RESEARCH · MARKET AND BUSINESS RESEARCH

Program Evaluation Research as a Distinct Mode of Applied Knowledge Generation

What program evaluation is, how it differs from and overlaps with research, the main types of evaluation from needs assessment to cost-effectiveness, and what the evidence shows about when evaluation findings are used.

ModQu research article · 1,873 words · 18 references (APA 7th edition)

Program evaluation uses research methods to answer a particular kind of question: is a program, policy or service doing what it is meant to do, and is it worth what it costs? Governments evaluate employment schemes and health programs; charities evaluate community projects; businesses evaluate training courses, customer service changes and internal initiatives. The methods are familiar from social research: surveys, interviews, administrative data, experiments and case studies. The purpose, the audience and the kind of conclusion are different. This article explains what makes program evaluation a distinct mode of applied knowledge generation, how it differs from and overlaps with research, the main types of evaluation, and the conditions under which evaluation findings are actually used.

What evaluation is

Scriven (1991) defines evaluation as the process of determining the merit, worth or value of something. The definition contains the feature that most clearly separates evaluation from other forms of inquiry: it ends in a judgement. A descriptive study of a job-training program might report that 60% of participants

found work within six months. An evaluation asks whether that result is good: good compared with what participants would otherwise have achieved, compared with the program′s targets, compared with similar programs, and in relation to its cost.

Rossi et al. (2019) describe program evaluation as the use of social research methods to systematically investigate the effectiveness of social intervention programs, in ways that are adapted to their political and organisational environments and designed to inform social action. Weiss (1998) similarly emphasises that evaluation compares a program against explicit or implicit standards, with the aim of contributing to its improvement. Both definitions combine three elements: systematic methods, a judgement against standards, and an intended use.

Fournier (1995) set out the general logic by which evaluators reach such judgements. They establish criteria of merit (what a good program of this kind should achieve), construct standards (how well it must perform on each criterion), measure performance, and synthesise the results into a judgement. Davidson (2005) built a practical methodology on this logic, arguing that evaluation is not complete until the evidence has been explicitly interpreted against defined criteria of value.

How evaluation differs from research

Whether evaluation is a kind of research, or a separate discipline that uses research methods, has been debated for decades. Mathison (2008) reviews the main positions and concludes that the two overlap heavily in methods but differ in purpose and in the kind of claims they make. Levin-Rozalis (2003) argues that the differences are substantive rather than merely technical: research aims to

produce general knowledge and is guided by theory and the researcher′s questions, whereas evaluation is driven by the needs of a particular program and its stakeholders, and its findings must be useful in that setting.

Wanzer (2021) surveyed evaluators and researchers about how they understood the relationship. Evaluators generally saw the two fields as overlapping, while researchers were more likely to see evaluation as a subcomponent of research. Evaluators perceived larger differences than researchers did, and the differences were greatest at the start of a study, in its purpose, questions and audience, and at the end, in making value judgements and disseminating results.

Research and program evaluation compared

Shared methods, different purposes: typical tendencies rather than fixed rules

Dimension	Research	Program evaluation
Purpose	Add to general knowledge	Judge merit or worth to inform action
Who sets questions	Mainly the researcher	Stakeholders, with the evaluator
Conclusion	Description or explanation	Evaluative judgement, recommendations
Generalisation	A central aim	Secondary; first be accurate locally
Timing	Set by the research process	Set by decisions and budget cycles
Success criteria	Validity, contribution to theory	Utility, feasibility, propriety, accuracy

Based on Levin-Rozalis (2003); Mathison (2008); Wanzer (2021); Yarbrough et al. (2011) ModQu.com

Figure 1. Typical differences between research and program evaluation. The two share methods; they differ mainly in purpose, in who sets the questions and in the kind of conclusion reached. Based on Levin-Rozalis (2003), Mathison (2008) and Wanzer (2021); diagram by ModQu.

Figure 1 summarises the typical differences. They are tendencies rather than rules. Some evaluations produce widely generalisable knowledge, such as large

randomised trials of social programs, and some research is commissioned for a particular client. The distinctions are still useful, because they explain why evaluations are designed, judged and reported differently from academic studies.

The questions come from stakeholders. In research, the investigator usually chooses the question. In evaluation, the questions are set, or at least negotiated, with funders, managers, staff and participants. An evaluator who answers an interesting question that no decision-maker asked has not done the job.

The conclusion is evaluative. Research typically concludes with a description or an explanation. Evaluation concludes with a judgement of merit or worth, and often with recommendations.

The context is particular. An evaluation is usually about one program in one setting at one time. Its first obligation is to be accurate about that program. Generalisation is a bonus, not the main aim.

Timing is set by decisions. Evaluation findings are needed when budgets are set, contracts are renewed or programs are redesigned. A rigorous answer delivered after the decision has been taken has little value. This creates pressure for the rapid methods discussed elsewhere in this series.

Quality includes usefulness. The Program Evaluation Standards (Yarbrough et al., 2011), developed by the Joint Committee on Standards for Educational Evaluation, judge evaluations on five attributes: utility, feasibility, propriety, accuracy and evaluation accountability. Accuracy is only one of the five. An evaluation can be methodologically sound and still fail the standards because it was not useful, not practical, or not conducted fairly and ethically.

Types of evaluation

Scriven (1967) introduced the distinction between **formative** evaluation, which is carried out during development or delivery to help improve a program, and **summative** evaluation, which judges a program′s overall merit, often to inform decisions about whether to continue, expand or end it. The distinction concerns the use of findings more than the methods: the same survey could serve either purpose.

Rossi et al. (2019) set out a sequence of evaluation questions that follows the logic of a program, from whether it is needed to whether it is worth its cost. Figure 2 shows this sequence.

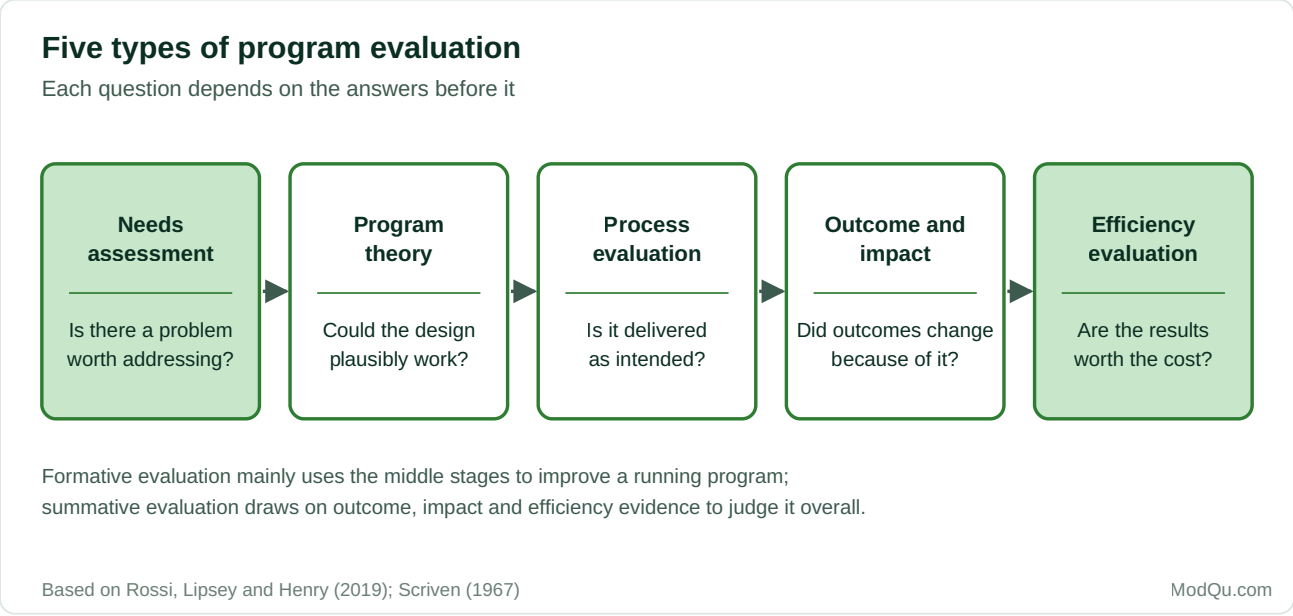


Figure 2. Five types of evaluation and the questions they answer. Each level depends on the ones before it: an impact evaluation of a program that was never properly implemented says little about the program's design. Based on Rossi et al. (2019); diagram by ModQu.

Needs assessment asks whether there is a problem that warrants a program, how large it is and whom it affects.

Assessment of program theory asks whether the program′s design is plausible: whether the activities could reasonably be expected to produce the intended outcomes. Funnell and Rogers (2011) describe how evaluators make a program′s theory of change explicit, usually as a logic model linking inputs, activities, outputs, outcomes and longer-term impacts.

Process or implementation evaluation asks whether the program is being delivered as intended, to the intended people, at the intended quality.

Outcome and impact evaluation asks whether the program achieved its intended outcomes, and whether those outcomes can be attributed to the program rather than to other factors. Attribution is the hardest question in evaluation and is where experimental and quasi-experimental designs are most often used.

Efficiency evaluation asks whether the outcomes justify the costs, using methods such as cost-effectiveness or cost-benefit analysis.

The sequence matters because each level depends on the ones before it. If an impact evaluation finds no effect, a process evaluation can show whether the program failed because its design was wrong or because it was never properly delivered. Weiss (1998) called these theory failure and implementation failure; they lead to very different decisions.

Approaches to evaluation

Evaluation has developed many competing approaches. Shadish et al. (1991) traced the field′s theoretical development through successive stages, from an early emphasis on rigorous methods to a later concern with the use of findings and the political context of programs. Alkin and Christie (2004) organised

evaluation theorists into an "evaluation theory tree" with three main branches: **methods** (emphasising rigorous designs for causal inference), **valuing** (emphasising how judgements of merit and worth are made and whose values count) and **use** (emphasising how evaluation can be designed so that its findings inform decisions).

Stufflebeam and Coryn (2014) review the major models, including Stufflebeam's own CIPP model, which organises evaluation around context, input, process and product. Theory-based approaches, including realist evaluation, focus on how and why programs work. Participatory approaches involve program staff and participants in designing and conducting the evaluation. The choice depends on the questions, the stakeholders and the resources available, rather than on any single approach being correct.

Evaluation use

If evaluation is defined partly by its intended use, then whether findings are used is central to its quality. Early studies found that evaluation results were often ignored, which prompted a large body of research on evaluation use.

Weiss (1979) described several models of how research and evaluation influence policy. In a problem-solving model, findings are used **instrumentally**, to make a specific decision; in an enlightenment model, they are used **conceptually**, gradually changing how decision-makers understand a problem; and in political and tactical models, they are used **symbolically**, to support a position already taken or to delay action. She argued that enlightenment, rather than direct problem-solving, is probably the most common way research enters policy.

Cousins and Leithwood (1986) reviewed 65 empirical studies of evaluation use. They found that use was associated with characteristics of the evaluation, such as

its relevance, quality, credibility and timeliness, and with characteristics of the decision setting, such as the information needs of users, their commitment to the evaluation and the political climate.

Patton (2008) built these findings into **utilization-focused evaluation**, an approach that begins by identifying the specific people who are expected to use the findings and designing every stage of the evaluation around their intended uses. Patton also described **process use**: the learning that happens among stakeholders because they took part in an evaluation, regardless of its final findings.

Evaluation in government and business

Many governments require evaluation of public programs. In Canada, the Treasury Board′s Policy on Results (Treasury Board of Canada Secretariat, 2016) requires federal departments to measure and evaluate the performance of their programs and to use the results to inform decisions about them. Evaluations under such policies typically examine relevance, effectiveness and efficiency.

The same logic applies in business, although the word ″evaluation″ is used less often. A company that introduces a new onboarding process, a sales training course or a customer loyalty scheme faces the same questions: was it needed, was it delivered as planned, did it change outcomes, and was it worth the cost? Businesses often measure only outputs, such as the number of staff trained or the number of loyalty cards issued, and stop there. An evaluation framework pushes the analysis on to outcomes and attribution: whether sales or retention changed, and whether the change would have happened anyway.

Practical steps for commissioning or conducting an evaluation include:

- **Identify the intended users and decisions.** Decide who will act on the findings and what they need to know, and by when.
- **Make the program theory explicit.** Write down how the program is expected to produce its outcomes, so that the evaluation can test each link.
- **Agree criteria and standards in advance.** Decide what would count as success before the data arrive, as Fournier (1995) and Davidson (2005) recommend.
- **Match the design to the question.** Use process methods for delivery questions and comparison or experimental designs for attribution questions.
- **Report a judgement, not just data.** State clearly how well the program performed against the agreed criteria, and how confident that judgement is.

Evaluation as a distinct mode of knowledge

Program evaluation generates knowledge that is local, evaluative and tied to action. It is local because it concerns a particular program in a particular setting. It is evaluative because it ends in a judgement of merit or worth against stated criteria. It is tied to action because it is designed to inform decisions by identifiable users within a time frame those decisions allow. These features do not make evaluation less rigorous than research; they mean that its rigour is judged against a wider set of standards, in which usefulness, feasibility and fairness sit alongside accuracy (Yarbrough et al., 2011).

References

Alkin, M. C., & Christie, C. A. (2004). An evaluation theory tree. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists′ views and influences* (pp. 13–65). Sage.

Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research, 56*(3), 331–364. https://doi.org/10.3102/00346543056003331

Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Sage.

Fournier, D. M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. *New Directions for Evaluation, 1995*(68), 15–32. https://doi.org/10.1002/ev.1017

Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. Jossey-Bass.

Levin-Rozalis, M. (2003). Evaluation and research: Differences and similarities. *Canadian Journal of Program Evaluation, 18*(2), 1–31. https://doi.org/10.3138/cjpe.18.001

Mathison, S. (2008). What is the difference between evaluation and research—and why do we care? In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 183–196). Guilford Press.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Sage.

Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2019). *Evaluation: A systematic approach* (8th ed.). Sage.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Rand McNally.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Sage.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Sage.

Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, and applications* (2nd ed.). Jossey-Bass.

Treasury Board of Canada Secretariat. (2016). *Policy on results*. Government of Canada.

Wanzer, D. L. (2021). What is evaluation? Perspectives of how evaluation differs (or not) from research. *American Journal of Evaluation, 42*(1), 28–46. https://doi.org/10.1177/1098214020920710

Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review, 39*(5), 426–431. https://doi.org/10.2307/3109916

Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Prentice Hall.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Sage.