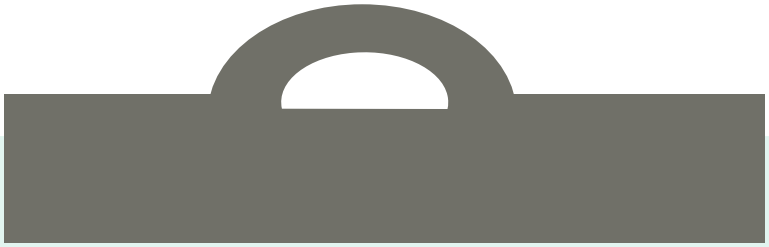
LOGIC MODELS: A TOOL FOR DESIGNING AND MONITORING PROGRAM EVALUATIONS – *Professional Learning Committees*
National Center for Education Evaluation and Regional Assistance, 2014

LOGIC MODELS: A TOOL FOR DESIGNING AND MONITORING PROGRAM EVALUATIONS – *Reading Strategy Program*
National Center for Education Evaluation and Regional Assistance, 2014

TECHNOLOGY AND SAFETY LOGIC MODEL
University of Wisconsin-Extension, 2009

TEACHING ABOUT THE ENVIRONMENT THROUGH SERVICE-LEARNING ACTIVITIES LOGIC MODEL
University of Wisconsin-Extension, 2009

LOGIC MODEL CHECKLIST



Use the following checklist to determine the quality of the logic model draft.

Question	Yes	Unsure	No
Do the outcomes represent meaningful benefits or changes for participants?			
Will the outcomes help you communicate the benefits of your program?			
Are your outcome goals clear and understandable?			
Are the outcomes participant-focused, rather than program-focused?			
Does your model include the outcomes of greatest importance to your key stakeholders?			
Is it reasonable, based on research, theory, or common-sense, that the program can influence outcomes in a substantial way?			
Does the model include all important program activities that participants receive?			
Does the model make appropriate connections between inputs, activities, outputs, and outcomes?			

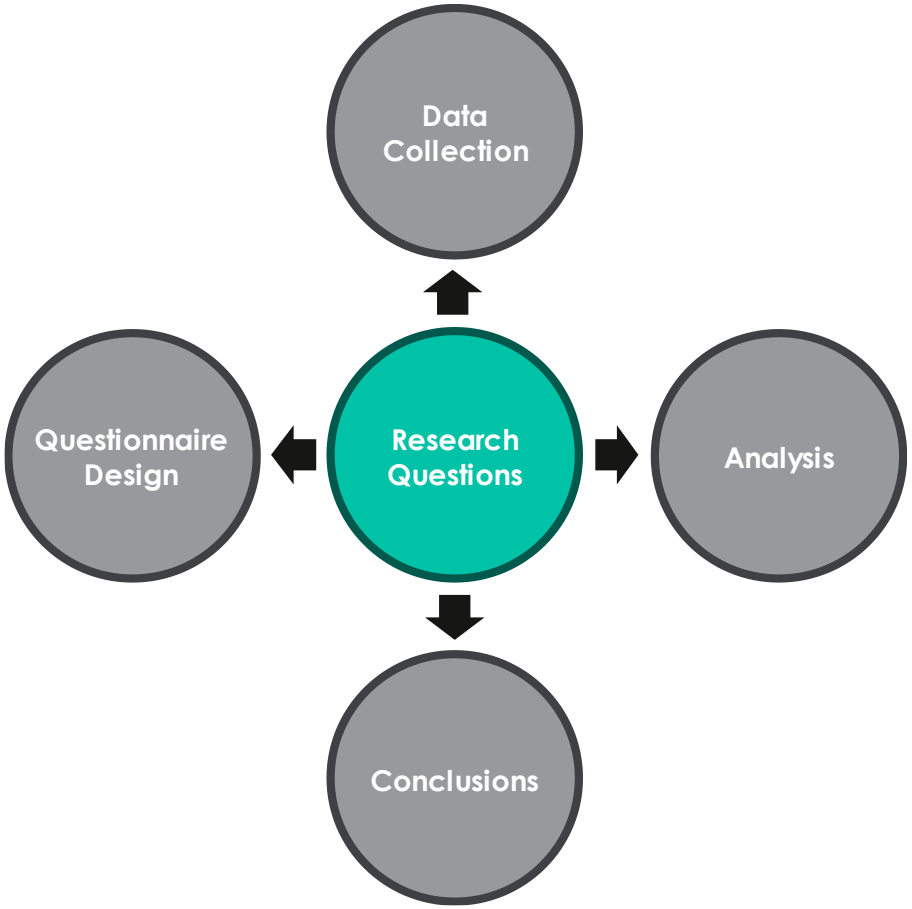
Source: Wilder Research, 2009.

OUTLINING RESEARCH QUESTIONS

Developing and outlining effective research questions for program evaluation requires a four-step process, driven by the logic model. Outline program evaluation research questions to inform data collection, analysis, questionnaire design, and conclusions.

CHECKLIST FOR OUTLINING RESEARCH QUESTIONS

- ❑ Test the program's theory of change as outlined in the evaluation logic model
- ❑ Align research questions with the specific purpose of the program evaluation
- ❑ Ensure research questions target any requirements or expectations from key stakeholders
- ❑ Design all research questions to be measurable
- ❑ Clearly define all research question to be targeted and specific



Source: Corporation for National and Community Service, p. 3.

OUTLINING QUESTIONS WITH PURPOSE

Review the following sample evaluation questions to determine which questions best align with the purpose of the evaluation.

Purpose	Sample Question
NEEDS ASSESSMENT	▪ What condition/situation/outcome is not working? In need of improvement? ▪ Why does that condition/situation/outcome exist? What is contributing to it?
ASSESSMENT OF PROGRAM THEORY	▪ Is our Theory of Action plausible based on the research literature? ▪ Do our stakeholders and/or subject matter experts think it's reasonable and feasible?
PROCESS EVALUATION	▪ Is the program reaching the targeted recipients? ▪ Is the program being implemented as planned/designed? ▪ Are implementation benchmarks being reached? ▪ How is the program progressing? Compared to last year? A month ago? ▪ What challenges have we faced? What improvements/changes in strategies are needed for us to reach intended outcomes?
IMPACT OR OUTCOME EVALUATION	▪ What were the intended outcomes of our program? What changes did we hope to achieve? Did we achieve those intended/hoped for outcomes? ▪ Did the program yield the same results for all participants? Was the program more effective for some? If so, why? ▪ What were (any) unintended outcomes of the program?
EFFICIENCY ASSESSMENT	▪ Do the benefits of our program outweigh the costs? ▪ Are we getting a return on our investment?

Additional Resources for Sample Research Questions

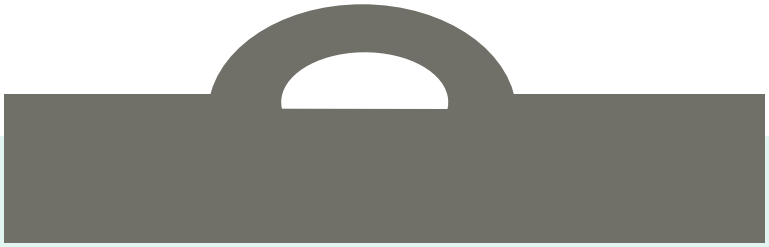
Specify the Key Evaluation Questions
Better Evaluation

How to Develop the Right Research Questions for Program Evaluation
Corporation for National & Community Service

Asking Program Evaluation Questions
American University

Source: Colorado Department of Education, 2016, p. 12.

OTHER STRATEGIC PLANNING TOOLS



- **Needs Assessment Research Questionnaire** – Aligned needs assessment survey questions with the appropriate stakeholder group.
- **Conducting a Needs Assessment** – Kentucky Department of Education's tool to guide districts through conducting a needs assessment.
- **Program Theory and Logic Models** – Assists directors in developing logic models by determining goals, developing evaluation activities, and constructing the model.
- **Best Methods for Evaluating Education Impact** – Discusses best practices for conducting outcome/impact evaluations.
- **Effectiveness and Efficiency Framework** – Focuses on evaluating programs for increasing student performance.
- **A Guide for Comprehensive Needs Assessment** – Walks directors through each step of the needs assessment process.

PLANNING FOR EVALUATION: WORKSHEET

PROGRAM EVALUATION PLANNING WORKSHEET

Below is a worksheet designed to facilitate your program evaluation planning process. These steps and worksheet create a framework to keep project evaluations organized, focused on common goals, and well-documented. The worksheet continues onto the next slide.

Step 1: What program or initiative are you planning to evaluate?

Provide a brief description:

WHAT ARE THE SELECTION CRITERIA FOR PARTICIPATION?
ARE THERE ANY POTENTIALLY CONFOUNDING VARIABLES?

Step 2: What is the purpose of evaluation?

Explain why the program is being evaluated and the overarching goals of the evaluation:

WILL THE PROGRAM REQUIRE FORMATIVE AND/OR SUMMATIVE EVALUATIONS?
WHAT OUTCOMES WILL BE MEASURED?

Step 3: Who will use the evaluation? How will they use it?

Table 1: Audience and Use of Evaluation

Who will use the evaluation? (Audience)	How will they use it?

Step 4: What key research questions will the evaluation seek to answer?

List succinctly in order or priority. These questions will be expanded upon in Step 6.

1. **3.** **5.**
2. **4.** **6.**

Source: Developed by Hanover Research.

PROGRAM EVALUATION PLANNING WORKSHEET

Step 5: When is the evaluation needed?

List a general timeframe and/or dates of any key meetings or presentations, if known.

Once steps **1-5** have been completed, the staff responsible for planning the evaluation may need to work with other divisions and/or external partners to complete the remaining steps of the planning process.

Step 6a: What resources will be required to answer the key research questions?

Fill out the tables below for each individual research question, creating additional copies of the table as needed. Use as many rows needed to describe each resource and/or data point that will be used.

Research Q. #1:

Information needed to answer	Source of information	Analysis required	Goals and outcomes of analysis	Staff responsibilities	Potential challenges

Step 6b: What stages are required for this evaluation? Who will be responsible for each stage? When will each stage be completed?

Use the table below to answer each question.

Stage	Staff responsible	Timeframe

Source: Developed by Hanover Research.

PROGRAM EVALUATION TOOLS AND PROFILES

Additional Program Evaluation Tools

- **Planning Guide for Online and Blended Learning**
- **OLC Quality Scorecard Suite** – Program evaluation scorecards for online learning
- **School Technology Needs Assessment**
- **Virtual Learning Program Rubric** – Rubric to evaluate virtual learning program quality and vigor
- **Evaluation Toolkit for Youth Programs**

Program Evaluation Profiles

Comprehensive Needs Assessment
Georgia Department of Education, 2017-2018

Conducting the Strengths and Needs Assessment
STEP Grant Program, 2011

IMPLEMENTING PROGRAM EVALUATIONS

OVERVIEW OF PROGRAM EVALUATION IMPLEMENTATION

Once planned, the implementation of a program evaluation relies on two key components: (1) process and design and (2) fidelity.

Implementation Process and Design

DETERMINE KEY CONSIDERATIONS IN DESIGN Considering which schools should pilot programs, which students will be eligible, and the timing of programs.

DEVELOP A FIVE-YEAR PLAN Outlining a step-by-step plan for the first five years of implementation.

Implementing with Fidelity

OUTLINE THE COMPONENTS OF FIDELITY Gauging program differentiation, adherence, exposure, quality, and responsiveness.

COLLECT AND EXAMINE THE DATA Conducting an analysis of fidelity through objective data collection techniques.

IMPLEMENTATION PROCESS AND DESIGN

KEY QUESTIONS

Schools to Receive the Program:

- Which schools should pilot the program?
- How will additional schools roll-out the program?

Rules for Students' Program Eligibility

- What should the cutoff(s) be for pre-program at-risk academic or behavioral outcomes?
- Should there be a subjective teacher recommendation component?

Amount of Time or Degree of Programming

- Should program dosage vary across subgroups of students?
- How to implement varying degrees of intervention (e.g., a three-tiered intervention)?

Source: Developed by Hanover Research.

POINTS TO CONSIDER

Schools to Receive the Program:

- Schools should be selected for programming in a way that reduces selection bias. For example, the program should not be implemented in all of the most at-risk schools first.
- School selection should reduce contamination with other programs.
- Choices in implementation may limit the availability of comparison groups.

Rules for Students' Program Eligibility

- A lottery for program participation would help identify a comparison group of non-program students, but entry into the lottery still presents a selection issue.
- Subjective measures are difficult to measure, make cutoffs unclear, and limit comparisons.

Amount of Time or Degree of Programming

- Varying dosage adds a dimension on which to evaluate program efficacy (which is good).
- However, in some cases, different program dosages or intensities are meant for different student subgroups.

Source: Developed by Hanover Research.

EVALUATION: FIVE-YEAR PLAN

A five-year plan to evaluate a program might start with formative analyses of program implementation and perceptions in the earlier years followed by summative analyses of program impact in the later years. This tables below and on the following slides describe steps for evaluation planning and execution by year.

Years of Implementation	Steps for Evaluation Planning and Execution
YEAR 0 (BEFORE PROGRAM)	▪ Solicit stakeholder (teachers, parents, etc.) feedback on program roll-out. ▪ Finalize program roll-out plan and determine criteria for student program eligibility. ▪ Determine the data points required for evaluation, including eligibility criteria. ▪ Identify data points that are not yet measured. ▪ **Do not implement the program until key outcome and control variables are measured.** ▪ Check the measured data for completeness (including available data in years prior to implementation). ▪ Identify research questions for the formative and summative evaluations of the program in later years. ▪ Plan and administer teacher training in program implementation. ▪ **Do not implement the program until all teachers of targeted students have been trained.**

Source: Developed by Hanover Research.

EVALUATION: FIVE-YEAR PLAN

The table below describes the two years immediately following program implementation. These years are best for formative evaluations of the program and subsequent adjustments.

Years of Implementation	Steps for Evaluation Planning and Execution
YEAR 1 (FIRST YEAR OF PROGRAM)	▪ Continue to collect and check required data points. ▪ If program participation follows rules (e.g., performance below a threshold score or percentile, only certain grade levels or classrooms implement the program), then data checks should encompass these points as part of a formative evaluation. ▪ Design, administer, and analyze surveys of program training, implementation to determine fidelity and perception. ▪ Make changes to training and implementation to ensure program fidelity. ▪ If perceptions indicate need to change program eligibility and roll-out, then document these changes and identify potential limiting effects on future summative evaluations, *or* reaffirm program eligibility to stakeholders to minimize changes to evaluation plan.
YEAR 2	▪ Continue data collection and quality check efforts. ▪ Update program training, implementation, and perception survey designs as needed. ▪ Administer and analyze updated surveys from Year 1. ▪ Make additional changes to training and implementation as needed.

Source: Developed by Hanover Research.

EVALUATION: FIVE-YEAR PLAN

The table below describes the years after the district has settled into program implementation, when there are no longer adjustments to program fidelity (labeled as "Year 4+"). Summative evaluations of program efficacy are more suitable for these later years.

Years of Implementation	Steps for Evaluation Planning and Execution
YEAR 3	▪ Continue data collection and quality check efforts. ▪ Conduct a quantitative summative evaluation of the program that uses a method that fits with program implementation. o A descriptive analysis of short-term trends might be more appropriate than a more rigorous analysis if not enough years of programming have passed. ▪ Related research questions may arise from the summative program evaluation (e.g., program effects for student subgroups of interest). Planning and conducting these additional studies will be important for a more complete understanding of program impact. ▪ Continue to gauge stakeholder perceptions of the program.
YEAR 4+	▪ Continue data collection and quality check efforts. ▪ Conduct quantitative and qualitative research to build on Year 3. ▪ When no remaining questions about the program exist, create a capstone report of the program impact, draw in related research on the return-on-investment of the program, and form recommendations for whether to continue the program.

Source: Developed by Hanover Research.

IMPLEMENTING WITH FIDELITY

IMPLEMENTATION WITH FIDELITY

When developing and implementing programs, include measures of fidelity during and after the implementation process to ensure that the final program operates and was delivered as intended.

Relying solely on student outcome data does not necessarily indicate how well the program was implemented; evaluation directors should **test how well the implementation aligns with the intended program to determine whether student assessment results are an accurate reflection of program effectiveness.**

In gauging the level of implementation fidelity, consider the following questions:

- **Program differentiation -** How well did the implemented program result in successful mastery of student learning outcomes?
- **Adherence -** Was the intended program delivered?
- **Exposure -** Were students exposed to the full program?
- **Quality -** How well was each element of the program delivered or implemented?
- **Responsiveness -** To what degree are students engaged or actively participating?

Source: James Madison University.

IMPLEMENTATION FIDELITY DATA

Consider each of the four common ways to collect implementation fidelity data and choose those that best align with the program's design and available resources.

Self Report

Program facilitators evaluate their own program; students evaluate their own responsiveness.

Advantages: Time and cost efficient (no additional staff or equipment needed)

Disadvantages: Risk of desirability bias (students may feel pressure to report high responsiveness and facilitators may feel pressure to indicate perfect adherence to the planned program.)

Outside Observation

Trained, independent evaluators (often posing as participants) observe and evaluate the program.

Advantages: Outside observers are less subject to social desirability; outside observers actually experience the program as a participant.

Disadvantages: Time and cost intensive. Auditing an entire program may take a long time. Additionally, hiring outside observers may be expensive.

Audio Recording

The program is audio recorded and reviewed by one or more evaluators at a later date.

Advantages: Cost effective, convenient, ability to review the data multiple times, ability to use multiple evaluators, ability to have "blind" raters.

Disadvantages: Limits observation – loss of visual data; reactivity to recording (students and/or facilitators may act differently if they know they are being recorded.

Video Recording

The program is video recorded and reviewed by one or more evaluators at a later date.

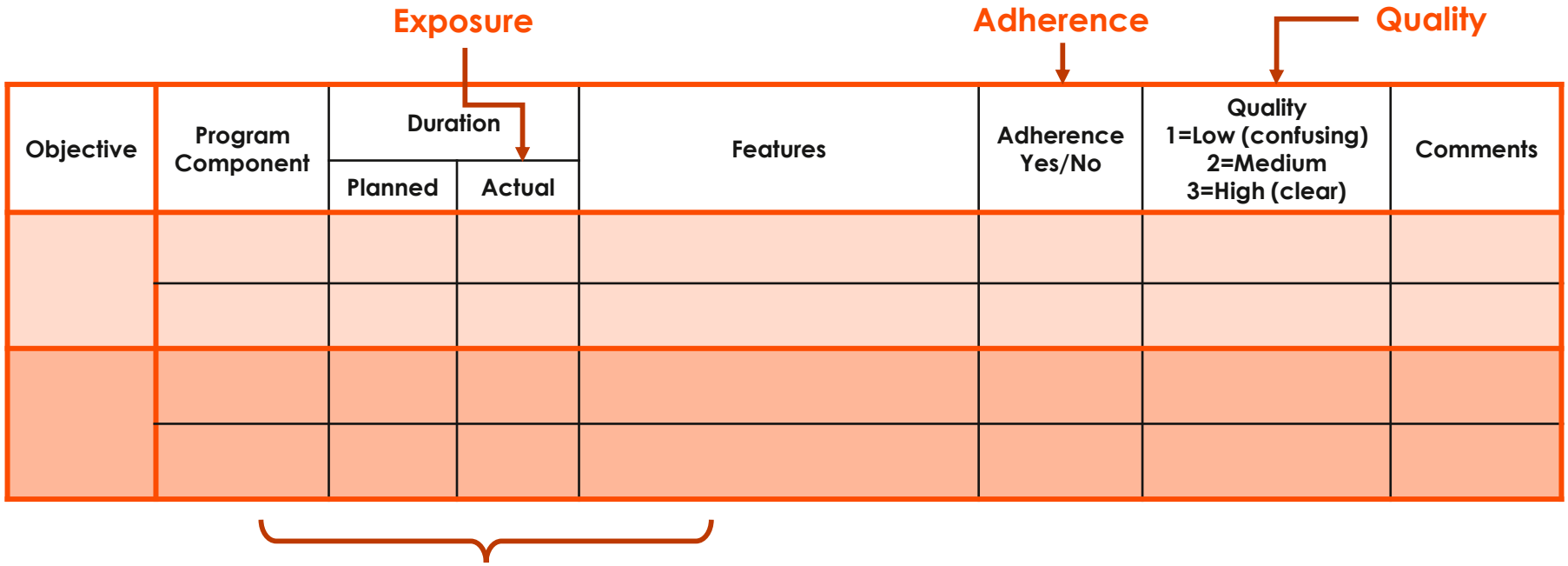
Advantages: Relatively cost effective, convenient, ability to review the data multiple times, ability to use multiple evaluators, richer depiction of the environment.

Disadvantages: Camera costs, reactivity to camera (students and/or facilitators may act differently if they know they are being recorded).

Source: James Madison University.

IMPLEMENTATION FIDELITY CHECKLIST

Use the following ***Implementation Fidelity Checklist*** to determine whether each objective and program component adheres to a high standard of quality and responsiveness. The checklist helps answer how well the program was delivered through four of the five components: program differentiation, exposure, adherence, and quality. Evaluators can assess the fifth component, responsiveness, through surveys administered to participants.



Additional Fidelity Checklists
RTI Action Network

Source: "Fidelity Workshop," James Madison University, p. 47.

IMPLEMENTATION FIDELITY CHECKLIST EXAMPLE

The example ***Implementation Fidelity Checklist*** below, from a leadership development program, illustrates how to use the tool to break down each objective and program component to assess the four components.

Objective	Program Component	Duration		Features	Adherence Yes/No	Quality 1=Low (confusing) 2=Medium 3=High (clear)	Comments
		Planned	Actual				
As a result of the Leadership Development Course, students will demonstrate improved leadership skills.	Presenter A leads discussion on "Challenging the Process"	45 min	50 min	Presenter explains that leaders seek new things they could learn and treat every day as if it were the first day of work.	Yes	2	
				Presenter has class note five things they could learn and share one with the class.	Yes	3	
	Presenter B leads discussion on "Inspiring Shared Vision"	45 min	45 min	Presenter details the importance of asking others about goals for the future.	Yes	2	
				Students are split into groups to share goals for the future with each other.	Yes	2	
	Presenter A discusses "Enabling Others to Act"	45 min	60 min	Presenter explains that important tasks should be delegated, particularly to those who could use practice to hone skills.	Yes	3	
				Students practice delegation with group members in hypothetical scenario	Yes	3	

Source: "Fidelity Workshop," James Madison University, p. 51.

EVALUATING LEARNING OUTCOMES WITH IMPLEMENTATION FIDELITY DATA

After conducting an implementation fidelity study and reviewing student learning outcomes (SLO) data, use the following matrix to determine whether the program was an effective determinant of student outcomes.

Outcomes Assessment	Implementation Fidelity: Low	Implementation Fidelity: High
Poor	The program was not implemented as planned and the SLOs were not met. Do not claim that current study was not effective. A new study should be conducted with high implementation fidelity to assess the effectiveness of the intended program.	The program was implemented as planned but the SLOs were not met. Low implementation fidelity can be ruled out as an explanation for why SLOs were not achieved. Some other reason contributed to low performance.
Good	The program was not implemented as planned but the SLOs were achieved. The intended program cannot be credited with contributing to student achievement. Further investigation on why students met SLOs is warranted.	The program was implemented as planned and the SLOs were met. There is evidence that the intended programming may be effective at achieving the SLOs.

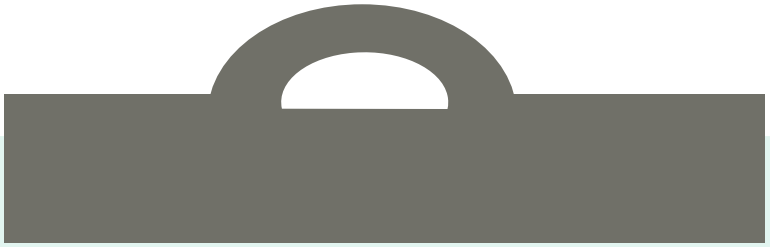
Source: Gerstner and Finney, 2013, p. 23.

VIDEO: IMPLEMENTATION FIDELITY

The video **Implementation Fidelity**, developed by James Madison University, can assist evaluation directors in:

- ✓ Understanding what implementation is and how it affects overall program evaluation.
- ✓ Considering which aspects of implementation fidelity to assess.
- ✓ Constructing an implementation fidelity checklist to assess each contributing factor.

FIDELITY OF IMPLEMENTATION TOOLS



- **Fidelity of Implementation:** A learning module from Vanderbilt University discussing the fidelity of implementing evidence-based practices and programs.
- **TIPS Fidelity of Implementation Checklist**: A checklist of tasks for implementation teams to complete to ensure that they conduct all elements of implementing a program evaluation with fidelity across responsibilities.
- **Implementation Fidelity Workshop**: A workshop slide deck that introduces how to incorporate implementation fidelity into the overall program evaluation cycle. This workshop includes an applied example at the university level.
- **Approaches to Measuring Implementation Fidelity in School-Based Program Evaluations**: An article that discusses various approaches to assessing implementation fidelity.
- **Measuring Implementation Fidelity:** An article demonstrating the importance of including implementation fidelity in the program evaluation implementation process
- **RTI Fidelity of Implementation Rubric**: A rubric that assists evaluators in assessing the fidelity of implementation specifically for RTI programs.
- **Fidelity of Implementation of the TASC ExpandED Model in 2011-12**: An analysis of the fidelity of implementation for a whole-school reform model in schools throughout the country.
- **Fidelity of Implementation Guidebook**: A guidebook for teachers for implementing 21st Century Skills with fidelity.

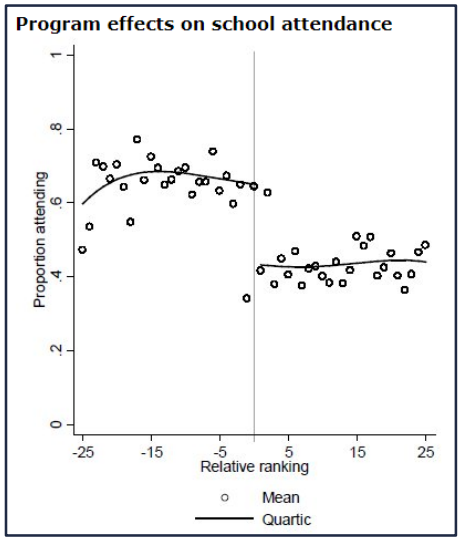
DETERMINING PROGRAM EFFICACY

OVERVIEW OF QUANTITATIVE METHODS

This section describes three methods of quantitative analysis, as defined below, for use in program evaluations. During the evaluation design, determine which quantitative methods best suit the program, evaluation goals, and available resources.

Regression Discontinuity (RD)

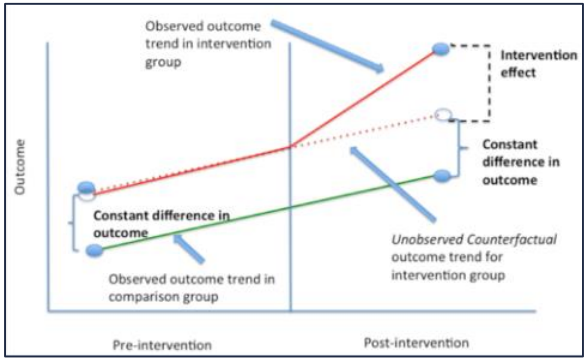
RD uses a pre-test, post-test two-group design to measure the impact of a program (treatment) based on a continuous eligibility index.



Source: "Regression Discontinuity," Better Evaluation.

Difference-in-Difference (DiD)

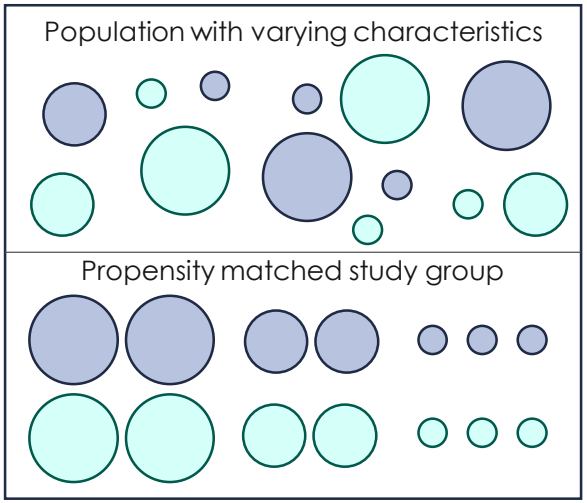
DiD compares the pre-program and post-program results of participants to the results of non-participants. Using DiD, participants are not randomly assigned.



Source: Columbia University Mailman School of Public Health.

Propensity Score Matching (PSM)

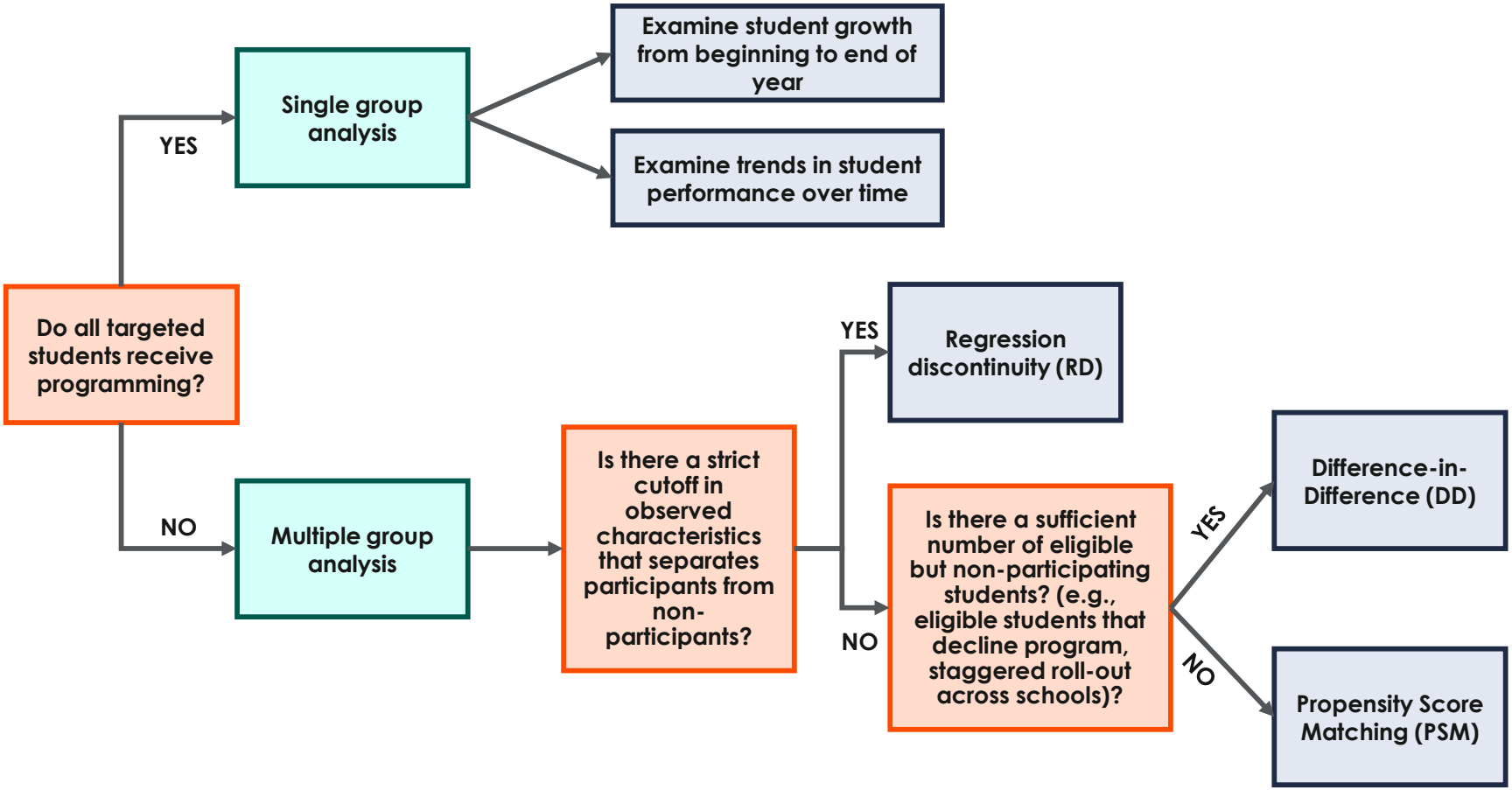
PSM compares participants and non-participants based on similar characteristics. PSM analyzes outcomes between each group and compares differences as a result of program participation.



Source: "Propensity Scores," Better Evaluation.

QUANTITATIVE METHOD DECISION TREE

Use the ***Quantitative Method Decision Tree*** to select which method of determining program efficacy best aligns with the structure of the program.



Source: Developed by Hanover Research.

REGRESSION DISCONTINUITY PROFILES

The profiles below briefly showcase regression continuity analyses in practice, one of which reflects best practices while the other provides a cautionary tale.

Hanover Profiles

Hanover Success Profile:
In an evaluation of math and literacy interventions, a district determined program eligibility based on the previous year's performance. Students participated in the program until they reached grade level standards.

The district identified cutoff scores by grade and verified that participation aligned with the cutoffs.

Hanover Word of Caution Profile:
In another instance, a district's program screener did not meet expectations; the screener data was not fully available for participants and non-participants. When available, screener data included a mixture of previous performance by program status. In some cases, program participation did not align with the cutoff.

DIFFERENCE-IN-DIFFERENCE PROFILES

The profiles below briefly showcase difference-in-difference analyses in practice, one of which reflects best practices while the other provides a cautionary tale.

Hanover Profiles

Hanover Success Profile:
In an analysis of a Kindergarten preparatory program, a district invited underperforming students to participate in the program, although not all enrolled. As a result, observable student outcome data included (i) participants versus invited non-participants and (ii) pre-program versus post-program.

Hanover Word of Caution Profile:
In a district analysis examining the impact of accelerated math on middle school student math course grades, collected data was not easily comparable as accelerated and non-accelerated students took different courses. Some students also did not have linked elementary math performance to control for pre-program math performance.

PROPENSITY SCORE MATCHING PROFILES

The profile below briefly showcases a propensity score matching analysis that reflects best practices. Three common challenges in completing PSM analyses are also listed.

Hanover Profiles

Hanover Success Profile:
In an evaluation of a reading intervention, a district aimed to measure pre-program reading ability, which was available for all students. The data also included student demographics and classifications. The evaluation used propensity score matching to conclude an absence of significant differences in pre-program ability and other student characteristics between program students and matched non-program students.

Three Common Issues:
Districts must be aware of three common issues limiting analysis through propensity score matching:

1. Matched non-program students who do not resemble program students.
2. Lack of common or observed pre-program measure of baseline performance (e.g., previous reading ability) between program and non-program students.
3. Limited information on which schools, grade levels, and/or student subgroups are targeted for the program. *This is a broader concern that also affects the ability of PSM to identify a good comparison group.*

USING DESCRIPTIVE ANALYSIS

When more rigorous analyses do not apply…

Descriptive analysis of student growth or performance trends over time can be helpful in situations where:

- all targeted students receive programming;
- the program has not been implemented for long enough to support a more rigorous analysis; and
- program implementation limits the application of more rigorous methods.

In some of these situations, Hanover recommends using **at least three years** of student performance to identify trends over time.

This type of analysis is more convincing when it evaluates student growth, either from beginning to end of year or from year-end performance across multiple years.

Source: Developed by Hanover Research.


ADDITIONAL RESOURCES

PROGRAM EVALUATION CHECKLIST

APEX Learning developed its own ***Program Evaluation Best Practices Checklist*** to determine district adherence to best practices of education programs. While not all items on the checklist may apply to every program, this checklist serves as a general guide to districtwide fidelity.

Leadership and Planning

- ❑ There is uniform direction and oversight provided by a district program coordinator.
- ❑ Each site has an active site coordinator and administrative support.
- ❑ An implementation plan has been documented and communicated.
- ❑ The guidance department is involved in student selection.
- ❑ Students are scheduled in classes by subject area.

Utilization

- ❑ Program types are defined (Credit Recovery, Remediation, Exam Prep, Other)
- ❑ Percentage utilization is greater than 75% of the order.
- ❑ Courses: Appropriate curriculum pathways are in use for each program.

Student Success

- ❑ Courses: Greater than 70% of enrolled students have accessed their course within the past 14 days.
- ❑ Courses: Greater than 70% of enrolled students are on schedule.
- ❑ Courses: Greater than 70% of enrolled students are passing.
- ❑ Courses: The completion rate is above 80%. Actual completion rate _____%.
- ❑ Courses: Greater than 70% of enrolled students have a passing (60%) Quality of Work (QOW).
- ❑ End-of-Course/Exit Exam/AP scores show improvement. From _____% passing to ____% passing.

PROGRAM EVALUATION CHECKLIST

Academic Integrity

- ❑ There is uniform direction and oversight The district/school has an Acceptable Use Policy in place.
- ❑ Staff user accounts are managed to promote security.
- ❑ The district has applied content filters.
- ❑ There is an established limit to assessment attempts.
- ❑ The average amount of time/days to complete is appropriate for the implementation.
- ❑ Courses: Course settings are enabled to support integrity.
- ❑ Course: Study Sheets or student notes are checked before a quiz is reset.
- ❑ Courses: At least 50% of teacher-scored work is included in the course outline.

Classroom Management

- ❑ Student computers are positioned so that the teacher can view most monitors.
- ❑ the teacher has classroom management software to view thumbnails of the screens.
- ❑ Course and classroom materials are managed (paper, printing, required materials, computers).
- ❑ There is an established system for students to request activity resets or tutoring.

Instructional Methods

- ❑ Courses are facilitated by a Highly Qualified Teacher within their area of certification.
- ❑ Courses: A student portfolio is required that includes all written work and notes.
- ❑ Courses: The use of Study Sheets is required.
- ❑ Teachers are available for synchronous guidance, tutoring, and support.
- ❑ Teachers provide small group and individual target instructional opportunities.

PROGRAM EVALUATION CHECKLIST

Motivate Students to Stay on Track

- ❑ Teachers develop positive relationships with students.
- ❑ Teachers help students set goals, plan, and pace their learning.
- ❑ Teachers conduct weekly progress check-ins with students.
- ❑ Tutoring or attendance is required for students below ___% progress.
- ❑ Learning contract conferences are conducted when students are consistently behind.
- ❑ Courses: Due dates are set in the Grade Book.
- ❑ Courses: Motivational reward systems are in place for students who are on track for on-time completion.

Additional Notes:

Communication

- ❑ Teachers regularly interact with students.
- ❑ Teacher feedback is evident on written work.
- ❑ Courses: The coach email report is used to keep parents/guardians informed of progress.
- ❑ Other methods of communication are used as needed (phone, email, video chat, etc.)

Data Management

- ❑ Teachers maintain a communication log or use the message center for Independent Study and Virtual program students.
- ❑ Courses: Teachers enter valid final grades (0-100, p, A-F) in the Grade Book.

WORKS CITED

WORKS CITED

"Difference-in-Difference Estimation." Columbia University Mailman School of Public Health. https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation

Finney, S. and J. Gerstner. "Fidelity Workshop." , James Madison University. http://www.jmu.edu/assessment/_files/Fidelity%20Workshop.pptx

Gerstner, J. and S. Finney. "Measuring the Implementation Fidelity of Student Affairs Programs: A Critical Component of the Outcomes Assessment Cycle." *Research & Practice in Assessment*, 8, 2013. http://www.rpajournal.com/dev/wp-content/uploads/2013/11/SF2.pdf

Giancola, S. "Evaluation Matters." U.S. Department of Education, 2014. https://www2.ed.gov/about/offices/list/oese/sst/evaluationmatters.pdf

Goldstein, N. "The Program Manager's Guide to Evaluation." Administration for Children and Families, 2010. https://www.acf.hhs.gov/sites/default/files/opre/program_managers_guide_to_eval2010.pdf

"Implementation Fidelity." James Madison University. http://www.jmu.edu/assessment/sass/AC-step-four.shtml

Mohajeri-Nelson, N. and T. Negley. "Best Practices in Program Evaluation." Colorado Department of Education, 2016. https://www.cde.state.co.us/fedprograms/programevaluationppt

"Propensity Scores." Better Evaluation. https://www.betterevaluation.org/en/evaluation-options/propensity_scores

"Regression Discontinuity." Better Evaluation. https://www.betterevaluation.org/en/evaluation-options/regressiondiscontinuity

Research, W. "Program Theory and Logic Models." Wilder Research, 2009. http://www.evaluatod.org/assets/resources/evaluation-guides/logicmodel-8-09.pdf

Sattler, C. et al. "Scaling Barriers to Ensure Success in Program Evaluation." Strategic Data Project. http://sdp.cepr.harvard.edu/files/cepr-sdp/files/program_evaluation.pdf?m=1450111809