

The Case for Enhancing the Quality of Primary Evaluations through Metaevaluation in Developing Country Contexts

Peter Wandwasi

<https://orcid.org/0000-0003-0619-5120>

University of Johannesburg

wandwasip@yahoo.com

Abstract

This article makes a case for metaevaluation in the context of further evaluative inquiry beyond primary evaluation level and advances the plausible perspective that metaevaluation improves the quality of primary evaluations and the evidence used to support decision-making. It argues that metaevaluation is a form of evaluative enquiry that is aimed at assessing the strengths and weaknesses of primary evaluations. Through so doing, metaevaluation makes a valuable contribution toward enhancing the credibility and reliability of primary evaluations and the evidence provided to decision-makers. Metaevaluation has been utilised in different contexts in some developed countries to assess the quality of evaluations of development interventions, although scanty evidence exists of its use in developing country contexts.

Keywords: metaevaluation; evaluation; constructivist evaluation; evaluative inquiry

Introduction

Metaevaluation is defined by Scriven (1969) as the evaluation of evaluations. It is an advanced specialised form of the evaluation discipline, and it is used to subject primary evaluations to further evaluative scrutiny in order to judge the strengths, weaknesses and credibility of evaluation evidence. Metaevaluation is driven by the principles of accountability, transparency, participatory social enquiry and epistemology, all of which provide for good programme evaluation. The objective of this article is to summarise the case for metaevaluation and motivate why it is necessary to use metaevaluation to improve the quality of primary evaluations and the evaluation evidence used to support decision-making.

This article is divided into eight sections. The first section specifies the conceptualisation parameters for metaevaluation within the context of evaluative

enquiry. Section two outlines a critical summary of the case for metaevaluation. The third section presents the theoretical basis for metaevaluation. The philosophical basis for metaevaluation methodology and a discussion of guidelines for conducting metaevaluation are, respectively, sketched out in the fourth and fifth sections. The importance of metaevaluation is discussed in section six. Concluding remarks and the list of references are provided in sections seven and eight, respectively.

Conceptualisation of Metaevaluation

Scriven (1969) first coined the term “metaevaluation” as a concept in the late 1960s. The term “evaluation of evaluations,” as defined by Scriven (1969), was created by Orata in 1940 (Hedler and Gibram 2009). It is important to note from the outset that metaevaluation does not mean meta-analysis. Stufflebeam (2001, 187) points out that “although metaevaluation and meta-analysis refer to different concepts, they are often inappropriately equated.” Weiss (1998, 48) defines meta-analysis as the “systematic summary of the results from a number of different evaluations of the same kind of programs.” The systematic summary of the results from a number of different evaluations of the same kind involves collecting and summarising empirical data from similar evaluation studies.

This process uses statistical approaches to combine the results from multiple evaluation studies of the same kind of programmes to increase statistical power in order to resolve improbabilities in circumstances where evaluation evidence from similar multiple evaluation studies differs. Rossi, Freeman, and Lipsey (1999) further explain that meta-analysis tries to determine clusters of findings in order to generate knowledge.

Stufflebeam (2001), in clarifying the key difference between metaevaluation and meta-analysis, describes metaevaluation as an assessment of the merit and worth of a given evaluation, whereas meta-analysis is explained as a form of quantitative synthesis of studies that address a similar research question. The process of metaevaluation can be carried out formatively when a primary evaluation is being conducted and summatively after a primary evaluation has been completed on the effectiveness of an intervention.

Cooksy and Caracelli (2005) describe metaevaluation as a systematic review of evaluations so that the quality of evaluations and the effectiveness of evaluation processes can be determined. However, while Cooksy and Caracelli (2005) describe metaevaluation as a systematic review of evaluations, using the checklist contained in the standards of the Program Evaluation Standards developed by the Joint Committee for Program Evaluation of 1994 and various national evaluation standards, it can be added here too that evaluation systems and the use and effectiveness of evaluation tools can also be subjected to metaevaluation in order to strengthen planning and management of evaluations in organisations.

In addition, Stufflebeam (2000, 95) explains that metaevaluation involves the “process of delineating, obtaining, and applying judgmental information about the utility,

feasibility, propriety and accuracy of an evaluation to report on its strengths and weaknesses.”

Stufflebeam (1974) suggests a range of conceptualisation premises that can be used to conduct metaevaluation activities. Six conceptualisation premises are adopted from Stufflebeam (1974), with modifications, for the purpose of this article. These conceptualisation premises are:

- Metaevaluation should assess the merit and worth of evaluations using appropriate criteria.
- Metaevaluation should assess the objectives, designs and quality of an evaluation.
- Metaevaluation should describe and judge evaluation work and recommend how evaluations can be improved and how the findings can appropriately be used.
- Metaevaluation should serve evaluators and all the persons who are interested in their work.
- Metaevaluation must collect, organise and analyse needed information and apply it to appropriate decision-making and learning processes.
- Metaevaluations must be technically adequate and useful.

Whereas these conceptualisation premises serve as technical hinges around which this article is organised, I submit that the effectiveness of the outcomes of any metaevaluation exercise depends largely on the extent and the effectiveness with which empirical evidence is generated, triangulated and used to provide multiple sources of reliable evidence. This process of triangulating multiple sources of information to construct plausible knowledge cues and conclusions about the overall outcomes of the metaevaluation exercise is largely dependent on the nature of selected evaluation cases as well as the type of methodological approach.

A Critical Summary of the Case for Metaevaluation

Metaevaluation, as defined by Scriven (1969) as the evaluation of evaluations in the preceding section, is a unique, advanced and specialised form of the evaluation discipline that critically assesses the strengths and weaknesses of a primary evaluation, thus enabling interested parties to better judge the reliability and credibility of evaluation as well as the evidence put forward for decision-making (Caracelli and Cooksy 2008; Shinkfield and Stufflebeam 2007; Stufflebeam 2000; Stufflebeam 2001).

There is a variety of definitions of the term “evaluation” reflecting different perspectives of various theorists regarding the purpose and functions of an evaluation (Alkin 1969; Alkin 2012; Borich 1990; Cronbach 2000; Glass 1969; Nevo 1983; Scriven 1969; Stufflebeam 1971; Tyler 1942; Tyler 1950; Wittrock and Wiley 1970). For example, Tyler (1942) propagates the perspective that evaluation is a process of determining the extent to which the objectives of an intervention are realised. Cronbach (2000), Alkin

(1969) and Stufflebeam (1971) assert that evaluation is a process for providing evidence for decision-making.

In addition, the Organisation for Economic Cooperation and Development (OECD) defines evaluation as the process of determining the worth or significance of an activity, policy or programme (OECD 2004). Moreover, evaluation is defined by the World Bank as “the process of assessing the performance of policies, programmes, projects and processes by identifying and disseminating the lessons learnt from experience and by making recommendations drawn from evaluation findings” (World Bank 2010, 26).

The diverse definitions of the purpose of evaluation demonstrate that the field of evaluation is perceived in slightly different ways from different intellectual perspectives. While there are many different theoretical perspectives regarding the purpose of evaluation, the common purpose of these is to assess the extent to which development projects have consistently achieved the intended change for social betterment. However, while there is no universal agreement on any one specific definition itself, it is clear, particularly from the perspective of evaluation of development interventions in developing country contexts, that funders are mainly interested in evaluations that can progressively judge the nature of the impact, efficiency and effectiveness of their funding support.

The diverse definitions of evaluation have also culminated in a variety of models of evaluation, which reflect different perspectives of various theorists regarding the purpose and functions of an evaluation. Owen and Rogers (1999) explain that a model is a prescription for undertaking an evaluation based on certain theoretical assumptions. Stufflebeam and Webster (1983) clarify that the proliferation in the evaluation models is a result of a range of attempts by evaluation theorists to classify evaluation models in terms of elements, such as assumptions, methodology and the engagement of stakeholders.

However, Orr, Cameron, and Day (1991) note that, although there has been a proliferation in evaluation models, a theoretical framework, which would make sense of this situation from the point of view of the evaluation practitioner, was missing. In order to address this missing link, Owen and Rogers (1999) developed a meta-model consisting of five categories of evaluative enquiry, which are designed to address the “why” (why an evaluation is conducted) question in evaluation from the perspective of an evaluation practitioner.

Owen and Rogers (1999, 60) argue that the five forms of evaluative enquiry demonstrate “five dominant styles of evaluation and that each of them produces useful knowledge for decision-making.” The five categories of evaluative enquiry, which form the theoretical grounds for conducting good evaluations include: proactive, clarificative, interactive, monitoring, and impact. All are briefly discussed here.

Owen and Rogers (1999, 60) explain that the “proactive” form of evaluative inquiry “takes place before a programme is designed and that it is intended to provide input to decisions regarding the best way to develop a programme in advance of the planning stage.” The “clarificative” form of evaluative enquiry focuses mainly on explaining the “internal structure and functioning of an intervention, the collection and analysis of data, the intended outcomes and program design, the underlying rationale, and plausibility of the program” (Owen and Rogers 1999, 42–43). These authors further explain that “interactive” evaluation is based on the assumption that those with a direct vested interest in the intervention should also control the evaluation of the intervention. The “monitoring” form of evaluative enquiry is undertaken when an intervention is ongoing and is intended to assess its progress. Finally, “impact” evaluation is used to assess the effects of a completed programme and is intended to inform decision-making on the worth of an intervention.

In addition, Owen and Rogers (1999) specify that, compared to other forms of evaluative enquiry, such as the proactive form of evaluative enquiry, clarificative evaluation involves the collection and analysis of data and seeks to address issues, such as whether or not the intended outcomes can be achieved by the nature of programme design, the underlying rationale for the programme, and its plausibility.

However, depending on the issues on which a metaevaluation exercise may focus, such as programme design, internal structure, objectives, rationale and outcomes, as well as the thrust for which a metaevaluation exercise is intended to resolve, any one of the above forms of evaluative enquiry can be adopted as an appropriate metaevaluation framework.

The Theoretical Basis for Metaevaluation

Metaevaluation is grounded in the theory of good programme evaluation (Alkin 2012; Cronbach et al. 1980; Shadish, Cook, and Leviton 1991; Weiss 1998). The theory for good programme evaluation is based on the vision of improving the well-being of society. The vision of improving the well-being of society through evaluation within the context of development is idealised in the context of solving the problems faced by society.

Shadish et al. (1991), for example, demonstrate that the process of solving the problems of society is guided by a problem-solving sequence. These authors explain that the problem-solving sequence involves the identification of a problem in society, generating and implementing alternatives to reduce the adverse effects caused by the problem, evaluating the effectiveness of these alternatives, and adapting alternatives that evaluation results suggest can significantly reduce the problem in society.

Furthermore, Alkin (2012) demonstrates that the evolution of the theory for good evaluation was driven mainly by the quest to promote accountability and social enquiry. This evolutionary process subsequently precipitated a surge in theoretical debate, which

focused mainly on three major considerations related to evaluation methodology, the manner in which data are to be valued and judged and the uses of the results of the evaluation process.

Shadish et al. (1991), for example, compound the three major considerations of use, methods and valuing into three stage theories. According to these authors, first stage theories emphasise the importance of social programming and the epistemological discovery of knowledge; second stage theories focus on the way evaluation is used, and its social utility; and third stage theories address the integration of enquiry and utility.

In the context of these three major theoretical considerations, a good theory of programme evaluation emerged, which suggests feasible practices that can be used to construct knowledge of the value of social development programmes, which in turn can be used to ameliorate the social problems to which programmes are relevant. The emergent social programming component of the theory assumes that social problem-solving can be improved by incremental improvements in existing programmes, better designing of new programmes, or terminating bad programmes and replacing them with better ones (Shadish et al. 1991).

The knowledge component of a good theory of social programme evaluation addresses the contestation between common sense and systematic observations, and provides answers regarding appropriate methods of constructing knowledge by answering questions, such as: “Why employ professional evaluators? Why not rely on gossip, cronyism, or newspaper reports or lobbyists? Do evaluators offer different or better knowledge than what is already available?” (Alkin 2012; Shadish et al. 1991, 42). In addition, Shadish et al. (1991) and Alkin (2012) advance the evaluative narrative, which propagates the understanding that the process of evaluation is a human investigative endeavour which involves evaluative complexities that are interdependent and often characterised by a nonlinear set of problem-solving activities.

Therefore, the three major considerations of use, methods and valuing demonstrate that humans identify options for improving their own lives by detecting the change, generating pieces of information about the effectiveness of the evaluand (for example, a project, programme, process or organisation), which can be used to reduce societal problems within a given context throughout the evaluation process.

In addition, there are also debates within the context of different forms of human behaviours as to whether or not evaluations should be values-free. While Scriven (1966) suggests that evaluations should be values-free in constructing knowledge, however, Shadish et al. (1991) argue that social programmes are not, in themselves, values-free, and that it is, therefore, impossible to evaluate in the political world of social programming without considering the importance of societal values.

In addition, since evaluation practitioners are action-oriented, they orientate their evaluation tasks with pragmatic concepts and methods, which serve as a useful guide in the execution of the tasks (Shadish et al. 1991). These authors further propose that in order to avoid undertaking evaluations that are incomplete, impractical or technically inferior, the discussion of the good theory of evaluation is important for evaluation practitioners.

In addition, I argue that such human investigative endeavour is grounded in the three major considerations of use, methods and valuing mentioned above, and that the process of evaluation often results in pieces of evidence that inherently reflect a variability of constructions, judgments and envisaged outcomes based on differing forms of behaviour. The reflections of mental constructions and judgements on the effectiveness of the evaluand—resulting from such forms of differing behaviour—follow the constructivist philosophical paradigm discussed below in the section on the philosophical basis for metaevaluation methodology, which advances the argument that humans generate knowledge and meaning based on their own interactions, experiences and ideas. For example, the use component of the good theory of evaluation implies that the evaluation results are used to contribute toward solving societal problems.

Therefore, underlying the use component of a good theory of evaluation, by extrapolation, implies that social interventions that are not effective in contributing to social betterment, based on evaluation results, can be terminated and replaced with better interventions. The practice component of a good theory of evaluation helps human investigative processes to rank the evaluands that have the greatest and most effective opportunity to improve the well-being of society. None of these conclusions are new. They are all logical implications of current knowledge on the topic that have been implicitly or explicitly written up in the past.

Philosophical Basis for Metaevaluation Methodology

Since metaevaluation (evaluation of evaluations) is a specialised form of the evaluation discipline, the plausible methodological approach must be situated within the context of evaluative enquiry and based on constructivist evaluation, which is grounded in the constructivist philosophical paradigm. Evaluation is one of the three basic disciplines of enquiry, the others being policy analysis and research (Guba and Lincoln 1989). Evaluation is a form of enquiry that focuses on the effectiveness of the evaluand. The evaluand is the subject of an evaluation and can be a project, programme, organisation or process.

The evaluation of the evaluand results in merit and worth judgments/constructions (Guba and Lincoln 1989; Guba and Lincoln 2001). Merit constructions/judgments focus on the intrinsic quality of the evaluand, and worth judgments/constructions focus on the extrinsic usefulness of the evaluand (Guba and Lincoln 2001). Constructions on the intrinsic quality of the evaluand focus mainly on determining the extent to which the project, programme, organisation or process is relevant to a specific context, and

constructions on the extrinsic usefulness focus on determining the outward beneficial impact of the project, programme, organisation or process to a specific localised context.

Guba and Lincoln (2001, 2) define constructivism as “a theory of knowledge that argues that humans generate knowledge and meaning from interactions between their experiences and their ideas,” whereas constructivist evaluation is a form of enquiry which results in constructions and judgments on the merit and worth of a specific evaluand, such as a project, programme, organisation or process.

The constructivism paradigm is based on three fundamental assumptions of knowledge. These assumptions of knowledge include ontological assumptions of relativism, epistemological assumptions of subjectivism, and methodological assumptions of hermeneutic-dialecticism (Guba and Lincoln 2001). Ontological assumptions of relativism propagate the claim that any point of view does not imply absolute truth, and therefore, such a point of view cannot be claimed to assume monolithic validity. For example, descriptions of factual evidence on the nature and effects of the project, programme, organisation or process are not presumed to represent absolute truth.

Baxter and Jack (2008) explain that the constructivist assumptions of relativism postulate the relativism of truth. Epistemological assumptions of subjectivism advance the claim that knowledge is subjective and dependent upon both one’s experiences and one’s ideas. The process of collecting and interpreting knowledge about the nature and effectiveness of the project, programme, organisation or process, therefore, culminates in subjective knowledge.

Methodological assumptions of hermeneutic-dialecticism promote the understanding that the process of the systematic assessment of the nature and effectiveness, as well as the interpretation of resulting knowledge from the evaluand, such as the project, programme, organisation or process, recognise the importance of the involvement of various interested parties, as well as the usefulness of the triangulation of tools of investigation and interpretation in constructing knowledge.

However, Stige, Malterud, and Midtgarden (2009) clarify that although the “constructivist paradigm recognises the importance of the subjective human creation of meaning; it does not reject the notion of objectivity.” Therefore, it can be inferred, based on this postulation on the importance of the subjective human creation of meaning, that objective views on reality can be described through evaluation case studies.

Guidelines for Conducting Metaevaluation

Evaluation literature proposes a variety of key guidelines for conducting metaevaluation. Stufflebeam (1974, 67), for example, proposes that metaevaluation should:

... assess the merit; serve the decision-making and accountability imperatives; assess goals, designs, implementation and results; provide judgmental information and appropriate recommendations; serve all persons who are involved in and affected by the evaluation; be conducted by both insiders and outsiders; be a process of delineating the questions to be addressed; obtaining the needed information and using the information in decision-making and accountability; and be technically adequate.

Patton (1997) specifies the questions on which a metaevaluation exercise must focus, such as: Was the evaluation well done? Is it worth using? Did the evaluation meet professional standards and principles? Patton (1997) further recommends that metaevaluation should be conducted with consideration of political contexts.

In addition, Scriven (2007) recommends the use of the evaluation checklist or standards of the PES developed by the Joint Committee for Program Evaluation of 1994. Wholey, Hatry, and Newcomer (2004) state that any evaluation must entail a description of the resources, activities, process and results caused by the programme, and identify issues, such as the relevance, effectiveness, efficiency and sustainability of the programme.

Furthermore, Bollen, Paxton, and Moroshima (2005) suggest that metaevaluation should focus on three necessary ingredients, which include information on inputs, information on results, and controls for confounding factors. These guidelines can be used to assess the quality of evaluations and serve to determine the degree to which an evaluation has complied with acceptable requirements.

Importance and Use of Metaevaluation

Since the development of the Programme Evaluations Standards by the Joint Committee on PES in 1994 (The Joint Committee 1994), there has been an upsurge of interest in the quality of evaluations of social interventions. Metaevaluation is used to assess the quality of either a single evaluation or a set of evaluations. The literature on metaevaluation outlines two broad categories of metaevaluation, which can be used to assess the quality of a single evaluation or a set of evaluations.

Shinkfield and Stufflebeam (2007) label these two categories as “formative” and “summative” metaevaluations. These authors explain that formative evaluations are intended to improve decision-making by helping evaluators to improve their planning and data-collection approaches during the process of evaluating. In addition to improving decision-making, it can be added that summative metaevaluation is also used to judge the strengths and weaknesses of a primary evaluation, so that readers can judge the reliability and credibility of primary evaluations.

Fitzpatrick, Sanders, and Worthen (2004) and Shinkfield and Stufflebeam (2007) explain that formative metaevaluation provides guidance on the quality of programme evaluations while in the process of evaluation. These authors argue that summative metaevaluation helps in preventing the dissemination of invalid conclusions, increasing

the utilisation of the results of the primary evaluation and improving the cost-effectiveness of the evaluation. Summative metaevaluations are conducted following an evaluation (Shinkfield and Stufflebeam 2007).

While Fitzpatrick et al. (2004) and Shinkfield and Stufflebeam (2007) explain that summative metaevaluation prevents the dissemination of invalid results of the primary evaluation, the researcher, in addition, argues that summative metaevaluation, overall, is used to validate primary evaluations, add credibility to them and enhance the confidence of the users of the results of these primary evaluations. In addition, summative metaevaluations help decision-makers to determine whether or not to make adjustments to the nature of interventions.

Wingate et al. (2010) explain that when a summative metaevaluation finds serious flaws in a primary evaluation, it can prevent decision-makers from taking action based on faulty information. In addition, summative metaevaluation assesses the quality of a completed evaluation, the appropriateness of the evaluation processes and the validity of its conclusions (Shinkfield and Stufflebeam 2007). Furthermore, Greene (1992), Cooksy and Caracelli (2005), and Caracelli and Cooksy (2008) propagate the understanding that summative metaevaluations help audiences to see the strengths and weaknesses of a primary evaluation, and judge its value.

Hanssen, Lawrenz, and Dunet (2008) also deduce that, while evaluation helps in improving programmes as well as learning to enhance the effectiveness of the evaluation profession, conducting metaevaluation improves the quality of evaluations, thus, enhancing the evaluation practice. Stufflebeam (2001) further argues that the field of evaluation has advanced sufficiently in its methodology and public service such that evaluators can and should subject their evaluations to systematic metaevaluation. In addition, Pierre (1983) argues that metaevaluation should also be concerned with the re-analysis of data generated in evaluations, in order to verify the credibility of early findings and clarify the inferences that may have been drawn.

Fitzpatrick et al. (2004) and Hanssen et al. (2008) all argue for the increased use of metaevaluation. In particular, Fitzpatrick et al. (2004) reiterate the lack of the use of metaevaluation and argue for its increased use in improving evaluation conduct and practice. Furthermore, Stufflebeam (2001) and Hanssen et al. (2008) strongly recommend the increased use of metaevaluation to describe and judge information concerning the strengths and weaknesses of an evaluation.

Wingate et al. (2010), Shinkfield and Stufflebeam (2007), Greene (1992), Cooksy and Caracelli (2005), and Caracelli and Cooksy (2008) generally agree in principle that summative metaevaluation determines the credibility of primary evaluations, improves learning and enhances decision-making. The researcher adapts the general narrative that metaevaluation enhances the effectiveness of both the evaluation practice and evaluation profession. Overall, it is within the context of advancing the evaluation

practice and profession that summative metaevaluation was adopted as being the most suitable form of metaevaluation for the purposes of this metaevaluation study.

In practice, metaevaluation has mainly been utilised in different developed country contexts to assess the quality of evaluations. Uusikylä and Virtanen (2000), for example, used metaevaluation to determine the quality of a case study of 15 evaluations of different European Structural Fund Programmes in Finland. Odom and Fewell (1983) conducted a metaevaluation of an evaluation of the Early Childhood Special Education programme in the USA. Apthorpe and Gasper (1982) used metaevaluation to assess the credibility of the evaluation of rural co-operatives in the Netherlands.

Conclusions

Situated in evaluative enquiry, metaevaluation (evaluation of evaluations) is a unique and specialised form of evaluation, which is grounded in the theory of good programme evaluation and idealised within the context of solving social problems faced by society. Metaevaluation is used to assess the quality of either a single evaluation or a set of evaluations and helps in judging the strengths and weaknesses of primary evaluations, so that interested parties are able to judge the reliability and evaluations, and thus can determine the quality of primary evaluations and the credibility of evaluation evidence used to make decisions. Metaevaluation has been utilised in different contexts in some developed countries to assess the quality of evaluations of development interventions, although scanty evidence exists of its use in developing country contexts.

References

- Alkin, M. C. 1969. “*Evaluating the Cost-effectiveness of Instructional Programs. Symposium presented at the Center for the Study of Evaluation.*” Report No. 25, University of California, Los Angeles. <https://files.eric.ed.gov/fulltext/ED031818.pdf>.
<https://doi.org/10.1037/e684812011-001>.
- Alkin, M. C. (Ed.). 2012. *Evaluation Roots: A Wider Perspective of Theorists’ Views and Influences*. Thousand Oaks, CA: Sage Publications.
- Apthorpe, R., and D. Gasper. 1982. “Policy Evaluation and Meta-evaluation: The Case of Rural Co-operatives.” *World Development* 10 (8): 651–668. [https://doi.org/10.1016/0305-750X\(82\)90091-2](https://doi.org/10.1016/0305-750X(82)90091-2).
- Baxter, P., and S. Jack. 2008. “Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers.” *The Qualitative Report* 13 (4): 544–559.
- Borich, G. D. 1990. *Decision Oriented Evaluation*. Oxford, England: Pergamon Press.
- Bollen, K., P. Paxton, and R. Morishima. 2005. “Assessing International Evaluations: An Example from USAID’s Democracy and Governance Program.” *American Journal of Evaluation* 26 (2): 189–203. <https://doi.org/10.1177/1098214005275640>.

- Caracelli, V. J., and L. J. Cooksy. 2008. "Meta-evaluation in Practice: Selection and Application of Criteria." *Journal of Multidisciplinary Evaluation* 6 (11): 1–15.
- Cooksy, L. J., and V. J. Caracelli. 2005. "Quality, Context, and Use: Issues in Achieving the Goals of Meta-evaluation." *American Journal of Evaluation* 26 (1): 31–42. <https://doi.org/10.1177/1098214004273252>.
- Cronbach, L. J. 2000. "Course Improvement through Evaluation." In *Evaluation Models: Viewpoints on Educational and Human Services*, 2nd edition, 235–247, edited by D. L. Stufflebeam, G. F. Madaus, and T. Kellaghan. Hingham, MA: Kluwer. https://doi.org/10.1007/0-306-47559-6_14.
- Cronbach, L. J., S. R. Ambron, S. M. Dornbusch, R. D. Hess, R. C. Hornik, D. C. Phillips, and S. S. Weiner. 1980. *Toward Reform of Program Evaluation*, 3. San Francisco: Jossey-Bass.
- Fitzpatrick, J. L., J. R. Sanders, and B. R. Worthen. 2004. *Program Evaluation: Alternative Approaches and Practical Guidelines*. Boston, MA: Pearson Publishing.
- Glass, G. V. 1969. "The Growth of Evaluation Methodology." Research paper no. 27, Boulder, Colo., University of Colorado: Laboratory of Educational Research.
- Greene, J. C. 1992. "A Case Study of Evaluation Auditing as Meta-evaluation." *Evaluation and Program Planning* 15 (1): 71–74. [https://doi.org/10.1016/0149-7189\(92\)90063-Z](https://doi.org/10.1016/0149-7189(92)90063-Z).
- Guba, E. G., and Y. S. Lincoln. 1989. *Fourth Generation Evaluation*. Thousand Oaks, CA: Sage Publications.
- Guba, E. G., and Y. S. Lincoln. 2001. *Guidelines and Checklist for Constructivist (aka Fourth Generation) Evaluation*. Beverly Hills, CA: Sage Publications.
- Hanssen, C. E., F. Lawrenz, and D. O. Dunet. 2008. "Concurrent Meta-evaluation: A Critique." *American Journal of Evaluation* 29 (4): 572–582. <https://doi.org/10.1177/1098214008320462>.
- Hedler, H. C., and N. Gibram. 2009. "The Contribution of Metaevaluation to Program Evaluation: Proposition of a Model." *Journal of MultiDisciplinary Evaluation* 6 (12): 210–223.
- Nevo, D. 1983. "The Conceptualization of Educational Evaluation: An Analytical Review of the Literature." *Review of Educational Research* 53 (1): 117–128. [doi:10.3102/00346543053001117](https://doi.org/10.3102/00346543053001117).
- Odom, S. L., and R. R. Fewell. 1983. "Program Evaluation in Early Childhood Special Education: A Meta-evaluation." *Educational Evaluation and Policy Analysis* 5 (4): 445–460. <https://doi.org/10.3102/01623737005004445>.

- OECD. 2004. *The Organisation for Economic Co-operation and Development (OECD)*. Paris: Routledge Publishing.
- Orr, R. R., S. J. Cameron, and D. M. Day. 1991. "Coping with Stress in Families with Children who have Mental Retardation: An Evaluation of the Double ABCX Model". *American Journal on Mental Retardation* 95 (4): 444–450.
- Owen, J. M., and P. Rogers. 1999. *Programme Evaluation: Forms and Approaches*. Thousand Oaks, CA: Sage Publications.
- Patton, M. Q. 1997. "Toward Distinguishing Empowerment Evaluation and Placing it in a larger Context." *Evaluation Practice* 18 (2): 147–163.
- Pierre, R. G. 1983. *Reanalyzing Programme Evaluation*. Basel: Jossey-Bass Publishers.
- Rossi, P. H., H. E. Freeman, and M. W. Lipsey. 1999. *Evaluation. A systematic Approach*. Thousand Oaks, CA: Sage Publications.
- Scriven, M. 1966. *Primary Philosophy*. Thousand Oaks, CA: Sage Publications.
- Scriven, M. 1969. An Introduction to Metaevaluation. *Educ Prod Rep* 2 (4): 36–38.
- Scriven, M. 2007. *Key Evaluation Checklist (KEC)*. Thousand Oaks, CA: Sage Publications.
- Shadish, W. R., T. D. Cook, and L. C. Leviton. 1991. *Foundations of Program Evaluation: Theories of Practice*. Thousand Oaks, CA: Sage Publications.
- Shinkfield, A. D., and L. Stufflebeam. 2007. *Evaluation Theory, Models and Applications*. San Francisco: CA: John Wiley. Thousand Oaks, CA: Sage Publications.
- Stige, B., K. Malterud, and T. Midtgarden. 2009. "Toward an Agenda for Evaluation of Qualitative Research." *Qualitative Health Research* 19 (10): 1504–1516. <https://doi.org/10.1177/1049732309348501>.
- Stufflebeam, D. L. 1971. *The Relevance of the CIPP Evaluation Model for Educational Accountability*. Abingdon, London: Routledge Publishing.
- Stufflebeam, D. L. 1974. *Meta-evaluation*. MI: Western Michigan University Evaluation Centre.
- Stufflebeam, D. L. 2000. *Foundational Models for 21st Century Program Evaluation*. Dordrecht, Netherlands: Springer.
- Stufflebeam, D. L. 2001. "The Meta-evaluation Imperative." *American Journal of Evaluation* 22 (2): 183–209. <https://doi.org/10.1177/109821400102200204>; [https://doi.org/10.1016/S1098-2140\(01\)00127-8](https://doi.org/10.1016/S1098-2140(01)00127-8).

- Stufflebeam, D. L., and W. J. Webster. 1983. "An Analysis of Alternative Approaches to Evaluation." In *Evaluation Models*, 23–43, edited by D. L. Stufflebeam. Dordrecht, Netherlands: Springer. https://doi.org/10.1007/978-94-009-6669-7_2.
- The Joint Committee on Standards for Educational Evaluation. 1994. *The Programme Evaluation Standards*, 2nd edition. Thousand Oaks, CA: Sage Publications.
- Tyler, R. W. 1942. "General Statement on Evaluation." *The Journal of Educational Research* 35 (7): 492–501. <https://doi.org/10.1080/00220671.1942.10881106>.
- Tyler, R. W. 1950. "Basic Principles of Curriculum and Instruction: Syllabus for Education 360." IL: Chicago.
- Uusikylä, P., and P. Virtanen. 2000. "Meta-evaluation as a Tool for Learning: A Case Study of the European Structural Fund Evaluations in Finland." *Evaluation* 6 (1): 50–65. <https://doi.org/10.1177/13563890022209118>.
- Weiss, C. H. 1998. *Evaluation*. Upper Saddle, NJ: Prentice Hall Publishing.
- Wholey, J. S., H. P. Hatry, and K. E. Newcomer (Eds). 2004. *Handbook of Practical Programme Evaluation*. San Francisco, CA Jossey-Bass Publishers.
- Wingate, L. A., C. L. S. Coryn, A. R. Gullickson, and L. J. Cooksy. 2010. The Program Evaluation Standards Applied for Meta-evaluation Purposes: Investigating Interrater Agreement and its Implications for Use." *Manuscript submitted for publication*.
- Wittrock, M. C., and D. E. Wiley. 1970. *The Evaluation of Instruction: Issues and Problems*. New York, NY: Holt, Rinehart and Winston Publishing.
- World Bank. 2010. *African Development Indicators 2008*. Washington, DC: World Bank Publications.