META EVALUATION PRACTICE IN AN ENGLISH DISTANT TRAINING PROGRAM IN JAKARTA

Efi Dyah Indrawati

efidyahindrawati@kemenkeu.go.id

Finance Education and Training Agency, Indonesia

ABSTRACT

Meta-**evaluation** is a means used to aggregate findings from a series of evaluations. It could be said like an audit of certain evaluation results using particular standards and criteria. The practice of meta-evaluation has not been widely carried out in the world, even more in learning programs organized by government institutions in Indonesia. This study is intended to evaluate the results of evaluation of a distant training program namely PJJ English for Customs Officers (EFCO Training) at Pusdiklat Keuangan Umum (General Finance Education and Training Center) or GFETC of Indonesia Ministry of Finance as the object of research. The study is expected to answer whether the evaluation of EFCO training meets the requirements of the utility, feasibility, propriety, and accuracy standards. Three meta-evaluators (1 internal and 2 external) gave a Yes or No tick to 290 items in the Meta Evaluation Checklist after they analyze the qualitative and quantitative evaluation data (documents) from the Reaction Evaluation (Level 1 Kirkpatrick Evaluation) and Learning Result Evaluation (Level 2 Kirkpatrick Evaluation) using Stufflebeam's Program Evaluation Standards. The conclusion is that Level 1 and Level 2 (Kirkpatrick) evaluation of EFCO training met the standards of utility, feasibility, propriety, and accuracy. Another conclusion is that the implementation of meta-evaluation enables GFETC to indicate specific aspects of a training program that need to be improved in quality, and that meta evaluation can be used for quality assurance and decision making. The limitation of the research is that it has not been used to evaluate Level 3 and Level 4 of Kirkpatrick's evaluation and that no panel review with stakeholders is conducted after the meta evaluation is completed. Recommendations of the research are the refined translation of standard items to check all the 4 levels of Kirkpatrick's evaluation and that the research should be conducted for more training programs within longer period of evaluation.

Keywords: meta-evaluation, 4 level Kirkpatrick's evaluation, utility, feasibility, propriety, accuracy

INTRODUCTION

The field of evaluation has progressed significantly in both methodology and public service to the extent that evaluators have the capability and should indeed subject their evaluations to systematic meta-evaluation. Meta evaluation is the evaluation of evaluation (Scriven, 1975). It is a professional imperative or obligations for evaluators (Stufflebeam, 2001) to delineate, obtain, and apply descriptive information and judgmental information about an evaluation's utility, feasibility, propriety. It also encompasses an assessment of its systematic nature, competence, honesty, respectfulness, and social responsibility to guide the evaluation and publicly disclose its strengths and weaknesses.

Various criteria have been set forth for what constitutes excellence in evaluation. In the context of educational program evaluation, the dominant criteria are the Program Evaluation Standards (PES), developed by Joint Committee on Standards for Educational Evaluation (JCSEE). Numerous institutions, including member countries of the Organization for Economic Co-operation and Development (OECD) like the United States, Denmark, and Turkey, adhere to these meta-evaluation standards. Nevertheless, there has been limited empirical study of evaluation standard and the practice of meta-evaluation has not been carried out in many other countries, including Indonesia.

In the context of English language training programs tailored for government officers of Indonesian government institutions, the application of meta-evaluation practices emerges as a critical area of study, as all government trainings are prepared using systematic process that involves evaluation. As government officers in Indonesia Ministry of Finance (MOF) increasingly recognize the importance of proficient English communication skills in diplomatic, administrative, and international contexts, the evaluation of the effectiveness of such training programs becomes imperative. Understanding the overview of meta-evaluation within the unique context of MOF officers serving for the Directorate General of Customs and Excise (DGCE) is essential for optimizing the design and delivery of these programs, identifying strengths and weaknesses, and ultimately informing evidence-based strategies for continuous improvement of the program. This research endeavours to study the application of meta-evaluation practices within the English for Customs Officers (EFCO) training in the General Finance Education and Training Center (GFETC), Jakarta, offering insights that can contribute to the refinement of training programs tailored to the specific needs of DGCE employees to improve MOF organizational performance. The study is based on the following questions:

1) Did the evaluation of EFCO training meet the utility standards?
2) Did the evaluation of EFCO training meet the feasibility standards?
3) Did the evaluation of EFCO training meet the proprietary standards?
4) Did the evaluation of EFCO training meet the accuracy standards?

By conducting this research, it is hoped that this will provide the first study on the meta evaluation process for Indonesia MOF and will improve the service of GFETC to stakeholders.

Meta evaluation Process

This study uses the eleven-step process first proposed by Stufflebeam (2000) and refined by Stufflbeam and Coryn (2014) as the followings:

1. *Staffing*: select qualified meta evaluators to ensure a thoughtful but constructive assessment
2. *Stakeholder Engagement*: Identify and arrange to interact with the meta evaluation's stakeholders for data collection and meta evaluation question development
3. *Standards*: Agree on standards, principles, or criteria to judge the evaluation system or particular evaluation
4. *Questions*: Develop the meta evaluation questions based on the standards and stakeholder feedback.
5. *Formal Agreements*: sign a memo of understanding or a formal meta evaluation contract to prevent any possible disagreements
6. *Existing Information*: Collect and determine all relevant, available information on the evaluation being meta-evaluated
7. *New Information:* Collect additional information as needed, including the program information and other documents

8. *Analysis and synthesis*: Analyze and synthesize the obtained information to determine if the existing criteria for each meta-evaluation question were fulfilled and to what degree.
9. *Reaching Conclusions*: draw a general conclusion from the assessment or evaluation system based on its conformity to relevant standards, principles, or criteria
10. Reporting: share the meta evaluation results through reports, correspondence, oral presentations, workshops, and other means
11. *Follow-Up*: As appropriate and feasible, help the client and other stakeholders interpret and use the results

Based on the context of study, those processes were applied within four phases: pre-planning, phase 1, phase 2, and phase 4. The Follow-Up step was not accomplished fully due to time limitation, yet the medium was made available for further assistance.

METHODOLOGY

The study employs an evaluation research design. The object of the study was the learning evaluation results of EFCO training year 2021 and 2022, which was carried out at Level 1 and Level 2 of the Kirkpatrick Evaluation Model. Level 1 Evaluation (Reaction) refers to the degree to which training participants consider the training favorable, engaging and relevant to their jobs. Level 2 Evaluation (Learning) refers to the degree to which participants acquire the intended knowledge, skills, attitude, confidence, and commitment based on their participation in the training (Catalanelo & Kirkpatrick, 1968). The evaluation of learning outcomes can be seen in changes in attitudes, improved knowledge, and/or increased skills of participants after completing certain training program (Kirkpatrick & Kirkpatrick, 2008).

Three meta evaluators were involved (1 internal and 2 external) to guarantee the evaluation credibility (Stufflebeam, 2004), selected based on their experience in evaluation and their distant degree from the program under evaluation. The qualitative data were collected from document analysis and interview with the stakeholders on the meta evaluation process. The quantitative data were obtained from the meta evaluators using meta evaluation checklist adapted and translated in Indonesian from Stufflebeam's (1999) based on the Program Evaluation Standards. The checklist consists of 4 main standards: utility (80 items), feasibility (30 items), propriety (70 items), and accuracy (120 items), all had "Yes" and "No" answers based on whether the standards in question existed in the program evaluation or not. All meta evaluators were reminded to correctly complete the checklist to ensure the data given were reliable.

The JCSEE defines the program evaluation standard as follows:
- **Utility Standards**
  The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs, and the sub-standards are
      U1 Evaluator Credibility Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.h6
      U2 Attention to Stakeholders Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.
      U3 Negotiated Purposes Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.
      U4 Explicit Values Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.
      U5 Relevant Information Evaluation information should serve the identified and emergent needs of stakeholders.

U6 Meaningful Processes and Products Evaluations should construct activities, descriptions, and judgments in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviors.

U7 Timely and Appropriate Communicating and Reporting Evaluations should attend to the continuing information needs of their multiple audiences.

U8 Concern for Consequences and Influence Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse.

- **Feasibility Standards**

The feasibility standards are intended to increase evaluation effectiveness and efficiency, and the sub-standards are

F1 Project Management Evaluations should use effective project management strategies.

F2 Practical Procedures Evaluation procedures should be practical and responsive to the way the program operates.

F3 Contextual Viability Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.

- **Propriety Standards**

The propriety standards support what is proper, fair, legal, right and just in evaluations, and the sub-standards are

P1 Responsive and Inclusive Orientation Evaluations should be responsive to stakeholders and their communities.

P2 Formal Agreements Evaluation agreements should be negotiated to make obligations explicit and take into account the needs, expectations, and cultural contexts of clients and other stakeholders.

P3 Human Rights and Respect Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders.

P4 Clarity and Fairness Evaluations should be understandable and fair in addressing stakeholder needs and purposes.

P5 Transparency and Disclosure Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations.

P6 Conflicts of Interests Evaluations should openly and honestly identify and address real or perceived conflicts of interests that may compromise the evaluation.

P7 Fiscal Responsibility Evaluations should account for all expended resources and comply with sound fiscal procedures and processes.

- **Accuracy Standards**

The accuracy standards are intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality. The sub-standards are

A1 Program Documentation The program being evaluated should be described and documented, so that the program is clearly identified.

A2 Context Analysis The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified

A3 Described Purposes and Procedures The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.

A4 Defensible Information Sources The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.

A5 Valid Information Evaluation information should serve the intended purposes and support valid interpretations.

A6 Reliable Information Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.

A7 Systematic Information Evaluations should employ systematic information collection, review, verification, and storage methods.

A8 Analysis of Quantitative Information Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

A9 Analysis of Qualitative Information Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

A10 Justified Conclusions Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequences.

A11 Impartial Reporting Evaluation reporting should guard against misconceptions, biases, distortions, and errors of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings

A12 Metaevaluation The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses

The collected data underwent scoring and analysis using Stufflebeam's (1999) 'Meta evaluation Checklist' guidelines. The judgement of each standard (criterion) was determined by referring to scoring tables, categorizing them as Excellent (93%), Very Good (68%), Good (50%), Fair (25%), and Poor (0%).

It is recommended that an evaluation be failed if it scores "Poor" on standards P1 Service Information, A5 Valid Information, A10 Justified Conclusions, or A11 Impartial Reporting.

FINDINGS AND DISCUSSION

The meta evaluators discovered that providing checklists for each evaluation level was a more straightforward and simpler task, considering the extensive and diverse nature of EFCO training evaluation documents. Therefore, the summary of the checklist scoring given is as follows:

Table 1 Summary of Checklist Scoring

Evaluator	Utility		Feasibility		Propriety		Accuracy	
	Level 1	Level 2	Level 1	Level 2	Level 1	Level 2	Level 1	Level 2
1	27	26	11	11	24	22	47	46
2	23	27	11	11	24	25	47	47
3	23	23	10	10	25	25	32	37
Evaluation	Excellent	Very Good	Excellent	Excellent	Very Good	Very Good	Very Good	Very Good

As for each standard fulfilment, the findings are shown on the following tables. As seen on Table 2, the utility standards, one evaluator assessed the evaluation as "Excellent" and the other two evaluators judged it as "Very Good". For the feasibility standards on Table 3, all evaluators approved the evaluation as "Excellent". This means that feasibility is the strongest aspect of this evaluation.

Table 2 Evaluation of Utility Standards

Meta Evaluator	Evaluation Score	Evaluation
1	93	Excellent
2	82	Very Good
3	86	Very Good

Table 3 Evaluation of Feasibility Standards

Meta Evaluator	Evaluation Score	Evaluation
1	92	Excellent
2	92	Excellent
3	92	Excellent

As for the propriety standards, Table 4 shows that all evaluators agreed that the evaluation is "Very Good". In the section of accuracy standards shown on Table 5, two evaluators judged the evaluation as "Excellent" and one evaluator assessed it as "Very Good".

Table 4 Evaluation of Propriety Standards

Meta Evaluator	Evaluation Score		Evaluation
1	86		Very Good
2	86		Very Good
3	89		Very Good

Table 5 Evaluation of Accuracy Standards

Meta Evaluator	Evaluation Score	Evaluation
1	98	Excellent
2	98	Excellent
3	63	Good

Other findings are that several sub-standards were evaluated as "Excellent" by all meta evaluators: P1 *Service Orientation*, A5 *Valid Information*, A10 *Justified Conclusions and Decisions*, and A11 *Impartial Reporting*. However, Evaluator 3 judged A2 *Context Analysis* as "Fair" and A12 *Meta evaluation* as "Poor". Dealing with context analysis, Evaluator 3 considered a wider context, yet the standard items were limited to technical, social, political, organizational and economic feature. As for the meta evaluation, it was previously agreed in the pre-planning phase that meta evaluation could be equated with the learning quality assurance at the training center. However, Evaluator 3 had his own personal belief of the inadequacy of quality assurance of EFCO training, whereas Evaluator 1 and Evaluator 2 both assessed sub-standard A12 *Meta evaluation* as "Excellent". This can be a limitation for this study as the selection of meta evaluators merely based on those who are considered competent in both evaluation theories and practices, but none of them is a professional evaluator with prior experience in meta evaluation.

There are still several other findings such as Evaluator 2 assessed "Fair" for U6 *Timely Reporting* for Level 2 Evaluation, Evaluator 1 gave a "Fair" on P5 *Complete and Fair Assessment* in Evaluation Level 2, and Evaluator 2 and Evaluator 3 gave "Fair" on P7 *Conflict of Interest* in both Evaluation Level 1 and Evaluation Level 2. Detailed discussion can be found on the original format of this study. In short, any judgement made by meta evaluators becomes an important evaluation record for the organizer and stakeholders of EFCO training.

An insight for the researcher is that during the preparation phase of the meta-evaluation, coordination involving all stakeholders requires quite a lengthy time. This is due to the difficulty in the coordination with the relevant evaluation staff and the Evaluation Division because of employee transfers and mutation, which means that the person in charge of administering the EFCO Training evaluation cannot be ascertained. In addition, there are some obstacles in translating the Meta Evaluation Checklist instrument to suit the context of learning programs in government agencies. To overcome this, the translation was carried out by a joint team within GFETC and also by juxtaposing each evaluation item with its original formulation in English to facilitate understanding. Apart from that, the Meta Evaluation Checklist is separated for Evaluation Level 1 and Evaluation Level 2 to make it easier for evaluators to assess standards according to their objects considering the large number of documents being examined within the given time for assessment. The meta evaluation instrument can actually be used to measure all evaluation levels in the Kirkpatrick's Evaluation Model, yet EFCO training is contextually limited to only Level 1 and Level 2 evaluation in the course design. Nevertheless, the translated Meta Evaluation Checklist still needs to be further refined for better implementation of meta evaluation practices in the future.

Furthermore, there are time constraints for onsite FGDs between stakeholders and meta evaluators. To overcome this, all coordination was carried out using online collaborative tools. Furthermore, researchers must ensure that evaluators have sufficient independent space to provide assessments with a perspective that is free from any influence. This is how meta-evaluations should be conducted, especially when the goal is to help an organization assess and reform its evaluation system (Stufflebeam, 2001). When the goal is to protect the public from being misinformed by a particular evaluation, examiners must maintain an appropriate distance to ensure an independent perspective. However, meta-evaluators must communicate appropriately with the audience or users of evaluation reports to safeguard the report's contribution, trustworthiness, importance, understanding, and informed use of the findings (Stufflebeam, 2001). Therefore, the use of Whatsapp, email, Google Sites, Google Forms, and Microsoft Teams in this research has met the demands of conducting an accountable evaluation.

The next insight from this study is that there are many areas of program evaluation that can be carried out using Stufflebeam's program evaluation standards. In order to be able to answer all the standard questions (a total of 290 items in this study), a very complete data evidence or evaluation documents are mandatory. This took a lengthy of time (3 months) as they have to be requested from different evaluation PICs. However, the advantage of the detailed standard question items makes it easier for researchers to indicate which parts of the EFCO training evaluation that need special attention. By looking at the results of A2 *Context Analysis*, for example, it can help GFETC to view whether the need analysis and program design are good or not. Thus, meta-evaluation can be used as a decision-making tool to improve certain learning program and as a tool for the quality assurance of a program evaluation.

CONCLUSIONS

It can be concluded that the evaluation of EFCO training met the standards of utility, feasibility, propriety, and accuracy. Some recommendations of this study are:

1. The meta evaluation can be used further to pinpoint specific area of improvement, especially when conducted separately between Level 1 Evaluation and Level 2 Evaluation.
2. Meta evaluation can be used for quality assurance and decision making.

3. GFETC needs the refined translation of standard items to check all the 4 levels of Kirkpatrick's evaluation
4. The meta evaluation research should be conducted for more training programs within longer period of evaluation.

REFERENCES

Glass, G.V., McGaw B., & Smith, M.L. 1981. *Meta-Analysis in Social Research*. London: Sage Publications. On *Meta Evaluation: An Overview by Daniel L. Stufflebeam.*

Kirpatrick, James D. & Wendy Kaiser Kirkpatrick. 2008. *Kirkpatrick's Four Levels of Training Evaluation*. Alexandria: ATD Press.

Ministry of Foreign Affairs, Danida. 2004. *Meta Evaluation: Private and Business Sector Development Interventions.* https://www.oecd.org/derec/denmark/36478191.pdf

Rallis, Sharon F. & Kathleen A. Bolland. 2004. *What is Program Evaluation? Generating Knowledge for Improvement.*
https://www.researchgate.net/publication/226571431_What_is_Program_Evaluation_Generating_Knowledge_for_Improvement.

Stufflebeam, D. L., Coryn, C. L. S. (2014). *Evaluation theory, models, and applications* (2nd ed.). Jossey-Bass/Wiley. https://psycnet.apa.org/record/2014-43319-000

Stufflebeam, D.L, George F. Madaus, & T. Kellaghan. 2000. Evaluation Models: Viewpoints on Educational and Human Services Evaluation. Dordrecht: Kluwer Academic Publisher. https://books.google.co.id/books?hl=id&lr=&id=5EAyBwAAQBAJ&oi=fnd&pg=PR7&dq=madaus+scriven+stufflebeam+CIPP&ots=9g3CAlaiFq&sig=VxNokrnZYXLFLpXjuyfTT5jx1Ac&redir_esc=y#v=onepage&q=madaus%20scriven%20stufflebeam%20CIPP&f=false.

Stufflebeam, Daniel L. 1971. *The CIPP Model for Evaluation.* https://link.springer.com/chapter/10.1007/978-94-010-0309-4_4.

Stufflebeam, Daniel L. 1978. *Meta Evaluation: An Overview.* https://journals.sagepub.com/doi/abs/10.1177/016327877800100102?journalCode=ehpa.

Stufflebeam, Daniel L. 2001. *Evaluation Checklists: Practical Tools for Guiding and Judging Evaluations.* https://journals.sagepub.com/doi/10.1177/109821400102200107

Stufflebeam, D. L. (2004). A note on the purposes, development, and applicability of the Joint Committee Evaluation Standards. *American Journal of Evaluation,* 25(1), 99-102 https://doi.org/10.1016/j.ameval.2003.12.002

The Commonwealth. 2016. Meta Evaluation: A Synthesis of Evaluation Studies 2005-2016. https://thecommonwealth.org/sites/default/files/inline/104%20-%20Meta-Evaluation%20Report%20-%202016.pdf.

Ulker, Kerimoglu & Berk (2019). *How to Conduct A Metaevaluation: A Metaevaluation Practice.* Canadian Journal of Program Evaluation, vol.38, no.1, pp.57-78, 2023 (ESCI) http://doi.org/10.3138/cjpe.71619

Yarbrough, D.B., Shula, L.M., Hopson, R.K., & Caruthers, F.A. (2010). *The Program Evaluation Standards : A guide for evaluators and evaluation users* (3rd. ed). Thousand Oaks, CA: Corwin Press.