

# CRITERIA OF EVALUATION

Authored by  
**mohammad looti**

November 12, 2025

## RECOMMENDED CITATION

mohammad looti (2025). *CRITERIA OF EVALUATION*. PSYCHOLOGICAL SCALES.  
Retrieved from <https://scales.arabpsychology.com/?p=68353>

## CRITERIA OF EVALUATION

**Primary Disciplinary Field(s):** Program Evaluation, Research Methodology, Social Sciences, Public Policy

### 1. Core Definition

The **criteria of evaluation** constitute the fundamental standards, metrics, or benchmarks utilized within academic, governmental, or organizational studies to specify, measure, or gauge the influence, efficacy, or overall results of a program, intervention, policy, or project. These criteria serve as the explicit rules against which evidence is weighed, determining whether stated objectives have been met and whether the resources expended were justified. They translate the often abstract goals of an endeavor into concrete, measurable indicators, thereby providing the foundation for empirical assessment and accountability. Without clearly defined criteria, evaluation efforts risk becoming subjective, leading to ambiguous findings that lack utility for decision-makers or stakeholders.

In essence, the criteria establish the definition of "success" or "failure" for the program under review. They are typically articulated during the early phases of the evaluation design and are directly linked to the program's theory of change or its underlying logic model. A well-designed evaluation will often utilize multiple criteria simultaneously--such as relevance, efficiency, and impact--to provide a holistic understanding of the program's performance across various dimensions. The selection of these standards is a critical, often negotiated process involving evaluators, program managers, funders, and beneficiaries, ensuring that the resulting assessment addresses the key informational needs of the users.

The utility of these criteria extends beyond simple judgment; they also guide data collection strategies and analytical approaches. If a criterion focuses on **sustainability**, for instance, the evaluation must gather data related to financial viability, institutional capacity building, and local ownership. Conversely, if the focus is primarily on **effectiveness**, the data collection centers on measuring direct outcomes against baseline conditions. Consequently, the clarity, appropriateness, and rigor with which criteria are established fundamentally determine the credibility and validity of the entire evaluation enterprise, making their precise definition a cornerstone of sound methodological practice.

### 2. Historical and Theoretical Context

The formal use of evaluation criteria emerged prominently in the mid-20th century, spurred by the growth of large-scale public sector programs in areas like education, social welfare, and international development. Early evaluation focused primarily on measuring **process** (fidelity of

implementation) and immediate **outcomes** (achievement of short-term goals). However, as programs became more complex and demanded greater public accountability, there was a theoretical need to assess broader impacts and longer-term viability. This shift necessitated the development of standardized criteria that could be applied across diverse sectors and institutional settings.

One of the most significant theoretical advancements came from international development organizations. The development of the standardized criteria by the Organisation for Economic Co-operation and Development's Development Assistance Committee (OECD/DAC) in the 1990s formalized five key criteria: Relevance, Effectiveness, Efficiency, Impact, and Sustainability. These criteria became the globally accepted framework for assessing Official Development Assistance (ODA) and have since been widely adopted or adapted by governments and non-profit organizations worldwide, establishing a common language for discussing evaluation results. This systematization moved evaluation from anecdotal assessment to a rigorous, comparative academic discipline.

Furthermore, theoretical approaches to evaluation, such as Utilization-Focused Evaluation (UFE) championed by Michael Patton, emphasize that criteria must be tailored to the specific needs and decision-making context of the intended users. This perspective highlights that criteria are not static, universal truths but instruments designed to maximize the utility of the evaluation findings. Therefore, the historical evolution reflects a movement from simple compliance checking toward complex, context-sensitive judgment rooted in stakeholder engagement and robust methodological frameworks.

### 3. Functions and Purpose in Evaluation

The primary function of establishing criteria is to provide a rational basis for making judgments about programmatic merit or worth. They transform subjective opinions about a program's success into objective statements supported by empirical evidence. By clearly stipulating what constitutes adequate performance *before* data collection begins, criteria manage expectations, reduce bias, and ensure that the evaluation addresses the most pressing policy questions relevant to the program's design and execution. This prescriptive function is indispensable for maintaining methodological integrity.

Secondly, criteria serve an essential accountability function. Public and private funders demand demonstrable returns on investment, and evaluation criteria provide the mechanism for demonstrating compliance with contractual obligations, legislative mandates, or stated organizational missions. For instance, a criterion related to **cost-effectiveness** ensures that program managers are held accountable not just for achieving results, but for achieving them economically. This aspect is crucial for securing continued funding and maintaining public trust in

the efficacy of social interventions.

Finally, criteria are invaluable tools for program learning and improvement. When an evaluation finds that a program is highly relevant but low in efficiency, these specific criteria pinpoint areas requiring managerial adjustment or redesign. They allow evaluators to diagnose failures and successes precisely, moving beyond a simple pass/fail grade to provide actionable feedback. Thus, criteria facilitate organizational learning by linking observed performance directly back to established benchmarks of acceptable performance.

#### 4. Typologies of Evaluation Criteria

Evaluation criteria can be classified in several ways, depending on their focus and domain of application. The most pervasive typology is the adaptation of the aforementioned OECD/DAC framework, which provides a comprehensive assessment scope for complex programs. These five criteria--Relevance, Effectiveness, Efficiency, Impact, and Sustainability--cover the entire program lifecycle, from initial conceptualization to long-term legacy. **Relevance** assesses whether the program's objectives align with the needs of the beneficiaries and the priorities of the stakeholders. **Effectiveness** measures the degree to which objectives were achieved. **Efficiency** looks at the relationship between inputs (resources) and outputs (results). **Impact** examines the broad, long-term, positive or negative changes resulting from the intervention. Finally, **Sustainability** assesses the likelihood that benefits will continue after the external funding ceases.

Beyond the OECD framework, criteria are often categorized based on the nature of the assessment:

**Process Criteria:** These focus on the fidelity and quality of program implementation (e.g., adherence to protocols, timeliness of service delivery, quality of staffing). They are essential for understanding *how* a program achieved its results, or why it failed to do so.

**Outcome Criteria:** These measure the immediate or intermediate changes experienced by participants (e.g., increased knowledge, behavioral modification, improved access to services).

**Macro-level Criteria:** These address overarching policy goals or organizational capacity (e.g., alignment with national strategic goals, institutional capacity strengthening, promotion of equity).

The strategic selection of these typologies ensures that the evaluation captures both internal program operations and external societal effects, providing layered insight into performance.

Furthermore, criteria can be distinguished by their source: **Internal Criteria** are derived from the program's own documented goals, mission statements, and logic models. **External Criteria**, conversely, are imposed by external mandates, regulatory bodies, or societal norms (e.g., minimum legal compliance standards, ethical guidelines, or benchmarks set by peer organizations). A rigorous evaluation typically balances both internal aspirations and external

accountability requirements, utilizing criteria that are both appropriate to the program's specific context and comparable to broader industry standards.

## 5. Characteristics of High-Quality Criteria

The value of an evaluation hinges heavily on the quality of its criteria. High-quality criteria must possess several key methodological characteristics, most importantly **reliability**, **validity**, and **utility**. Reliability ensures that if the criteria were applied repeatedly under similar conditions, they would yield consistent results; they must be clearly defined and unambiguous to minimize interpreter variation. If a criterion measuring "community engagement" is vaguely worded, different evaluators might apply it differently, leading to inconsistent findings and undermining the credibility of the judgment.

**Validity** refers to the degree to which the criteria accurately measure what they are intended to measure. Criteria must be conceptually sound and empirically operationalizable. For example, if a program aims to improve long-term economic stability, using only short-term employment rates as the sole criterion for effectiveness may lack construct validity, failing to capture the durable systemic changes desired. Therefore, criteria must be theoretically grounded in the literature pertaining to the intervention being studied.

Finally, **utility** ensures that the criteria chosen are relevant and useful to the intended audience. Criteria should address the questions that decision-makers are actually facing. An evaluation using highly scientifically rigorous criteria that program managers cannot understand or act upon sacrifices utility for academic purity. High-quality criteria are thus pragmatic, striking a balance between methodological rigor and practical applicability, ensuring that the findings inform genuine policy or managerial changes.

## 6. Relationship to Program Theory and Logic Models

Evaluation criteria are inextricably linked to the program's underlying theory of change and its operationalized depiction in a logic model. The **program theory** hypothesizes the causal links between inputs, activities, outputs, outcomes, and ultimately, impact. The criteria selected for evaluation must directly reflect these hypothesized links. If the theory posits that training (activity) leads to improved skills (output), which in turn leads to better job performance (outcome), the evaluation must establish criteria that measure these specific stages.

The logic model provides the structural framework necessary to organize the criteria hierarchically. Criteria related to **efficiency** are typically applied to the input and activity stages (e.g., cost per service unit), while criteria related to **effectiveness** are applied to the outcome and impact stages (e.g., percentage achievement of desired behavioral change). This direct mapping ensures that the evaluation is assessing the mechanisms specified by the program design rather than irrelevant

metrics. A failure to align criteria with the program theory can lead to judging a program based on metrics it was never designed to achieve, resulting in misleading conclusions.

Consequently, establishing criteria is often one of the first steps in refining or validating the program theory itself. If stakeholders cannot agree on measurable criteria for success, it suggests that the underlying theory or objectives are ill-defined or contradictory. The process of criterion selection thus serves as a crucial diagnostic tool, forcing clarity and consensus regarding the ultimate purpose and expected trajectory of the intervention being evaluated.

## 7. Challenges in Developing and Applying Criteria

Developing appropriate and unbiased criteria presents significant methodological and political challenges. Methodologically, the greatest difficulty lies in establishing criteria for measuring subtle or long-term social impacts (e.g., criteria for assessing empowerment, institutional trust, or cultural change), which often lack clear, quantifiable indicators. These complex criteria risk becoming either too abstract to measure reliably or too specific to capture the full scope of the desired change.

Politically, the process is often fraught with potential conflicts of interest. Stakeholders may advocate for criteria that maximize the appearance of success, potentially downplaying or ignoring critical criteria related to unintended negative consequences or poor efficiency. Evaluators must navigate these political pressures to ensure criteria are comprehensive and equitable. Furthermore, criteria must be dynamic enough to accommodate changes in the operating environment or evolving needs of the beneficiaries, requiring ongoing review and potential recalibration throughout the life of a multi-year program.

A common pitfall is the reliance on **unreliable criteria**, which, as the source content suggests, can "completely void any results that had been previously determined upon them." If the standards used to judge performance are inconsistent, poorly defined, or susceptible to manipulation, any data collected against those standards loses its evidentiary weight. This necessitates rigorous methodological testing of indicators and adherence to transparent protocols during the criteria development phase to safeguard the integrity of the evaluation findings.

## 8. Consequences of Flawed Criteria

The adoption of flawed, inappropriate, or unreliable evaluation criteria carries severe negative consequences for policy, accountability, and resource allocation. Firstly, it leads to erroneous conclusions about program effectiveness. A program judged as successful based on easily met, low-bar criteria may continue to receive funding despite failing to achieve meaningful societal impact, representing a significant waste of public resources. Conversely, a highly effective program measured against unattainable or irrelevant criteria may be prematurely terminated.

Secondly, flawed criteria distort incentives for program managers and implementers. If the evaluation criteria emphasize only short-term outputs (e.g., number of workshops held) and ignore long-term outcomes (e.g., sustained behavioral change), managers will naturally prioritize activities that maximize the measured output, potentially sacrificing quality and long-term goals. This phenomenon, often referred to as "teaching to the test," subverts the genuine goals of the intervention.

Ultimately, the failure to use robust, valid criteria undermines the foundational principle of evidence-based decision-making. If the evaluation data cannot be trusted because the standards of judgment were deficient, policymakers lose confidence in the evaluation process itself. This erosion of trust can lead to decisions driven by ideology or anecdote rather than objective analysis, hindering continuous improvement and resource optimization within public service sectors.

### Further Reading

[OECD/DAC Criteria for Evaluating Development Assistance and Promoting Sustainable Development](#)

[Better Evaluation: Evaluation Criteria](#)

[Patton, M. Q. \(2015\). Qualitative Research & Evaluation Methods: Integrating Theory and Practice. SAGE Publications.](#)

[Program Evaluation \(Wikipedia\)](#)