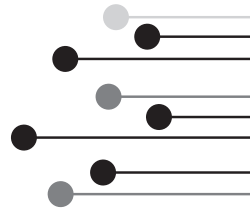# Choose Implementation Monitoring Methods and Compile the Comprehensive Implementation Monitoring Plan

In this step, you and your planning team will review evaluation methods applied to implementation monitoring, select specific methods to address implementation monitoring questions, and organize the methods into a plan using a logic model.

Consider options for implementation monitoring design and methods.

This step consists of reviewing and considering options for implementation monitoring design and methods, including data sources, sampling, design, data collection tools, data collection procedures, criteria for evidence of implementation, triangulation, data management, and data analysis/synthesis. A comprehensive plan ideally will use both *qualitative* and *quantitative methods* and multiple data sources, within the confines of available implementation monitoring resources. Internal Review Board (human subject) issues must also be considered and addressed prior to any data are collected; these will be discussed more fully in Step 10 when the data collection plan is carried out.

Qualitative and Quantitative Methods

Qualitative methods involve an inductive approach to gathering information about the *how* and *why* of human behavior through observation, interviews, focus groups, storytelling, and open-ended interview

Learning Objectives

By the end of the chapter you will

1. Consider options for implementation monitoring design and methods

2. Select multiple implementation monitoring methods for each implementation monitoring question

3. Compile the comprehensive implementation monitoring plan

4. Organize the implementation monitoring plan using a logic model

questions, in contrast to surveys with close-ended questions. Conversely, **quantitative methods** entail collecting data that are in numerical form or can be changed to numerical form for mathematical/statistical analysis. McDavid, Huse, and Hawthorn (2013) compared qualitative and quantitative approaches to evaluation, summarized in Table 8.1. The quantitative column is heavily oriented toward outcome/impact evaluation, which does not apply directly to implementation monitoring. Many of the principles expressed, however, do apply to the quantitative approaches recommended in this textbook. The quantitative focus of this textbook should be evident by now with its emphasis on conceptual and logic models, implementation monitoring questions, and quantitative data collection tools. This is in large part an attempt to create balance with the long tradition of using largely qualitative approaches in process evaluation including program, policy, or practice implementation monitoring. It is essential, however, to use both qualitative and quantitative methods.

The quantitative methods provide numerical data to specify level of implementation based on predefined program, policy, or practice elements and can be used in outcome analysis to adjust for level of implementation. Qualitative approaches enable the planning team to fully understand the setting, stakeholders' perceptions of the innovation, adaptations to the innovation, and both positive and negative unintended effects of the implementation process. Quantitative methods are well suited to capture expected elements, whereas qualitative methods are very useful for unexpected elements including some contextual factors; both are needed in a comprehensive approach.

Common qualitative data collection methods include, but are not limited to, open-ended questions in interviews, focus groups, direct observation, and content analysis of video. Common quantitative methods include, but are not limited to, surveys, checklists, attendance logs, self-administered forms, and project archives (Steckler & Linnan, 2002). Baranowski and Stables (2000) presented both qualitative and quantitative aspects of data collection for each of the components of implementation monitoring. Qualitative aspects are largely descriptive and document types of approaches including messages and incentives used to recruit and maintain participants, contextual factors, quality and depth of program, policy, or practice delivery, barriers experienced, changes or adjustments made to the program, policy, or practice during implementation, and participant reactions and preferences, whereas quantitative elements are numerical, including counts and levels. The approach taken in this textbook builds on these quantitative approaches.

Table 8.1 Comparing Qualitative and Quantitative Approaches to Evaluation

	Qualitative Work Characteristics	Quantitative Work Characteristics
Overall approach	Inductive approach to data gathering, interpretation, and reporting	Hypotheses and questions, which may be embedded in logic models, are tested.
Perspective	Holistic approach that looks to understand the context and implementation process and to interpret results	Finding patterns that either corroborate or disconfirm hypotheses and/or answer evaluation questions
Understanding	The subjective lived experiences of stakeholders (their truths)	How social reality as assessed by the evaluator corroborates or disconfirms hypotheses and answers evaluation questions
Data	Natural language throughout the process	Measurement procedures that lend themselves to numerical representations of variables
Data collection	In-depth, detailed, and focused	Representative samples
Sample size	Purposive sampling, small samples to examine a specific phenomenon in detail	Larger sample sizes, to gather evidence for overall implementation
Data collection tools	Evaluator as primary measuring instrument, qualitative interview, and focus group guides	Measuring instruments are quantitative and constructed to be valid and reliable
Approach to setting	Naturalistic, does not explicitly manipulate the setting	Evaluator control to improve objectivity

Source: Adapted from McDavid, Huse, and Hawthorn (2013, p. 201).

Methods and Design
Elements for Implementation Monitoring

Methods and design elements include data sources, sampling, design, data collection tools, criteria for evidence of implementation, data collection procedures, data management, and data analysis/synthesis. These are reviewed and summarized in Table 8.2, which presents qualitative and quantitative examples.

Data Sources

Data sources refer to *from where* or *from whom* information will be obtained; the selection of data sources may be related to, but is not the same as, sampling.

Table 8.2 Examples of Qualitative and Quantitative Implementation Monitoring Method Components

Methodological Component	General Definition	Quantitative Examples	Qualitative Examples
Data sources	Source of information (e.g., who and/or what will be surveyed, observed, interviewed, etc.).	Possible data sources include participants, teachers, or other staff delivering the program, policy, or practice, records, the environment, written policies, etc.	
Sampling	How participants, settings, and/or activities will be chosen, as well as how many will be chosen.	Quantitative sampling is optimally designed to be representative (ideally, random sampling).	Qualitative sampling is generally purposive (select specific cases for an in-depth view).
Design	Timing of data collection: when and how often data are to be collected and from what group(s) (e.g., intervention, control, or both) data are collected.	Observing intervention and control classroom activities at least twice per semester with at least 2 weeks between observations.	Conducting focus groups with participants in the last month of the program.
Data collection tools or measures	Instruments, tools, and guides used for gathering data.	Surveys, checklists, observation forms, interview guides, etc.	
Data collection procedures	Protocols for how the data collection tool will be administered.	Detailed description of how to do environmental observation, record reviews, face-to-face or phone interviews, mailed surveys, focus groups, etc.	
Criteria for evidence of implementation	Values on rating scale, percentages, or indices that indicate acceptable level of implementation.	Applies primarily to quantitative indicators; a rating of 3 or higher on a 4-point scale; 80% of participants with "agree" or "strongly agree" responses; index score of at least 8 out of 10.	In some cases, presence of theme may serve as qualitative evidence.
Data management	Procedures for collecting and entering data from field; quality checks on raw forms and data entry.	Staff turn in participant sheets weekly; implementation monitoring coordinator collects and checks surveys and gives them to data entry staff.	Interviewers transcribe information and turn in tapes and complete transcripts at the end of the month.
Data analysis/ synthesis	Statistical and/or qualitative methods used to analyze and/or summarize data.	Statistical analysis and software that will be used (e.g., frequencies and chi squares in SAS, SPSS).	Type of qualitative analysis and/or software that will be used (e.g., NUD*IST, InVIVO).

Source: Adapted from Saunders, Evans, and Joshi (2005).

Data sources include individuals who are reflecting their own perspectives and reactions and organizations reporting on their environmental policies or practices. Data sources may also include observation of environments and/or activities as well as reviews of organizational records. A new program, policy, or practice is experienced by many stakeholders and can be viewed from multiple perspectives; therefore, it is often recommended that multiple data sources be used to examine important elements (Bouffard, Taxman, & Silverman, 2003; Helitzer, Yoon, Wallerstein, & Dow y Garcia-Velarde, 2000; Resnicow et al., 1998). For example, ENRICH focused on enhancing the physical activity and nutrition environment in children's residential homes and sought to understand the perspectives of multiple stakeholders including children, implementing staff, direct care staff, and organizational administrators (Saunders et al., 2013).

Identifying the people, locations, and/or records to interview, observe, and/or review is largely a project-specific activity, but King and colleagues (King, 1987) provided several pointers. These include focusing on key people or sources who have the information in which you are interested, such as implementers, participants, and others who have active roles, and asking stakeholders to nominate individuals and other sources who are likely to have the information that is needed.

Identifying data sources that reflect organizational level perspectives can be challenging. There are several common approaches used to assess organizational policy, including reviewing written documents, interviewing or surveying organizational informants, and interviewing or surveying many people within the organization. Each approach has strengths and weaknesses, as well as its appropriate uses. For example, reviewing written documents is an effective way to assess formal policies (McGraw et al., 2000), but it may not capture informal practices and the extent to which policies are enforced, which may require interviewing. Interviewing or surveying an organizational informant as a representative of the organizational perspective is a common practice; however, it is important to select an informant who has the appropriate policy and/or practice perspective. Finally, interviewing or surveying many respondents within an organization may be appropriate for some organizational measures, such as climate or culture; however, for understanding policy, this approach may yield a diversity of opinions rather than a unified perspective. This diversity of opinion is informative in its own right, but it may be difficult to create a coherent variable from these data.

Obtaining information from a variety of data sources will likely require a great deal of cooperation from the stakeholders (King, 1987). It is essential to maintain positive and effective working relationships with all stakeholders by

being mindful of potential respondent burden when collecting implementation monitoring data. To ignore the potential burden is to jeopardize relationships, quality of the collaboration, quality of the data, and the program, policy, or practice itself.

Sampling

Sampling refers to *how* participants—including individuals and organizations, settings such as classrooms and other environments, and/or activities such as specific sessions or events that provide information about the implementation process—will be selected, as well as *how many* will be selected. For many change efforts that target change in programs, policies, or practices, sampling will need to be done at multiple levels that may include coalitions, organizations, groups, specific settings, and/or individuals. If there are multiple organizations with multiple settings and multiple individuals within each setting, sampling becomes somewhat more complex as the planning team will need to determine how to sample as well as how many participants to sample at each level. For example, what are the strategies for sampling if there are 48 recreation centers, each with multiple outdoor playgrounds and play areas, all operated by multiple personnel? How many of the organizations should be sampled? How does one select the specific observation areas at each site? How should staff be selected for interviews? These decisions are ideally driven by data sources needed to address the implementation monitoring question and are often constrained by available resources.

The manner in which the sample is selected should enable the planning team to draw meaningful conclusions about the question being examined. For most quantitative applications, the planning team is not interested in a single perspective about a program per se, but rather, a reflection of intervention participants in general. For example, a single individual's satisfaction or dissatisfaction with training may be less informative than the level of satisfaction of all or most participants who attended training. Similarly, from an intervention perspective, one policy environment is potentially an interesting case study, but the planning team is often interested in patterns or results involving many policy environments.

As with outcome evaluation, the evaluator using quantitative methods should avoid sampling in ways that create bias, particularly systematic bias. For example, if training takes place in multiple sessions over time and attendance drops over time, assessing only those present at the final training session may inadvertently select for those favorably predisposed because those who were dissatisfied may not be present. Ideally, sampling should be structured in a

manner to reflect the full population or group. If it is not possible to sample every individual and population or group, which is frequently the case, the optimal approach is random sampling, as this increases one's ability to generalize to the population, though in practice random sampling may be challenging.

In some cases, it may also make sense to take a stratified sample based on features of sites at the organizational level or demographic characteristics at the individual level that could affect how the program is implemented (King, 1987). For example, if organization size is an important influence on implementation and/or outcomes, stratifying by and sampling from both large and small organizations is a reasonable strategy. Having a full understanding of organizational and community factors will facilitate decisions along these lines. Similarly, at the individual level, if gender is known to have an influence on outcomes, then sampling should ensure perspectives of both genders for implementation monitoring. Having a full understanding of the program, policy, or practice focus and population of interest is essential for sampling at the individual level.

For example, level and type of physical activity, as well as influences on physical activity, vary by sex at nearly all ages. This means that boys and girls participating in the same afterschool physical activity intervention may have very different experiences. Fully understanding population reach and how the afterschool program was received will require sampling males and females.

To address qualitative evaluation questions, purposive sampling may be appropriate. For example, if the planner or evaluator wishes to understand nonparticipation from the perspective of nonparticipants in an initiative, sampling should draw from organizations and/or individuals who have not participated and who are willing to share their perspectives. Often qualitative approaches may call for understanding the perspectives of a limited number of participants in far greater depth; these are not intended to be generalizable, but rather to paint a very detailed picture that is generally unobtainable with quantitative approaches.

Design

Design refers to *when* and *from what groups* data are to be collected. From what groups data are collected in implementation monitoring refers to the intervention and control or comparison conditions; in contrast, sampling refers to how specific units of interest are selected into either condition, as described above. It is very common in implementation monitoring to collect data only from the intervention or program condition, and this may be appropriate in many cases;

however, if resources allow, it is advisable to collect the same or analogous information in control or comparison conditions as well. This will enable the planning team to examine the role of specific organizational and broader community factors as well as secular trends on implementation and outcomes in both conditions. Collecting implementation monitoring data in the control condition necessitates using language that does not require awareness of the program, policy, or practice. For example, rather than asking about the Lifestyle Education for Activity (LEAP) team, which is specific to the innovation, the planning team would ask about a "committee that plans or coordinates activities related to physical activity."

Monitoring contextual factors in intervention and control conditions is important because contextual and external influences can have positive or negative effects on implementation processes and study outcomes in both groups. Outside influences on the intervention group could be confounded with intervention effects. An example of a positive influence would be increased federal funding at the state level or new federal regulations that promote policies or practices similar to those promoted in the innovation. Examples of negative influences include staff turnover, construction, or disasters. Similarly, contextual and external influences can affect the control condition positively, mimicking the program, policy, or practice and reducing the distinction between the intervention and control conditions, or negatively. If these influences are assessed in both conditions, it is possible to document and describe them objectively in real time and to control for them analytically.

Concerning when data should be collected, baseline or preimplementation data ideally should be collected in both the intervention and control or comparison conditions within the same time frame. Depending on the implementation monitoring question and the implementation process, it may make sense to collect data periodically during the implementation process or at a single point later in the intervention timeline. The exact timing of the data collection depends on the question being answered as well as feasibility issues. If the purpose of data collection is to assess level of implementation, consideration needs to be given to at what points in time implementation can be best reflected. For example, if full implementation of a policy change in a nonprofit organization is expected to take 3 months, it would not make sense to collect data after 1 month. Similarly, if implementation is expected to result in organizational or environmental change, the time frame in which this is likely to occur needs to be considered as part of timing of data collection. King and colleagues (King, 1987, p. 51) provided a series of useful questions concerning timing of data collection:

- Do you wish to look at the program periodically in order to monitor whether the program implementation is on schedule?

- Do you intend to collect data from any individual site more than once?

- Do you have reasons to believe that the program will change over the course of the evaluation?

- If so do you want to write a profile of the program throughout history that describes how it evolved or changed?

Data Collection Tools

Data collection tools refer to all instruments, measures, checklists, observational tools, and interview/focus group guides used for gathering implementation monitoring data. Quantitative data collection tools for implementation monitoring differ from other quantitative measures only in their application to implementation monitoring; therefore, all measurement considerations such as validity and reliability apply to these scales (McDavid et al., 2013, Chapter 4). It is difficult to find standardized tools in implementation monitoring that have established validity and reliability, in large part because most implementation monitoring instruments are specific to the program, policy, or practice intervention under investigation (McGraw et al., 2000).

King and colleagues (King, 1987) described instrument validity in innovation implementation as a four-part question that addresses the extent to which the description of the program presented by the instrument is accurate, relevant, representative, and complete. An accurate instrument creates a picture of the program that is very close to what one would see on-site. Relevant measures focus on the most critical features of programs, those that are most likely related to the program outcomes. A representative depiction of the program presents a typical feature of the program and variations across sites and over time. And, finally, a complete picture includes all relevant and important program features. Table 8.3, adapted from King (1987) compares four common methods for collecting implementation data: examining records, conducting observations, using self-administered questionnaires, and conducting interviews. In a comparison of methods to assess school-based curricula, Resnicow and colleagues (1998) reported that implementer self-report was not related to outcomes, but that multiple observation and interviews were.

Table 8.3 Comparing Four Methods for Collecting Implementation Data

Method	Advantages	Disadvantages	Examples
Examine records	– Can be collected without additional demand on participants – Often viewed as objective and more credible – Sets down events at the time of occurrence that increases credibility	– May be incomplete – May be time-consuming to extract relevant information – May be ethical or legal constraints – May be burdensome if not routinely collected	Systematic accounts of regular occurrences, often part of the organization's record keeping; may also include records (e.g., field notes and other documentation) kept by project staff
Conduct observations	– May be seen as highly credible when collected by trained, objective observer – Observers provide a point of view that is different than those connected to the program.	– May alter what takes place – Time is needed to develop instruments and train observers. – Conducting observations is time-consuming. – May encounter scheduling problems	One or more trained observers use a checklist or other instrument to observe events, activities, and/or the environment.
Use self-report measures: questionnaires	– May address a variety of questions – May be answered anonymously – May allow respondent time to think before responding – May be given to many people at distant sites and simultaneously – May impose uniformity on information obtained	– Are not flexible – May limit people's ability to express themselves and capture unique circumstances – Getting surveys returned may be difficult.	A written tool administered in person, electronically, or via mail to which participants or staff respond, often using a rating scale (quantitative); open-ended responses are options (qualitative).
Use self-report measures: interviews	– May be used with a variety of people who have difficulty with written questions – Permits flexibility and ability to pursue anticipated responses	– Time-consuming – Interviewer may inadvertently influence responses.	Participant responds to questions in person or over the telephone; it may be open-ended (qualitative) or close-ended (quantitative).

Source: Content adapted from King, Morris, and Fitz-Gibbon (1987).

Ideally, the quantitative data collection tools used to monitor implementation are conceptually based, reflecting the conceptual framework of the program, policy, or practice—that is, reflecting complete and acceptable delivery/installation. Therefore, at a minimum, it should be possible to establish content and/or face validity and, with sufficient resources, to establish concurrent, predictive, and construct validity. Similarly, reliability for checklists, surveys, observational tools, record review instruments, and other measures may be established through test-retest methods or interrater reliability (DeVellis, 2012; King, 1987). All data collection tools should be pilot tested prior to use in data collection.

Qualitative methods have alternate criteria for judging quality and credibility. McDavid and colleagues (McDavid et al., 2013) compared traditional positivist/postpositivist quantitative and interpretivist/constructivist and critical change qualitative approaches. The qualitative criteria clearly reflect subjectivity openly as well as the necessity of social, political, moral, and/or historical values in context. The emphasis is on trustworthiness, authenticity, and representation of multiple perspectives in contrast to measurement validity and reliability. See Chapter 5 in McDavid et al. (2013) for more in-depth coverage of the use of qualitative methods in program evaluation.

Data Collection Procedures

Data collection procedures used in quantitative approaches refer to the detailed protocols for the administration of data collection tools. It is important that all procedures are clearly documented and that all data collectors are trained to administer the tools systematically. In large projects with ongoing data collection, data collectors may need periodic "recertification" to ensure systematic and high-quality data collection. Data collection procedures should be pilot tested in similar conditions prior to use in data collection. Data collection features for qualitative approaches should be consistent with the theoretical perspectives of the qualitative approach, as discussed in McDavid et al. (2013).

Criteria for Evidence of Implementation

Criteria for evidence of implementation refers to standards that are set for complete and acceptable delivery and that may include values on rating scales, percentages, or indices that indicate desirable levels of implementation. For quantitative data, it is optimal to set standards for desirable levels that provide evidence of implementation prior to data collection. What constitutes an acceptable level will vary by project and setting; this should be discussed and agreed on by the planning team and stakeholders. Establishing criteria for acceptable levels of implementation prior to data collection may prevent

problems such as collecting data that are not applicable or are difficult to interpret and therefore have little meaning or use.

At the most basic level, the planning team can set criteria for acceptable levels of implementation on individual quantitative data collection tools by determining acceptable score(s) on the rating scale. This should be informed by the meaning of the response options on the form. For example, in LEAP, there was a 4-point response scale that examined records for evidence of implementation of essential elements: 0 = not found in records; 1 = some activity documented; 2 = organized activity documented; and 3 = organized activity highly consistent with LEAP philosophy and theory. The planning team, working with stakeholders, set acceptable levels at 2 or higher, that is, any organized activity. Criteria for evidence of implementation may be revisited with caution in the analysis phase. It is not acceptable to manipulate criteria for evidence of implementation after-the-fact to create a more or less favorable report of implementation. However, if the collected data have a restricted or skewed range so that reported ratings are 2s and 3s with no 0s and 1s, the planning team may wish to revisit the definition of the evidence level.

If there are multiple data sources with corresponding data collection tools examining the same program, policy, or practice element, the planning team will next examine or **triangulate** multiple data sources/tools that provide multiple perspectives on one program element. **Data triangulation** refers to using two or more data sources to examine evidence of implementation. Similarly, if the program, policy, or practice has multiple elements within a component, criteria for evidence of implementation will need to be established at this level as well. In other words, the acceptable level of overall implementation when multiple data elements are triangulated must be determined to examine implementation of a program component (see Figure 8.1). Figure 8.1 illustrates how multiple tools can be used to assess each element of the program, policy, or practice, and then how multiple elements can contribute to understanding overall innovation implementation. Criteria for evidence of implementation are needed at each of these levels.

The approach described above enables the planning team to determine implementation for program, policy, or practice components using multiple data sources and tools to assess evidence for complete and acceptable delivery/installation of program, policy, or practice core elements. The purpose here is to illustrate the importance of thinking about criteria for evidence of implementation in a manner that reflects the complexity of the data.

Evidence of implementation can be determined for each component of a program (e.g., most to least implemented elements) and for each organization

or organizational unit (e.g., school- or classroom-level implementation). For example, in LEAP, we reported the most (Emphasizes lifelong physical) to least (Family involvement) implemented innovation elements, and classified each school into a "higher" and "lower" implementing category (Saunders, Ward, Felton, Dowda, & Pate, 2006).

Figure 8.1   Levels of Evidence for Implementation
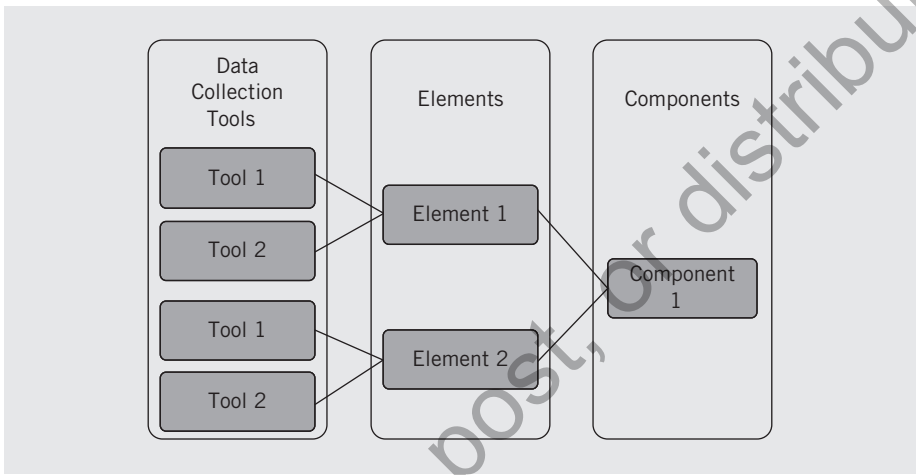
Data Management

Data management refers to the process of getting raw data, collected in the field, through the data entry process and into summarized form. It is important to plan carefully for this aspect of methodology, as poor data management can create tremendous amounts of unnecessary work. In the worst cases, poor data management can compromise data quality, rendering data useless. It is important in the planning stages, therefore, to determine preliminary procedures to ensure the needed resources are in place when data collection begins. These largely pertain to having sufficient time and qualified personnel. Data management will be discussed in more depth in Step 9.

Data Analysis/Synthesis

Quantitative **data analysis and synthesis** will be discussed in depth in Step 11. For planning purposes, it is important to determine preliminary approaches to conducting data analysis and synthesis. The specific analysis

or synthesis approach depends on the implementation monitoring questions, but often begins with descriptive data and basic statistical operations such as calculating means of multi-item scales, summing index scores, and triangulating multiple data sources (illustrated in Tables 8.5 and 8.6). Attention to this step prior to data collection can prevent collection of data that are difficult to summarize and can prompt planning about ways to meaningfully combine and synthesize large volumes of information. If large amounts of data accumulate prior to addressing this, the planning team will likely be overwhelmed.

After data are analyzed or synthesized, they should be put into a "digestible" form for stakeholders. Reporting and using data entails a description of how the information will be summarized, to whom it will be distributed, and for what specific purposes it will be used. Although reporting data to project staff and stakeholders and using the information are not methods per se, it is essential to think about how the data will be used as part of the planning process for methods. Thinking carefully about how this step can prevent collection of unnecessary information that will never be used.

Select multiple implementation monitoring methods for each implementation monitoring question.

The recommended elements of the implementation monitoring plan for each component of the program, policy, or practice include at a minimum fidelity, completeness, and reach, and may also include dose received, recruitment documentation, and contextual factor documentation. Each program, policy, or practice component may have different implementation monitoring plan elements and different methods; therefore, each must be addressed in the implementation monitoring plan. For example, in a school-based program, one component may target students, and another, the students' parents/guardians. The elements that constitute fidelity, completeness, reach, and context, as well as approaches to recruitment, will likely differ between these two components of the innovation.

Planning begins with an implementation monitoring question and consideration of complete and acceptable delivery/installation. The final implementation monitoring plan is the culmination of an iterative process in which the planning team considers implementation monitoring resources, program characteristics, and setting characteristics as implementation monitoring questions and methods are refined and prioritized (see Figure 8.2).

This section will highlight developing data collection tools and establishing criteria for evidence of implementation, as the planning team often finds these elements of the comprehensive plan challenging.

Figure 8.2 Illustration of the Iterative Process of Planning Implementation Monitoring Methods
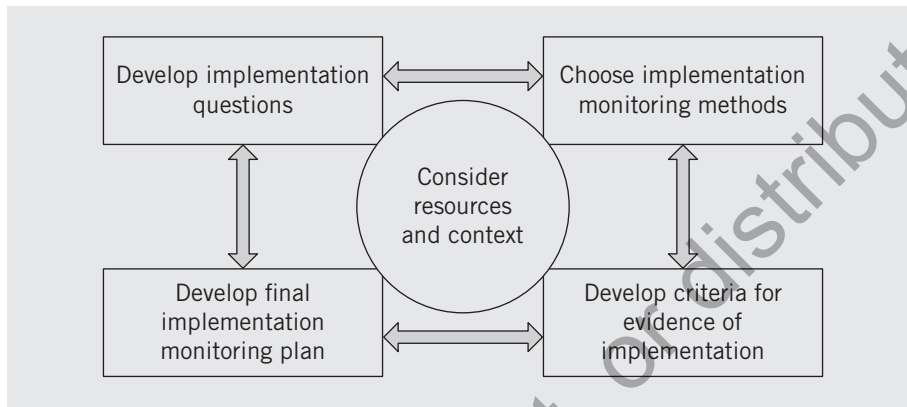
Data Collection Tools

The identification or development of data collection instruments should be guided by complete and acceptable delivery/installation. In LEAP, the essential elements that characterized LEAP PE and the healthy school environment were assessed using multiple quantitative data collection tools including rating scales, checklist observation of the environment and classroom activities, and review of written records and documentation using a rating scale (Saunders et al., 2006). In essence, the LEAP essential elements, which reflected complete and acceptable installation of the LEAP intervention, served as a framework for instrument development. For example, one of the essential elements for instructional practice was gender separation in physical education (PE). Accordingly, items appropriate to staff rating scales, observational checklists, and record review rating scales were developed based on this item (see Table 8.4). Multiple data collection methods and sources are recommended due to the complexity of settings and the genuinely varying perspectives of different stakeholders (Bouffard et al., 2003; Helitzer et al., 2000; Resnicow et al., 1998). Prior to use, all instruments should be pilot tested and all data collectors trained in their use.

Worksheet 8.1 provides a template for choosing and summarizing data sources and tools, including rating scales for implementation monitoring, based on the conceptual definition of complete and acceptable delivery/installation.

Worksheet 8.1 Data Sources, Sampling, and Tools Based on the Definitions of Complete and Acceptable Delivery/Installation of the Innovation

Element and Definition of Complete and Acceptable Delivery/Installation	Data Sources/ Sampling	Tools and Rating Scale	Sample Items
Component A—Fidelity			
Component A—Completeness			
Component B—Fidelity			
Component B—Completeness			

LEAP Case Illustration

Table 8.4 summarizes data sources and tools used in the LEAP project; all tools were based on the LEAP essential elements (Saunders et al., 2006).

Criteria for Evidence of Implementation

Criteria for what constitutes evidence of implementation are established through a series of steps that begin with determining the criteria for a single data source and then for multiple data sources used to monitor implementation of an essential or core program, policy, or practice element, defined by complete and acceptable delivery/installation. If an innovation component is made up of multiple elements, then criteria must be set at this level, also. This process is repeated for all components; criteria can also be set for the number of components that define complete and acceptable overall implementation.

LEAP Case Illustration

This sequence of steps will be illustrated by the assessment of long-term implementation or sustainability of instructional practices in LEAP (Saunders et al., 2012). There were seven essential elements comprising complete and acceptable delivery/installation of instructional practice in LEAP. Specifically, the LEAP PE elements were gender separation, fun classes, physically active classes, appropriate teaching methods, teaching behavioral skills, lifelong physical activity emphasis, and noncompetitive physical activity included.

Table 8.4 LEAP Data Sources and Tools Based on the Definitions of Complete and Acceptable Delivery/Installation of the Innovation

Element and Definition of Complete and Acceptable Delivery/Installation	Data Sources	Tools and Rating Scale	Sample Items
LEAP PE: Characterized by – Gender separation – Fun classes – Physically active classes – Appropriate teaching methods – Teaching behavioral skills – Lifelong physical activity emphasized – Noncompetitive physical activity included	Written records maintained by LEAP staff including training activities, training attendance, field notes, school files, and communication	*Record Review* (35-item rating scale) 0 = not found in records 1 = documents indicate some activity 2 = documents indicate organized activity 3 = documents indicated organized activity that is highly consistent with LEAP theory	Rate evidence for: *Instruction* – Lifelong physical activity is emphasized – Teaching behavioral skills *Environment* – School physical activity team – Administrative support for physical activity
Healthy School Environment: Characterized by – School administrator support for physical activity promotion – Active school physical activity team – Physical activity-promoting messages in the school	Observation of physical education (PE) class and school environment	*Observational checklist* (25 items) 0 = no or none 1 = sometimes 2 = most of the time 3 = all of the time	*Instruction* – Students are organized into small, enduring groups *Environment* – Girls are linked to out-of-class physical activity opportunities via school media messages.
	LEAP staff made systematic assessments based on observation and results documented in written records.	*LEAP Criteria* (36-item rating scale) 0 = no 1 = partially 2 = yes, completely	*Instruction* – Are noncompetitive activities included in PE? *Environment* – Does school have an active wellness team?

Source: Adapted from Saunders et al. (2006).

The active LEAP intervention had concluded several years prior to this implementation assessment; therefore, LEAP intervention staff were not available as data sources. Instead, the LEAP planning team used both qualitative and quantitative methods to tap into evidence of possible sustained implementation. Specifically, qualitative methods included interviews of PE teachers and focus groups of ninth-grade girls currently in PE classes. Many, but not all, of the PE teachers were involved with LEAP in prior years. As expected, none of the ninth-grade girls interviewed during the follow-up period had been exposed to the LEAP intervention during its active phase. The observational tool used to observe the ninth-grade PE class was identical to the quantitative tool used for the active intervention.

Single Data Collection Tool

An essential element was considered to be present if it was observed "most" or "all" of the time (i.e., rated 2 or 3 on the observational checklist) or identified in transcripts of focus groups or interviews by two independent coders.

Multiple Data Collection Tools

An instructional essential element was considered to be present in the school if two of the three data sources (observational checklist, focus groups, interviews) identified the element.
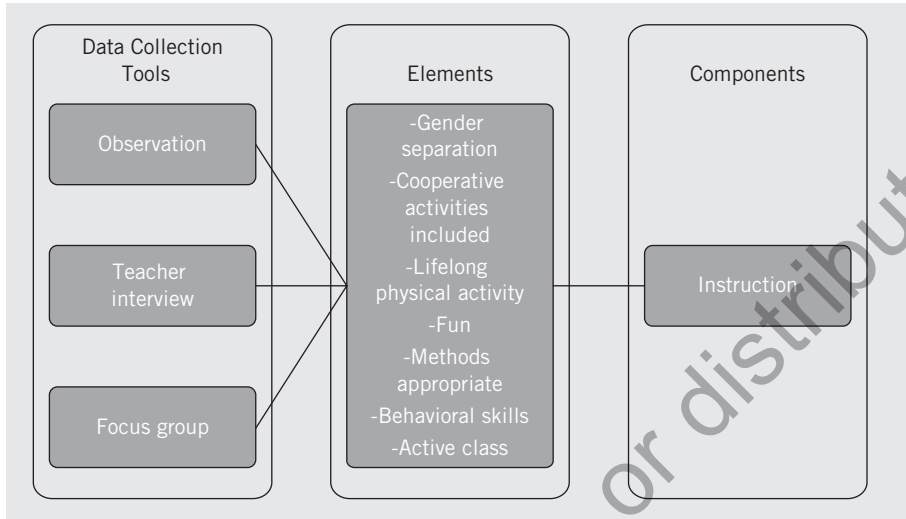
Multiple Essential Elements

LEAP-like instructional practices were considered to be present in a school if a majority (four out of seven) of the instructional essential elements were present.

Figure 8.3 applies this multilayer, multistep process illustrated in Figure 8.1 to LEAP.

Showing the results of this process will illustrate how to define criteria for evidence of implementation. Presenting data in tables and applying the criteria are part of data analysis/synthesis and will be discussed in more depth in Step 11. Table 8.5 presents the data from the three data sources (numbered 1, 2, and 3) by school (lettered A through K) and by essential element (left-hand column). A check is placed in each cell column when data from a data collection tool provides evidence of implementation using the criteria described above. From this table, the patterns of implementation by school and by component become clear. For example, in School A, for the essential element "gender separation in physical education," the PE teacher interview, former LEAP team players interview, and ninth-grade PE observation met the criteria and therefore provided evidence for implementation of this element at follow-up.

Figure 8.3 Levels of Evidence for Implementation for One Component in LEAP

Data Collection Tools
- Observation
- Teacher interview
- Focus group

Elements
- Gender separation
- Cooperative activities included
- Lifelong physical activity
- Fun
- Methods appropriate
- Behavioral skills
- Active class

Components
- Instruction

If the planning team requires only an assessment of overall implementation at the organizational level, without consideration of specific essential elements or components, another strategy is to use information from all data sources to rank order organizations based on scores on quantitative data collection tools. LEAP implementation assessment at the close of the active intervention illustrates this approach (Ward et al., 2006). Multiple data sources and quantitative data collection instruments were used to assess instructional practices as well as the school environment.

Each school received a mean scale score for each of four data collection tools with all items combined, and schools were ranked based on the scores; that is, each school received four rankings, one for each data source. These data were considered ordinal rather than numeric. Therefore, criteria for evidence of implementation was defined as the top two-thirds of schools for a given data source; schools consistently ranked in the top two-thirds were considered "higher implementers," whereas schools consistently ranked in the bottom third were considered "lower implementers." The results of this process illustrate the application of criteria for evidence of implementation (see Table 8.6). Note that the school codes presented in Table 8.6 are not the same as the school codes in Table 8.5, although many of the same schools were involved.

Table 8.5 Using Multiple Data Sources to Establish Evidence of Implementation for Instructional Component of LEAP by School

Essential Element	A			B			C			D			E			F			G			H			I			J			K		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Gender-separate physical education (PE) classes	✓	✓	✓	✓	✓						✓								✓	✓			✓			✓			✓	✓		✓	
Cooperative activities are included	✓	✓			✓	✓					✓							✓	✓	✓	✓		✓		✓				✓			✓	
Lifelong physical activity is emphasized	✓		✓		✓	✓				✓	✓				✓	✓				✓							✓	✓	✓	✓	✓	✓	✓
Classes are fun and enjoyable	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓					✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
Teaching methods are appropriate	✓	✓	✓	✓	✓	✓				✓	✓				✓			✓		✓	✓		✓	✓	✓						✓	✓	✓
Behavioral skills are taught	✓	✓				✓			✓			✓									✓			✓			✓			✓			✓
At least 50% of class is active	✓		✓	✓		✓			✓			✓			✓			✓						✓	✓	✓					✓		

Source: From Saunders et al. (2012).

1 = PE teacher interview; 2 = former LEAP team players interview; 3 = ninth-grade PE observation

✓ = evidence for presence of the indicated element for a given data source

Shaded = strong evidence for implementation

Table 8.6 LEAP Intervention Schools ($n = 12$) Ranked From Highest to Lowest Index Score for Level of Implementation of Essential Elements (Year 2) Using Multiple Data Sources

Rank	Record Review	PE Observations	LEAP Criteria	LEAP Criteria PE
1	G	G	I	C, G, L
2	C	L	G	B, J
3	A, B	F	C, J	A, F
4	F	A	B, L	H
5	D*	B	F	K*
6	J, L	D*	A	I*
7	H*	J	D*	E*
8	K*, E*	C	K*	D*
9	I*	I*	H*	
10		H*	E*	
11		K*		
12		E*		

Source: From Saunders et al. (2006).

Note: Schools ranked in the lower third are shaded. Schools with an asterisk are assigned to low implementation group.

Compile the comprehensive implementation monitoring plan.

The planning team is now ready to compile the final comprehensive implementation monitoring plan, considering the level of resources, program characteristics, and setting characteristics. Resource considerations include the availability of qualified staff to develop and implement all aspects of the implementation monitoring, as well as the time needed for planning; pilot testing instruments and protocols; and collecting, entering, analyzing, and reporting data. It is also important to consider how data collection might be disruptive to program, policy, or practice implementation or the organization's regular operations and might create excessive staff and/or respondent burden. Greater amounts of resources, including time, are needed for large and complicated innovations characterized by multiple components, large numbers

of collaborators, and multiple geographic sites. It is best to be realistic about the amount of data that can be collected and used, given the level of resources available and the context.

In practice, elements of the implementation monitoring plan are developed individually and then summarized into a final, comprehensive plan. The draft plan will include a description of data sources, sampling, tools and procedures, timing of data collection, data synthesis, criteria, and reporting. Worksheet 8.2 provides a template for a comprehensive implementation monitoring plan.

LEAP Case Illustration

Table 8.7 provides a LEAP example of a comprehensive implementation monitoring plan.

Organize the implementation monitoring plan using a logic model.

This is an optimal time to use the logic model to organize the comprehensive evaluation plan. To do this, an additional row that specifies measures identified in the comprehensive implementation monitoring plan is added to the logic model figure from previous chapters. Use Worksheet 8.3 as a template for summarizing the comprehensive evaluation plan using the logic model. Anything that is worth evaluating should be in the logic model, and anything in the logic model should be evaluated.

LEAP Case Illustration

The LEAP logic model with rows for the comprehensive implementation monitoring plan is presented in Table 8.8.

Worksheet 8.2 Template for Draft of Final Implementation Monitoring Plan

	Implementation Monitoring Question	Sampling	Data Sources	Tools and Procedures	Timing of Data Collection	Criteria for Evidence of Implementation	Analysis/ Synthesis	Reporting
Fidelity								
Dose delivered								
Dose received								
Reach								
Recruitment								
Context								

183

Table 8.7 LEAP Summative Implementation Monitoring Plan Example (partial)

Project Definition of Complete and Acceptable Delivery/Installation		Evaluation Question	Sampling and Design	Data Sources	Tools and Procedures	Timing	Criteria	Analysis/ Synthesis	Report
Complete-ness and fidelity	Instructional essential elements*	To what extent were instructional essential elements implemented?	All intervention school PE classrooms	Instruction activities	Independent process evaluator observed classes	At least 2 visits per class per year (fall and spring)	Rated 2 or 3 on 0–3 scale	Mean score	Data from multiple data sources were triangulated in a summative report for investigators, funder, and publication
	Environment essential elements**	To what extent were environmental essential elements implemented?	All intervention school environments	Physical environment	Independent process evaluator observed physical environment	At least 2 visits per school per year (fall and spring)	Rated 2 or 3 on 0–3 scale	Mean score	
	Instructional and environment essential elements	To what extent were all essential elements installed?	All documents on intervention schools	Records	Independent process evaluator rated documentation for essential elements	Once per year	Rated 2 or 3 on 0–3 scale	Mean score	

Project Definition of Complete and Acceptable Delivery/Installation	Evaluation Question	Sampling and Design	Data Sources	Tools and Procedures	Timing	Criteria	Analysis/ Synthesis	Report
Organization change (process outcome) Instructional and environment essential elements	To what extent did instructional practices and the school environment reflect essential elements?	All intervention and control school PE classrooms and environments	Assistant principal	Process evaluator interviewed assistant principal	Once per year (spring)	Not applicable (process outcome, rather than indicator)	Mean score	Examined relationship between implementation and organizational outcome in publication

*LEAP PE—characterized by:

–Gender separation

–Physically active classes

–Teaching behavioral skills emphasized

–Noncompetitive physical activity included

–Fun classes

–Appropriate teaching methods

–Lifelong physical activity

**Healthy School Environment—characterized by:

–School administrator support for physical activity promotion

–Active school physical activity team

–Physical activity–promoting messages in the school

185

Worksheet 8.3 Template for Summarizing and Organizing an Evaluation Plan Using a Logic Model

	Inputs	Outputs			Outcomes/Impacts		
	What We Invest	**What We Do**	**Who Is Reached and Expected Effects**	**Changes Made by Change Agent**	**Short-Term Changes, Impact on Influence Variables**	**Medium-Term Change in Individual Behavior Outcome**	**Long-Term Change on Health Outcomes**
Logic Model	Resources provided enable us to . . .	conduct certain activities that will . . .	reach the change agents and provide them with tools and skills needed to	carry out the program, policy, or practice, which will result in . . .	change in influence variables, which will affect the . . .	individual behavior, which, if sustained, will	have positive effects on health outcomes.
Evaluation Question							
Evaluation Component and Measures							
	ACTION MODEL—Implementation monitoring				CHANGE MODEL—Outcome evaluation		

186

Table 8.8 LEAP Outcomes Logic Model, Illustrating Use of Logic Model to Organize Implementation Monitoring Plan

	Inputs	Outputs			Outcomes/Impacts		
	What We Invest	What We Do	Who Is Reached and Expected Effects	Changes Made by Change Agent	Short-Term Changes, Impact on Influence Variables	Medium-Term Change in Individual Behavior Outcome	Long-Term Change on Health Outcomes
Logic Model	Resources provided enable us to . . .	conduct certain activities that will . . .	reach the change agents and provide them with tools and skills needed to . . .	carry out the program, policy, or practice, which will result in . . .	change in influence variables, which will affect the . . .	individual behavior, which, if sustained, will . . .	have positive effects on health outcomes.
Evaluation Question	To what extent did LEAP staff provide all innovation components, materials, and equipment through training, technical assistance, and ongoing support?		To what extent did LEAP team members attend training? Have the confidence and skills needed to carry out LEAP in their school?	To what extent did LEAP team members install all 11 LEAP instructional and environmental essential elements?			
Evaluation and Measurement	Dose delivered		Reach and dose received	Fidelity and completeness	Influence variables (e.g., self-efficacy)	Individual behavior: physical activity	N/A
	Documentation		Attendance records, training evaluation	Observation, record review, LEAP staff rating	LEAP outcome/impact assessment		
	ACTION MODEL—Implementation monitoring				CHANGE MODEL—Outcome evaluation		

187

Your Turn: Qualitative and Quantitative Methods

Your planning team is debating about whether to use qualitative or quantitative methods for implementation monitoring, and you have been asked to make a recommendation to the team. There are members on your planning team who feel that the philosophical differences between quantitative and qualitative approaches are such that these methodologies are not compatible (see Table 8.1), yet this textbook recommends using both. Make a persuasive argument to your planning team that both methods are needed, providing specific examples to make your case.

KEY POINTS FOR CHOOSING IMPLEMENTATION MONITORING METHODS

- Ideally, both qualitative and quantitative methods are used in implementation monitoring.

- Implementation monitoring methods include considering data sources, design, data collection tools or measures, data collection procedures, criteria for evidence of implementation, data management, and data analysis/synthesis.

- Planning implementation monitoring methods is an iterative process.

- Evidence of implementation will likely need to be established at multiple levels.

- The final implementation monitoring plan must consider available resources, characteristics of the program, and characteristics of the setting.

- The comprehensive implementation monitoring plan can be organized by the logic model.