Behavioral Health Service Division

Legislative Finance Committee

Supplemental Evaluation Guideline Resource for Behavioral Health Service Providers and Evaluators

June 1st , 2025

Table of Contents:

I. Overview ……...…. 3

II. Types of Evaluations ……...… 4

III. Logic Models ……..…... 6

IV. Guiding Principles for Evaluators ……...… 8

V. Performance Measures …….…..…..… 10

VI. References ……..……...……...… 12

Appendix I: Sample Evaluation Form for BHSD Funded Training

Attachment I: BHSD Metrics for Program Inventory

OVERVIEW

This document serves as a supplemental resource to help guide providers and evaluators on the structured set of initial evaluation guidelines for behavioral health services in New Mexico. To that end, this document includes a concise overview of key evaluation concepts, covering various types of evaluation, logic models, core principles that should guide effective evaluation practices especially as it relates to the selection of performance measures, an overview of the types of performance measures (e.g., structure, process, and outcome), and the performance measures selected by the Health Care Authority – Behavioral Health Services Division (BHSD).

The primary purpose of evaluation is to gather information to understand the quality and effectiveness of programs, including the types of services provided and the reach by population, to:

- inform data-driven decisions about resource allocation and program continuation
- identify strengths, gaps and areas of improvement
- ensure that objectives are met
- demonstrate accountability and transparency

It is important to distinguish between evaluating (1) **training** and (2) **services (e.g., clinical programs, interventions)**. The evaluation of **training** is focused on assessing the effectiveness of the training itself, including learning objectives, application of knowledge, and impact of overall performance. Evaluation of training is required for the provision of CEUs. An example of a form used to evaluate training for CEUs is included in **Appendix I**. In contrast, **service evaluation** assesses the quality, impact, and effectiveness of the services/interventions provided by an organization.

For an evaluation to be effective, it must be congruent with the purpose and intended outcome of what is being evaluated (Clark et al, 2019).

Evaluation plans should be designed with sensitivity to the capacities of frontline providers, ensuring that data tracking requirements are feasible and not overly burdensome.

TYPES OF EVALUATION

There are primarily four types of evaluation – formative, process, summative, and outcome - each with a distinct purpose (Rossi, Lipsey, & Freeman, 2004). The four types of evaluation assess effectiveness and impact at different stages.

1. Formative Evaluation

Purpose: To improve the design or performance of a program before full implementation.

When: Conducted during program development or early implementation.

Key Questions:

- Who is the target population?, What is the program vision (intention?), and How are they expected to be served?
- Is the program well-designed to meet its objectives?
- How can it be improved?
- What are the needs of the target population?

---

2. Process Evaluation (also called Implementation Evaluation)

Purpose: To assess how a program is being implemented.

When: During or shortly after implementation.

Key Questions:

- Is the program being delivered as planned?
- Who is being reached?
- What resources are being used?

---

3. Summative Evaluation

Purpose: To measure the short-term and intermediate effects of a program.

When: During program or post-implementation.

Key Questions:

- Did participants experience the desired outcomes?
- Did the program achieve its intended objectives?

---

4. Outcome Evaluation

Purpose: To determine the long-term effects and broader impact of a program.

When: Post implementation.

Key Questions:

- What long-term changes are attributable to the program?
- Did long-term outcomes differ among subgroups (e.g., males/females, Hispanics/non-Hispanics)?

Table 1: Overview of the Types of Evaluation

Evaluation Type	Purpose	When Conducted
Formative	To make early improvements, evaluate the quality, and to ensure that the program is aligned with its intended goals	Before/during development
Process	To explore how a program is being implemented	During implementation
Summative	To demonstrate the effectiveness of a program	During and/or post implementation
Outcome	To understand the longitudinal impact across individuals during a certain timeframe	Post implementation

Economic evaluations are also conducted by public agencies interested in determining the most efficient way to use resources to achieve stated objectives (National Academies of Sciences, Engineering and Medicine (2023). Compared to the various types of evaluations listed in Table 1, economic evaluations focus on costs. **Cost-benefit analysis** (CBA) and **cost-effective analysis** (CEA) are the two most common types of economic evaluations (Drummond et al., 2015, Yates and Marra, 2017).

LOGIC MODELS:

a) **What is a Logic Model /Why use it?**

A logic model is a visual tool that outlines how a program is intended to work, linking its resources (inputs), activities, outputs, and intended outcomes. Through a framework, a logic model helps clarify the relationship between what a program does, the steps a program takes to reach its goals, and the project's goals (Hayes, Parchman and Howard, 2011).

One reason to use a logic model is that it may help with project management, resource allocation and strategic planning. According to the Kellogg Foundation, the development of a logic model is a "conscious process that creates an explicit understanding of the challenges ahead, the resources available, and the timetable in which to hit the target" (W.K. Kellogg Foundation, 2004).

a) **Main Components of a Logic Model:**

i. **Inputs** are the resources required to carry out the program, such as financial resources, human resources, and materials and equipment. Inputs may also include partnerships and collaborations.

ii. **Actors** are the individuals or entities that are responsible for the implementation of programming. (ex., hospitals, courts, behavioral health agencies, etc.)

iii. **Activities** are the actions or processes required to implement the program. These include creating new materials, training staff, conducting outreach, providing services, monitoring progress and evaluating impact.

iv. **Outputs** are the direct, tangible results of the activities. These include (1) the number of people reached (e.g., number of participants in training, clients served); (2) products created (e.g., training materials, patient materials, reports); and (3) services delivered (e.g, number of counseling sessions, number of support groups, number and type of evidence-based practice delivered).

v. **Outcomes** are the expected changes or effects of the program and are typically categorized based on when they are expected to occur and how directly they result from program activities.

i. **Immediate Outcomes** (also called short-term outcomes) are the first changes expected as a result of program activity. Immediate outcomes are direct and measurable effects of program activities. Examples of immediate outcomes include changes in knowledge, attitudes or skills.

ii. **Intermediate Outcomes** occur after immediate outcomes and include for example changes in behavior that result from a change in knowledge, attitudes or skills.

iii. **Long-term Outcomes** (also called ultimate or distal outcomes) take place well after a program has been running. Long-term outcomes are broad and influenced by multiple factors beyond the program and include for example changes in health status and social conditions.

GUIDING PRINCIPLES FOR EVALUATORS

The American Evaluation Association (AEA) established five principles that guide the professional ethical conduct of evaluators (AEA, n.d.). When programs engage external evaluators, those selected should be aware of and adhere to these guiding principles. The guiding principles are fully described on the AEA website (https://www.eval.org/About/Guiding-Principles) and are summarized Table 2.

Table 2: A Summary of the Guiding Principles for Evaluators

Guiding Principle	Description
Systematic Inquiry	Evaluators must conduct data-based inquiries that are thorough, methodical, and contextually relevant.
Competence	Evaluators must provide skilled professional services to stakeholders.
Integrity	Evaluators must behave with honesty and transparency in order to ensure the integrity of the evaluation.
Respect for People	Evaluators must honor the dignity, well-being, and self-worth of individuals and acknowledge the influence of culture within and across groups.
Common Good and Equity	Evaluators must strive to contribute to the common good and advancement of an equitable and just society.

I. Guiding Principles of Evaluation, As It Relates to the Selection of Performance Measures

A committee affiliated with the Institute of Medicine (IOM) developed 20 principles to guide the development of evaluation plans (IOM, 2013). The recommendation from the committee is to *"balance the principles based on "context, end user needs, available resources, and other constraints that may appear"* (pg. 79). We selected 10 of the 20 principles, that were most relevant to the selection of performance measures, and adapted their wording slightly to better support evaluation planning for behavioral health services in New Mexico. The term *"performance measure"* refers to a quantifiable metric used to assess processes, track progress, and evaluate effectiveness. Related terms commonly found in evaluation literature include indicator, metric, and outcome.

Table 3: Guiding Principles of Evaluation, as it Relates to the Selection of Performance Measures*

Guiding Principle	Definition/Explanation
Accuracy	When determining performance measures, select those that are more likely to be free from error or bias. Accuracy is related to both reliability (replicability), which is the consistency of an indicator/measure to yield similar results under varying conditions, and validity, which is the extent to which an indicator/measure directly and without error represents a specific concept, construct, or variable.
Comparability	When determining performance measures, select those that are more likely to be comparable among different agencies, populations, and communities.
Context	When interpreting performance measures, it is important to consider the environment in which an intervention is being implemented.
Coordination and Partnership	When selecting performance measures and interpreting and disseminating the results of performance measurement it is important to ensure that all partner perspectives are involved.
Feasibility	When determining performance measures, select those that agencies are capable of measuring with available resources.
Parsimony	When determining performance measures, and when several performance measures could provide similar information, the most succinct and simplest should be selected.
Relevance	When determining performance measures, select those that provide practical, timely, meaningful information consistent with identified and emergent needs of end users.
Scalability	When determining performance measures, select those that can be expanded to a larger population, or another agency, and still be maintain accuracy and feasibility.
Sustainability	When determining performance measures, select those that are likely to be tracked over an extended period of time after external support and funding is terminated.
Utility and Value	When determining performance measures, select those that are useful, relevant, and responsive to the full range of end users.

**Adapted from IOM (Institute of Medicine). 2013. Evaluating obesity prevention efforts: A plan for measuring progress. Washington, DC: The National Academies Press; https://www.ncbi.nlm.nih.gov/books/NBK202506/table/tab_C_1/?report=objectonly*

PERFORMANCE MEASURES

Performance measures typically evaluate the **structure**, **process**, or **outcomes** of care and are expected to be important, scientifically acceptable, usable, and feasible (Ross and Siu, 2007, Donabedian, 1966, AHRQ, 2015).

> **Structure measures** the potential capacity of an organization to deliver high-quality care and refer to the resources that are needed for service delivery. Structure measures include, for example, the number of providers, facility characteristics, policies and procedures, staff job requirements and staff-to-patient ratios. Structure measures are often related to process and outcome measures. A well-structured system can facilitate better processes and lead to improved outcomes.
>
> **Process measures** include what is done to and for people and refer to internal functioning and adherence to best practices within a program, rather than results experienced by clients. Process measures assess the activities and processes involved in achieving program goals, ensuring that interventions are reaching the intended audience and being implemented as planned. Process measures are believed to influence the results or outcomes of that service and include, for example, the number of individuals served and the number of individuals who completed the program.
>
> **Outcome measures** include those that capture the results of care. They include, for example, service use (e.g., retention, length of stay, 30-day readmission), changes in symptoms and functioning, and client or family perceptions of care. Other outcomes include changes in quality of life and substance use.

Evaluations should also collect data on the target population's characteristics, for example the age range of the population served, the region served, ethnicity/race etc. A clear characterization of the target population enables policymakers and decision-makers to assess the reach of a training or intervention and to determine the effectiveness of services using key performance measures (Lund and Matthews, 2024).

The HCA-BHSD and the Legislative Finance Committee (LFC), New Mexico Legislature have developed a framework for measuring the performance of state funded behavioral health services in New Mexico. The performance measures are summarized in Table 4 and include a mixture of structure and process measures, as well as a target population descriptor measure. **Attachment 1** describes the full array of programs and performance measures used.

Table 4: Overview of Performance Measures of State-Funded Behavioral Health Services in New Mexico Selected by the Health Care Authority

Performance Measure	Type of Measure
Number of Providers	Structure
Average Duration of Program	Structure
Total Capacity of the Program	Structure
Number of Participants Eligible to Complete the Program	Structure
Total Participants Served	Process
Total Participants Completing the Program	Process
Number and Type of Evidence Based Practices Delivered to Participants	Process
Number of Programs Matched to Results First Clearinghouse	Process
Number of other types of programs (e.g., promising practices and research-based programs).	Process
Age of Participants Served	Target Population Descriptor

REFERENCES

AHRQ (2015). Types of Health Care Quality Measures. Content last reviewed July 2015. Agency for Healthcare Research and Quality, Rockville, MD. https://www.ahrq.gov/talkingquality/measures/types.html

American Evaluation Association (n.d.) Guiding Principles for Evaluators. https://www.eval.org/About/Guiding-Principles

Clarke, G. M., Conti, S., Wolters, A. T., & Steventon, A. (2019). Evaluating the impact of healthcare interventions using routine data. *Bmj*, *365*.

Donabedian A. Evaluating quality of medical care. Milbank Q. 1966;44:166–206. [PubMed] [Reference list]

Drummond MF, Sculpher MJ, Claxton K, Stoddard GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. 4th ed. Oxford, UK: Oxford University Press; 2015.

Hayes H, Parchman ML, Howard R. A logic model framework for evaluation and planning in a primary care practice-based research network (PBRN). J Am Board Fam Med. 2011 Sep-Oct;24(5):576-82. doi: 10.3122/jabfm.2011.05.110043. PMID: 21900441; PMCID: PMC3266837.

IOM (Institute of Medicine). 2013. Evaluating obesity prevention efforts: A plan for measuring progress. Washington, DC: The National Academies Press

Lund, J. L., & Matthews, A. A. (2024). Identifying target populations to align with decision-makers' needs. *American Journal of Epidemiology*, *193*(11), 1503-1506.

National Academies of Sciences, Engineering, and Medicine. 2023. *Review of Four CARA Programs and Preparing for Future Evaluations*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26831.

Ross, J. S., & Siu, A. L. (2007). The importance of population-based performance measures. *Health Services Research*, *42*(1 Pt 1), 1.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). Evaluation: A systematic approach. Sage Publications.

W.K. Kellogg Foundation. Using logic models to bring together planning, evaluation and action: logic model development guide. 2004.

Yates BT, Marra M. Introduction: Social return on investment (SROI). Evaluation and Program Planning. 2017;64:95–97