https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yildz Teknik Üniversitesi IP Address:193.140.7.203

How to Conduct a Metaevaluation?: A Metaevaluation Practice[1]

Esra Kerimoğlu
Yildiz Technical University, İstanbul, Turkey

Muazzez Nihal Öykü Ülker
İstanbul Technical University, İstanbul, Turkey

Şaban Berk
Marmara University, İstanbul, Turkey

Abstract: *A metaevaluation is a quality cross-check to examine the conduct of an evaluation and validate the results. Of the few metaevaluation studies, almost none have reported on the metaevaluation procedure through a practical example evaluation. This study reports on the strengths and weaknesses of a program evaluation study in terms of the four main standards: utility, feasibility, propriety, and accuracy. It includes a metaevaluation process that involves both quantitative and qualitative analysis of data from eight meta-evaluators. It was found that while the evaluation study had very good utility and accuracy standards, the feasibility and propriety standards were only fair.*

Keywords: *metaevaluation, metaevaluation practice, evaluating the evaluation, program evaluation*

Résumé : *Une métaévaluation et une vérification de la qualité pour examiner la façon dont une évaluation a été menée et pour valider les résultats. Parmi les quelques études effectuées sur les métaévaluations, presque aucune n'a parlé de la procédure de métaévaluation par l'intermédiaire d'une évaluation par exemples pratiques. La présente étude traite des points forts et des points faibles d'une étude d'évaluation de programme pour ce qui est de quatre normes principales : utilité, faisabilité, pertinence et exactitude. Elle inclut un processus de métaévaluation qui comprend une analyse quantitative et qualitative des données de huit métaévaluateurs/évaluatrices. On a conclu que l'étude d'évaluation avait de très bonnes normes d'utilité et d'exactitude, mais que l'atteinte des normes de faisabilité et de pertinence était moyenne.*

Mots clés : *métaévaluation, pratique de la métaévaluation, évaluation de l'évaluation, évaluation de programme*

Corresponding author: Esra Kerimoğlu, Davutpasa Campus, Faculty of Education, Yildiz Technical University, 34220 Esenler, Istanbul, Turkey; esrak@yildiz.edu.tr

© 2023 *Canadian Journal of Program Evaluation / La Revue canadienne d'évaluation de programme*
38.1 (June / juin), 57–78 doi: 10.3138/cjpe.71619

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

INTRODUCTION

Societies that do not improve their programs based on science and technological developments and the needs of the individual and society are educating people as if they were in the past (Berk, 2018a). Program improvements are made after evaluation of the program elements and a review of stakeholder expectations (Şentürk & Berk, 2019). Achievement of program objectives by learners within prescribed criteria will define program success. This, however, may not be the case every time. It is required to evaluate the program in order to determine whether there are elements that fail to achieve the anticipated objectives or perform as planned in the wake of implementation of a program, and which elements of the program caused failures if any (Demirel, 2017). Not only is maintaining program improvement and effective expansion regarded as the benefit of making program evaluations, but so, too, are supporting policymakers and senior leaders in their actions of strategic decision-making (e.g., budgetary expense plans, [re]allocation of resources), contributing to organization-level accountability, enlightening stakeholders and the like (Bourgeois & Whynot, 2018).

Evaluation studies in the Turkish education system are in a developing phase. Turkey has only recently seen the emergence of organizations similar to the Council for the Accreditation of Educator Preparation in the United States, which evaluates departments, schools, and universities preparing teachers and other educators. Over the past 10 years, the Association for Evaluation and Accreditation of Teacher Education Programs (EPDAD) has gained prominence. This association established standards for teacher education undergraduate programs across the country (EPDAD, 2016; 2021). By comparison, curriculum evaluation studies in Turkey are typically carried out at two levels: (1) by the Ministry of National Education (MoNE), through its institutions when deemed necessary, and (2) by academicians and other scholars through theses, papers, or presentations (Özdemir, 2009).

While many evaluations have been conducted on program effectiveness, adequacy, and deficiencies, these evaluations also need to be reviewed to ensure they were conducted in line with the specific evaluation standards. Therefore, how can evaluators be sure of the quality of their evaluations and how can the virtues and shortcomings of an evaluation be assessed? The answer is through metaevaluations.

Metaevaluations are basically audits of an evaluation's accountability to determine its soundness; that is, they are a type of quality cross-check to examine the correctness of the evaluation study and the accuracy of the results. This concept was first mooted by Scriven in 1969, who defined a metaevaluation as "an evaluation of the evaluation or the evaluator" (Scriven, 1975). Cooksy and Caracelli (2009) have more recently defined it as a systematic review to determine the quality of the methods used in the evaluation and the validity of the results. Snow (2001) claimed that as metaevaluations were focused on an evaluation's quality, they would need to be conducted by a third party (meta-evaluator) to eliminate any misleading elements that might be in the report and ensure

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

evaluation "synthesis," which is of critical importance and seen by many as a key metaevaluation objective (Leuww & Cooksy, 2005; Stufflebeam & Shinkfield 2007; Wingate, 2009). This "synthesizer" function of metaevaluations, according to the Commonwealth Secretariat, enables making broader operational suggestions for use in strategic planning processes. By the same token, the Organisation for Economic Co-operation and Development (OECD, 2011) emphasizes the overarching role of metaevaluation as a synthesis, pinpointing its significance for many purposes: (a) aggregating findings from a series of evaluations, (b) gauging the extent to which program or project objectives are implemented, (c) guiding on high-quality implementation and sustainability, (d) determining the impact of results at a country scale, (e) providing insights into taking decisions on macro-level policy, and (f) enabling cross-country comparisons. Apart from providing an overall assessment of evaluations, it accentuates key recommendations to be implemented over the period, evaluates the degree of the successful implementation, and proposes a monitoring system of as-yet-unfulfilled follow-up work (Bobin, 2017).

The study findings of Kürüm-Yapıcıoğlu et al. (2016) indicate that there are important theoretical shortcomings in the field of curriculum evaluation in Turkey, and thus, there is a critical need to improve the calibre of the research being done in this area. Regardless of conceptually understanding program evaluation quality or assessing its validity (Yarbrough, 2017) and many other previously mentioned benefits for conducting metaevaluations, there have been found no studies to date that have clearly shown how a metaevaluation is conducted in the Turkish context, which is the primary aim of this study. This research sought to explain the metaevaluation process using a practical example from Şentürk's (2017) evaluation of a third-grade science curriculum (SC), which was initially put into practice in Turkey in the 2014–2015 academic year.

Metaevaluation Conduct

Metaevaluations can be conducted internally or externally (Stufflebeam, 2000). An internal meta-evaluator examines their own program evaluation using a validated checklist to assess its accountability (Cooksy & Caracelli, 2009), and could also perform both formative and summative evaluations (Stufflebeam, 2001) to assess the systematicity of their evaluation steps. External metaevaluations can be conducted by one or more independent evaluators; however, more than one meta-evaluator is considered important to guarantee evaluation result soundness. Stufflebeam (2004) also noted that the quality of an evaluation could be assessed by program sponsors, clients or other stakeholders using the Joint Committee on Standards for Educational Evaluation (JCSEE) standards without the assistance of a professional evaluator. Earlier, Scriven (1975) had suggested that the bias and errors in evaluation studies could be highlighted using metaevaluations that could involve (a) the evaluator evaluating their own work, (b) experts evaluating the evaluation, or (c) establishing two different evaluation groups that would evaluate each other's evaluations. In the third JCSEE standards edition, internal

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

metaevaluation and external metaevaluation were added to the evaluation accountability criterion subdimensions (Yarbrough et al., 2010), and, in line with these changes, in 2012, Patton was awarded an Outstanding Evaluation Award by the American Evaluation Association for his metaevaluation on an internal evaluation of the Paris Declaration on Aid Effectiveness by the OECD (Patton, 2017), which was a primary example of an evaluator evaluating their own work.

Because there have been many studies in the program evaluation field, it is possible to gain a common understanding by reviewing research findings on the effectiveness of certain programs; however, this requires a metaevaluation to examine the effectiveness of these evaluations. Just as the program effectiveness is assessed using program evaluation, the evaluation effectiveness is assessed using a metaevaluation, which is generally done by measuring the evaluation accountability based on several criteria. Alkin (2012) claimed that this accountability was related to "answerability" in the widest sense, that is, an evaluation should have satisfactory answers to a confirmatory analysis that exposes its deficiencies. Patton (2018) questioned how a weak and sloppy evaluation could be recognized and how this conclusion could be presented to others, with the answer being through a metaevaluation. Stufflebeam and Coryn (2014) claimed that metaevaluations are useful for the development of program evaluation models as the results could be used to determine why various evaluation approaches were successful or unsuccessful and to develop new theories. For example, Gardner (2019) conducted a metaevaluation as the last step in his research with the aim of developing a new program evaluation model.

Metaevaluation Process

Various metaevaluation methods have been suggested; however, the following eleven-step process first proposed by Stufflebeam (2000) and refined by Stufflebeam and Coryn (2014) has provided a general road map for many metaevaluation studies.

1. *Staffing:* Select meta-evaluators with varied technical qualifications and content knowledge to ensure a sound but critical assessment.
2. *Stakeholder Engagement:* Identify the stakeholders to be included in the metaevaluation for data collection and metaevaluation question development.
3. *Standards:* Decide on the validated, reliable and easy-to-implement foundation standards or principles on which to build the metaevaluation.
4. *Questions:* Develop questions based on the standards and stakeholder feedback. For example, Russ-Eft and Preskill (2008) conducted a metaevaluation that was based entirely on questions that had been developed from document reviews and interviews with stakeholders, that is, no standards had been referred to.
5. *Formal Agreements*: Sign a formal agreement between the meta-evaluators and the metaevaluation results users to prevent any possible disagreements.

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

6. *Existing Information:* Collect all relevant information on the evaluation being meta-evaluated.
7. *New Information:* Collect additional information on the evaluation and the program that was/is being evaluated. Some meta-evaluators have also collected program information and documents from scratch to review the initial data collection process (Hartmann & Loizides, as cited in Cooksy & Caracelli, 2009).
8. *Analysis and Synthesis*: Analyze and synthesize the collected information to determine whether the current standards for each metaevaluation question were met and at what level.
9. *Reaching Conclusions*: Derive a general conclusion from the target evaluation in concert with the predetermined criteria.
10. *Reporting*: Report the detailed metaevaluation results and share them with the relevant stakeholders through oral presentations, webinars, focus group meetings and workshops.
11. *Follow-Up*: Assist other stakeholders to understand and interpret the findings and use the results.

As each metaevaluation varies depending on the context, the metaevaluation steps also vary; however, basically, a metaevaluation involves a re-analysis of one or more of the completed program evaluation steps and a comparison of the results (Patton, 2001).

Study Context

In the Turkish primary school context, science has recently undergone a renewal in accordance with scientific advancements, contemporary issues, and changes in the Turkish education system, paving the way for improvements to the primary school science curriculum. The adoption of the 4 (elementary school) + 4 (middle school) + 4 (high school) compulsory education system in 2013 resulted in revisions of curricula. The salient feature of the 2013 curriculum is that science is a compulsory subject for primary school children beginning in third grade, as opposed to the previous starting point of the fourth grade. As the science curriculum for third graders was developed and implemented for the first time in the Turkish education system, we studied the evaluation of this curriculum.

The 2013 curriculum includes science courses which are taught by classroom teachers and last three hours each week in the third and fourth grades of primary school (Aydın-Ceran, 2021). The vision of the 2013 SC was defined as "to educate all students as scientifically literate individuals" (Ministry of National Education, 2013, p. i). According to the 2013 SC, the teacher acts as a facilitator and guide in the learning and teaching process and students are expected to research, question, explain, and discuss the source of information. The inquiry-based learning strategy is used in the design of learning environments to help students learn science concepts in a lasting and meaningful way. The testing and evaluation approach is based on continuous feedback. It is recommended to use tools and materials that

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

are readily available, affordable, and technology-integrated. Students in the third grade should be familiar with the fundamental concepts of science (MoNE, 2013).

After this third-grade SC was introduced and implemented, a few evaluation studies were conducted to make a judgement about its worth and merits (Gedik, 2017; Güven, 2016; Şentürk, 2017). Of the authors of these evaluation studies, we could access Şentürk to be able to present her evaluation study to the meta-evalutors. Before moving on, it is worth pausing briefly here to look at Şentürk's (2017) evaluation study. Şentürk conducted the evaluation of third-grade SC in her master's thesis, which served as a partial fulfillment for her graduate degree. In Şentürk's evaluation study, a mixed (eclectic) model was employed, whereby more than one model was applied depending on the nature of the research. While the Tyler Model was utilized to determine the students' achievement levels to the course outcomes, Stake's Congruence–Contingency Model was used to evaluate the objectives, content, teaching–learning process, and testing and evaluation aspects that make up the elements of the program. The evaluation included both quantitative and qualitative data, which were obtained from 100 third-grade pupils and 200 elementary school teachers instructing third graders. Now that we have concisely introduced the evaluation study, we can then pick up the purpose of this meta-evaluative effort. The study is based on the following questions:

1. What are the main steps in a metaevaluation?
2. To what extent did the SC evaluation meet the utility standards?
3. To what extent did the SC evaluation meet the feasibility standards?
4. To what extent did the SC evaluation meet the propriety standards?
5. To what extent did the SC evaluation meet the accuracy standards?

Significance of the Study

This metaevaluation of an evaluation that used predetermined standards and information on the metaevaluation process is significant because it is the first comprehensive study on the metaevaluation process. Moreover, a step-by-step metaevaluation procedure was described, which does not exist in the Turkish context and is rare in the international context.

METHODOLOGY

Participants

As meta-evaluators, 20 experts/graduate students who had studied or were studying at the Curriculum and Instruction program at a public university, in Turkey, were invited by email to participate in the metaevaluation. Metaevaluators were selected based on their experiences on evaluation. They had taken at least one program evaluation course and had had experience in conducting evaluations as part of a research project or thesis study because there is neither a public or private entity nor a civil society organization for metaevaluation studies and, thus, no metaevaluation practitioners in Turkey. Eight participants (seven females and

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yildz Teknik Üniversitesi IP Address:193.140.7.203

one male) volunteered to take on the meta-evaluator roles. Three of whom were primary education graduates, two of whom were from English-language teaching, and one each of whom were from computer education and instructional technology, science education, and guidance and psychological counselling. This participant diversity ensured that the meta-evaluators were not too close to the program area under evaluation (Fitzpatrick et al., 2010).

Data Collection Tools and Procedure

Both qualitative and quantitative data were collected. The quantitative data were collected using the metaevaluation checklist, which was developed by transforming the Turkish program evaluation standards suggested in Yüksel's (2010) doctoral dissertation into a checklist, as had been done in numerous other studies (Wingate, 2009). In his adaptation study, the opinions of 138 academicians having doctorate degrees in the field of curriculum and instruction were applied. As a result of the adaptation process, 30 standards and 110 indicators belonging to four basic standards (*utility*, *feasibility*, *propriety*, *accuracy*) promulgated by the JCSEE were reduced to 23 standards and 94 indicators under the umbrella of the same four standards, since some of those were acknowledged as inappropriate for the context of Turkey.

The metaevaluation checklist had four main standards—utility standards, feasibility standards, propriety standards, and accuracy standards—and 94 items that had "yes" and "no" answers based on whether the standards in question existed in the program evaluation or not. These standards also had different subdimensions, as shown in Figure 1.

Yüksel (2010) defined these standards as follows:

The *utility standards* ensure that the evaluation considers the stakeholders' needs and the evaluation objectives. The *feasibility standards* ensure that the evaluation is reasonable, fair, and economical. The *propriety standards* ensure that the evaluation is based on legal, impartial and ethical principles. The *accuracy standards* ensure that the program evaluation findings overlap with the objectives of this evaluation and provide functional information.

Before the quantitative data were collected, the participants were informed about the implementation and reminded to correctly complete the checklist to ensure that the data were reliable. Şentürk presented her thesis for about 30 min, at which time the participants began completing the developed metaevaluation checklist. The meta-evaluators (the eight participants) then questioned Şentürk about her evaluation or the issues that were not explicitly mentioned in her study to ensure that there was maximum interaction and transparency.

While Scriven (2009) felt that checklists were a good idea, he also stated that evaluators needed to think broadly, holistically, and critically about the underlying assumptions rather than seeing metaevaluations as simply filling out a checklist. Thus, after the data obtained from the questionnaire was analyzed, the results were shared with the meta-evaluators, evaluator, and supervisor. They were asked to write a follow-up reflection paper about the metaevaluation results in order to

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

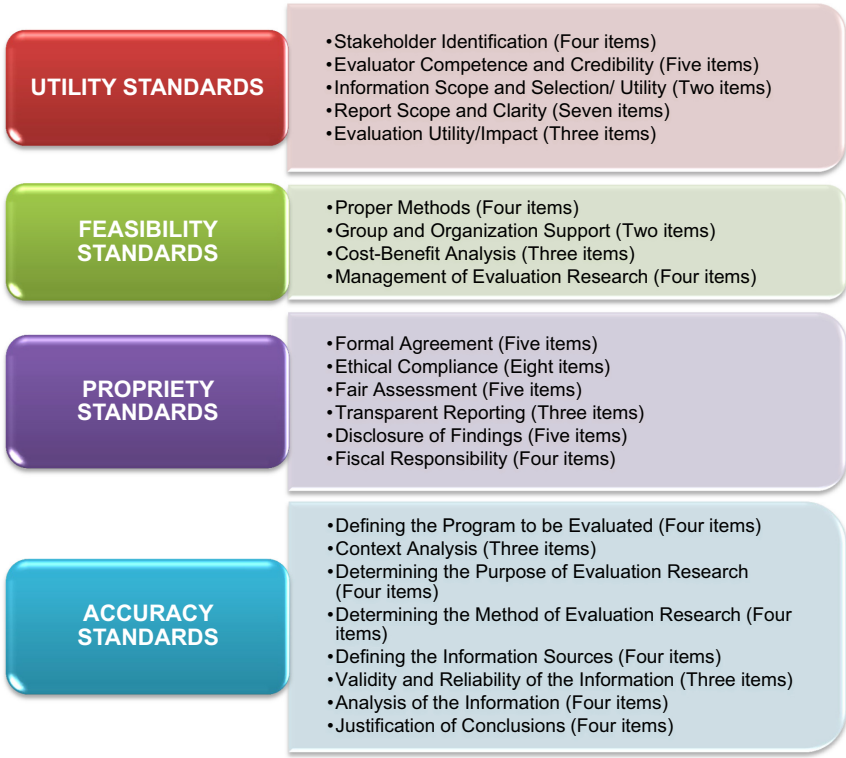

Figure 1. Subdimensions of the Turkish program evaluation standards checklist

validate them and explain their score-setting. The metaevaluation process was conducted based on 7 of Stufflebeam and Coryn's (2014) 11 steps, as follows:

1. Staffing: Eight educational experts were selected as the meta-evaluators. Most of them have graduate degrees in the Curriculum and Instruction Program, and the rest are preparing their thesis in the same field. The common feature of all is that they had taken or were currently studying the Program Evaluation course and participated in an evaluation study during their graduate education. The criteria are important for ensuring that the meta-evaluators have theoretical knowledge and are capable of putting it into practice.
2. Standards: When international metaevaluation studies were examined, Yüksel's (2010) Turkish program evaluation standards, the Turkish version of the most frequently used JCSEE standards, were employed.
3. Questions: The questions in this study were developed to examine the extent to which the SC evaluation conducted by Şentürk (2017) met the utility, feasibility, propriety, and accuracy criteria.

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

   a. Were the evaluation process and products worth meeting the stakeholders' needs?
   b. Was the evaluation procedure carried out effectively and efficiently?
   c. Was the evaluation conducted compatible with legal, unbiased and ethical issues?
   d. Did the evaluation findings serve the purposes of the evaluation?

4. Existing Information: Şentürk's presentation to the meta-evaluators on the SC evaluation process she followed was used as a primary source. The master's thesis was also examined.
5. Analysis and Synthesis: The data collected in this study were analyzed using Stufflebeam's (1999) metaevaluation checklist. At first, the calculations were made on a participant basis based on the responses from each meta-evaluator and the information synthesis step was completed after the general evaluation of the subdimensions for each criterion/standard.
6. Reaching Conclusions: Each participant's overall evaluation of the current utility, feasibility, propriety, and accuracy criteria standards of Şentürk's (2017) SC evaluation reached a common conclusion.
7. Reporting: All metaevaluation process data and the results were reported and presented orally at a conference of experts in the field.
8. Follow-Up: As a continuation of this metaevaluation study, every one of the experts' questions about the research was answered by the researchers to ensure an in-depth understanding and interpretation of the research.

The three steps of Stufflebeam and Coryn's (2014) 11-step metaevaluation process not used in this study were *stakeholder engagement*, *formal agreement*, and *new information*. The reason that the stakeholder engagement step was not employed was to eliminate any disadvantages from when a person was both an evaluator and a meta-evaluator for the same study because the teachers, the main assessment stakeholders, were also the meta-evaluators and because the other stakeholders were not deemed to have the necessary competencies to be meta-evaluators. As this metaevaluation was not for a third party, such as the government, a private institution, or a specific organization, no formal agreement was needed. Stufflebeam and Coryn (2014) state that the acquisition of the new information step was only necessary if needed; however, no additional information on the evaluated curriculum within this study was needed.

Data Analysis and Interpretation

As stated, the collected quantitative data were analyzed by adapting Stufflebeam's (1999) metaevaluation checklist (including the scoring of each subdimension and the general scoring standards) to the model JCSEE standards. Each of the standards in each subdimension were classified as "Excellent," "Very Good," "Good," "Fair," or "Poor" depending on the participants' responses to the checklist. Table

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

1 gives an example of the *Stakeholder Identification* subdimensions for the utility standards, for which the calculation for this classification was as follows.

If a meta-evaluator wrote "yes" for all statements in Table 1, then it was classified as excellent if they wrote "yes" for three, it was classified as very good if they "yes" for two, it was classified as good if they wrote "yes" to one, it was classified as fair, and if they wrote "no" to all, it was classified as poor.

Then the standards (utility, feasibility, propriety, accuracy standards) to which subdimensions belong were calculated using the formulas given in Table 2: four times the number of items was evaluated as "Excellent," three times the number of items was evaluated as "Very Good," two times the number of items was evaluated as "Good," and one times the number of items was evaluated as "Fair." The final value obtained was again interpreted based on the intervals in the second column of Table 2.

The highest score for each standard was when all subdimensions were multiplied by 4. The maximum score that could be achieved under the utility standards

Table 1. An example of the utility standard's subdimension of stakeholder identification

US.1.	Stakeholder Identification	Yes	No
US.1.1.	Are those supporting the program evaluation research identified?		
US.1.2.	Are evaluation experts to plan, implement and evaluate program evaluation research identified?		
US.1.3.	Are the groups or individuals to whom the tools of the program evaluation research are to be implemented defined?		
US.1.4.	Are the groups or individuals to be directly or indirectly affected by the results of the program evaluation research identified?		

• 4 Excellent • 3 Very Good • 2 Good • 1 Fair • 0 Poor

Table 2. Calculation of standards in the evaluation of a meta-evaluator

Standard Scoring	Strength of the Evaluation's Provisions for the Target Standard	
Number of "Excellent" ratings _____ × 4 = _____	93–100	Excellent
Number of "Very Good" _____ × 3 = _____	68–92	Very Good
Number of "Good" _____ × 2 = _____	50–67	Good
Number of "Fair" _____ × 1 = _____	25–49	Fair
Total score = _____	0–24	Poor
Total score (TS) ÷ Maximum score (MS) × 100 = _____; (TS ÷ MS) × 100		

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

Table 3. Calculation of all meta-evaluators' evaluation of standards

Scoring the Standard in Question	Strength of the Evaluation's Provisions for This Standard	
Number of meta-evaluators having chosen the degree of "Excellent" (0–8) ____ × 4 = ____		
Number of meta-evaluators having chosen the degree of "Very Good" (0–8) ____ × 3 = ____	93–100	Excellent
	68–92	Very Good
Number of meta-evaluators having chosen the degree of "Good" (0–8) ____ × 2 = ____	50–67	Good
	25–49	Fair
	0–24	Poor
Number of meta-evaluators having chosen the degree of "Fair" (0–8) ____ × 1 = ____		
Total Score = ____		
Total Score (TS) ÷ Maximum Score (MS) × 100 = _____; (TS ÷ MS) × 100		

in the checklist was 20 as there were five subdimensions. For example, if a meta-evaluator judged two out of the five utility standard subdimensions as "excellent" (2 × 4 = 8), two as "good" (2 × 2 = 4), and one as "fair" (1 × 1 = 1), the total score for that meta-evaluator would be 13 (8 + 4 + 1). The utility standard was then calculated as (13 ÷ 20) × 100 = 65 using this formula (TS ÷ MS) × 100, which meant that overall the meta-evaluator assessed the utility standard as "Good" based on the scale in the left column of Table 2.

After each meta-evaluator evaluated the four standards as excellent, very good, good, fair, and poor, the general evaluations from the eight meta-evaluators for each standard were calculated according to Table 3 using the calculations from Stufflebeam's (1999) "Metaevaluation Checklist."

The quantitative data are supported and validated using verbatim quotations from the meta-evaluators', the evaluator's, and the supervisor's reflection papers.

FINDINGS

As the application of the metaevaluation steps were reported in the Introduction and Methods sections, they are not repeated here.

The findings from the metaevaluation of Şentürk's (2017) evaluation of the third-grade SC are presented in the following and the SC evaluation for the utility standard, the first of the program evaluation standards, is shown in Table 4.

As shown in Table 4, one of the eight meta-evaluators evaluated the utility standard as "Excellent," six evaluated the utility standard as "Very Good," and

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

Table 4. The SC evaluation's level of meeting the utility standard

Evaluation for Utility Standards		
Meta-Evaluator	**Evaluation Score**	**Evaluation**
1	90	Very Good
2	70	Very Good
3	70	Very Good
4	65	Good
5	70	Very Good
6	85	Very Good
7	70	Very Good
8	95	Excellent

one evaluated the utility standard as "Good." Therefore, the compilation calculations from the eight meta-evaluators shown in Table 3 [$(1 \times 4) + (6 \times 3) + (1 \times 2) = 24$; $(24 \div 32) \times 100 = 75$] evaluated the overall utility standard as "**Very Good**" (75).

The *Stakeholder Identification* and *Information Scope and Selection/Utility* subdimensions were evaluated as "Excellent" by all participants, the *Evaluator Competence and Credibility* subdimension was evaluated by one participant as "Very Good" but "Excellent" by the others, five participants evaluated the *Report Scope and Clarity* subdimension as "Good," two evaluated it as "Very Good" and one evaluated it as "Excellent," and seven participants evaluated the *Evaluation Utility/Impact* subdimension as "Good" and one evaluated it as "Excellent."

These quantitative results were validated by the meta-evaluators, the evaluator, and the supervisor. A meta-evaluator stated that the academic training of the evaluator positively affected the utility standard:

> The competence and credibility of the evaluator was found to be high since she took the academic training on program evaluation. In addition, this situation ensured that the program evaluation research was carried out in line with the evaluation objectives and the needs of the stakeholders.

In comparison, the supervisor of the thesis thought that the fact that the subdimensions of the utility standard and the content of a master's thesis are very similar causes this standard to be evaluated as "Very Good":

> The dimensions covered within the scope of the utility standard also constitute the content of a master's thesis. That is, in the content of each master's thesis, both the stakeholders and the quality of the evaluators are defined, and why these evaluators are preferred, etc. In addition, the findings and conclusion part of the evaluation report must be clear and understandable since they are checked by a supervisor and three jury members, otherwise the thesis will not be approved.

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

The SC evaluation's for the feasibility standard, the second of the program evaluation standards, is shown in Table 5, from which it can be seen that one meta-evaluator evaluated the standard as "Very Good," five evaluated it as "Good" and two evaluated it as "Fair", with the compilation result $[(1 \times 3) + (5 \times 2) + (1 \times 1) = 14; (14 \div 32) \times 100 = 43.75]$ being "**Fair**."

The *Cost–Benefit Analysis* subdimensions of this standard were evaluated as "Fair" by all participants, and the *Management of Evaluation Research* subdimension was evaluated as "Excellent" by all participants. The *Proper Methods* subdimension was evaluated as "Good" by five participants, "Very Good" by two participants, and "Fair" by one participant, and the *Group and Organization Support* subdimension was evaluated as "Good" by five participants, "Poor" by two participants, and "Excellent" by one participant.

One of the meta-evaluators explained the rationale for the evaluation of the *Cost–Benefit Analysis* subdimension as "Fair" as follows:

> Since no funds are provided for program evaluation studies in Turkey, this standard was evaluated as fair due to the deficiencies in the cost–benefit analysis subdimension. The evaluator carried out this study with academic motivation and in line with her own possibilities and limitations.

Similarly, the evaluator confirmed the quantitative results for this standard as follows:

> Educational institutions in Turkey use the curriculum prepared by the Ministry of National Education. For this reason, there is no second Science Curriculum with which the aforementioned program can be compared. A comparison is not made; therefore, it is not decided which one is a less costly and more effective program.

The SC evaluation for the propriety standard, the third of the program evaluation standards, is shown in Table 6, from which it can be seen that it was evaluated

Table 5. The SC evaluation's level of meeting the feasibility standard

Evaluation for Feasibility Standards		
Meta-Evaluator	**Evaluation Score**	**Evaluation**
1	43.75	Fair
2	62.50	Good
3	56.25	Good
4	56.25	Good
5	56.25	Good
6	43.75	Fair
7	75	Very Good
8	50	Good

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

Table 6. The SC evaluation's level of meeting the propriety standard

Evaluation for Propriety Standards		
Meta-Evaluator	**Evaluation Score**	**Evaluation**
1	45.83	Fair
2	41.66	Fair
3	41.66	Fair
4	45.83	Fair
5	41.66	Fair
6	41.66	Fair
7	41.66	Fair
8	41.66	Fair

as "Fair" by all meta-evaluators, with the compilation result [$(8 \times 1) = 8$; $(8 \div 32) \times 100 = 25$] also being "**Fair**."

The *Formal Agreement* and *Fiscal Responsibility* subdimensions were evaluated as "Poor" by all participants, the *Transparent Reporting* subdimension was evaluated as "Excellent" by all participants. The *Ethical Compliance* subdimension was evaluated as "Very Good" by seven participants and "Excellent" by one, the *Fair Assessment* subdimension was evaluated as "Very Good" by five participants and "Excellent" by 3, and the *Disclosure of Findings* subdimension was evaluated as "Poor" by seven participants and "Good" by one.

One of the meta-evaluators expressed the evaluation of the *Formal Agreement* and *Fiscal Responsibility* subdimension as "Poor" as follows:

> In our current system, there are no formal contracts or financial compliance procedures when evaluating curricula.

Another meta-evaluator explained the reason for the evaluation of *Disclosure of Findings* subdimension as "Poor" as follows:

> [I]t is related to the fact that although the findings are discussed with the field experts in the thesis defense procedure and in a congress, the decision makers are not reached, and the results are not presented. On the other hand, getting into touch with decision makers and presenting research findings in Turkey is a very difficult and tough process.

The reason for the positive findings in the other subdimensions of the propriety standard was stated by the supervisor as follows:

> [D]ue to the fact that they are issues that are followed meticulously in a graduate study. For example, Ethical Compliance is an issue that is too sensitive to be considered and it is carefully followed from the first word to the last word of the study. Otherwise, failure to do so has serious consequences. Therefore, it is quite understandable that this part was evaluated as very good.

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yildz Teknik Üniversitesi IP Address:193.140.7.203

The SC evaluation for the accuracy standard, the fourth of the program evaluation standards, is shown in Table 7, from which it can be seen that it was evaluated by two of the meta-evaluators as "Excellent," with the others evaluating it as "Very Good"; therefore, the compilation [$(2 \times 4) + (6 \times 3) = 26$; $(26 \div 32) \times 100 = 81.25$] score for the accuracy standard was "**Very Good**" as shown in Table 3.

Identifying the Program to be Evaluated, *Defining the Information Sources*, and *Analysis of the Information* subdimensions were all evaluated as "Excellent" by all participants. The *Context Analysis* subdimension was evaluated as "Good" by five participants, "Excellent" by two, and "Fair" by one; the *Determining the Purpose of Evaluation Research* subdimension was evaluated as "Excellent" by six participants and "Very Good" by two; the *Determining the Method of Evaluation Research* was evaluated as "Excellent" by six participants and "Very Good" by two; the *Validity and Reliability of the Information* subdimension was evaluated as "Good" by six participants and "Excellent" by two; and the *Justification of Conclusions* subdimension was evaluated as "Excellent" by half the participants and "Very Good" by the other half.

The opinions of a meta-evaluator regarding getting positive findings related to the accuracy standard are as follows:

> I think that making the whole process transparent and meticulous, from defining the program to be evaluated to analyzing the data and justifying the results, has positive contributions in the evaluation of the accuracy standard.

In this regard, the supervisor made the following statement:

> The fact that all the components of the accuracy standard are also indispensable elements of a thesis study may be among the reasons why the evaluation of this standard is very high (excellent-very good).

Table 7. The SC evaluation's level of meeting the accuracy standard

Evaluation for Accuracy Standards		
Meta-Evaluator	**Evaluation Score**	**Evaluation**
1	81.25	Very Good
2	93.75	Excellent
3	84.37	Very Good
4	87.50	Very Good
5	81.25	Very Good
6	81.25	Very Good
7	84.37	Very Good
8	96.87	Excellent

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

CONCLUSION AND DISCUSSION

This research, which could be classified as an external and summative metaevaluation, was conducted to demonstrate the process of reviewing a completed evaluation using evaluation standards. A metaevaluation of each evaluation should be conducted as it can inform evaluators of the standards that must be followed for the successful execution of an evaluation and provide a resource for assessing the evaluation quality from an objective perspective. Metaevaluations provide detailed and objective measures that contribute to program evaluation accountability and justifiability. Morris (2007) commented that metaevaluations were powerful mechanisms that ethically improved evaluation quality because no matter who conducted the evaluation, whether it be an independent evaluator or the evaluator themselves, metaevaluations provide moral protection against critical review.

It was concluded that Stufflebeam and Coryn's (2014) 11 metaevaluation steps could not be wholly applied to this study. The *formal agreement*, *stakeholder engagement*, and *new information* steps were not employed. As Stufflebeam (2001) pointed out, these steps were basically general processes to be considered when conducting metaevaluations and therefore did not need to be applied to every metaevaluation, and if needed, new steps could be added.

This study also examines Şentürk's (2017) SC evaluation using Turkish Evaluation Standards adapted from JCSEE to the Turkish context in order to highlight the evaluation's strengths and weaknesses. The metaevaluation came to a conclusion about the evaluation's merit and worth. The results of the metaevaluation of Şentürk's (2017) SC evaluation for the *utility standard* was "Very Good," with its *Stakeholder Identification*, *Evaluator Competence and Credibility*, *Information Scope and Selection/Utility*, and *Evaluation Utility/Impact* subdimensions considered to be applicable to Turkey. Since this program evaluation study was conducted by a graduate student for her master's thesis, the evaluation objectives were discreetly specified as the supervisor and metaevaluators mentioned. This may account for the favourable results of the Utility standard. The subdimensions for *Report Scope and Clarity*, had lowers score than the other subdimensions, possibly because the summary report, main report, and technical report were not separately prepared, the meta-evaluators' purposes were academic, and the evaluation was not presented by an independent evaluator, which meant that it was not clear whether a new stakeholder perspective was developed. The results are extremely close to Patton's (2012) metaevaluation of the evaluation of the Paris Declaration, particularly the subdimensions of *Evaluator Competence* and *Credibility*, despite the fact that he utilized different, but similar, standards.

The *feasibility standard* was judged as "Fair," possibly because *Cost–Benefit Analyses* are not generally considered for academic evaluation studies in Turkey, and there is no *Group and Organization Support*, which has been commented on in the past (Daloğlu, 1996; Kuru, 1987) and in more recent studies (Berk, 2018b; Türk, 2019). However, the evaluator clarified this "Fair" evaluation by stressing the absence of a second science curriculum to compare with, as well. However, the

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yildz Teknik Üniversitesi IP Address:193.140.7.203

"Excellent" and "Very Good" evaluations for the *Proper Methods* and *Management of Evaluation* subdimensions indicated that these elements were meticulously followed in academic evaluations. In a research-based evaluation, feasibility standards are more likely related to methodology. Jacob and Desautels's (2014) findings on methodology criteria to be mostly quite high (80+%) as in ours; however, some subdimensions were very low (33%).

The *Propriety Standards* were also only evaluated as "Fair" as the evaluation was inadequate because of the low evaluations for the *Formal Agreement* and *Fiscal Responsibility* subdimensions. This was because formal evaluation environments are not supported by state, organizations/institutions, program users or sponsor stakeholders in Turkey. The *Disclosure of Findings* subdimension was also insufficient as evaluation results are generally not discussed with stakeholders such as field experts and decision makers except at conferences and symposiums that are usually attended for academic reasons. One meta-evalutor also mentioned about the difficulty of contacting decision makers in Turkey to deliver study findings. However, ethical compliance, fair evaluation and transparent evaluation reporting were found to meet the standards, with ethical compliance in particular being emphasized in all research and especially in evaluation studies as it is a legal and institutional obligation (ÜAK Code of Ethics, 2019). The supervisor stated that the process was carried out in a very sensitive manner in terms of ethics due to the fact that the study was a thesis study, an official study. Sanders (1999) who conducted a metaevaluation of a single program using the same JCSEE standards as in this study reported that some subdimensions of propriety standard were "Very Good" and others were "Fair," which is consistent with the findings of our study. Also, a study by Perry (2008) yielded similar findings with ours.

The *accuracy standards* were found to be the strongest aspect of this evaluation because of the diverse range of detailed data and information sources and the data analysis techniques used to assess whether the functional information matched the program objectives. Similarly, qualitative findings supported this result that the entire procedure was methodically and openly carried out, and the subdimensions of the accuracy standard match with requirements of a thesis study. As Gökmenoğlu (2015) emphasized, the accuracy standard subdimensions are indispensable elements in all evaluations. Lynch et al. (2003) found results that were so similar to ours since all the accuracy standard subdimensions were excellent.

In sum, Şentürk's evaluation generally corresponded with the metaevaluation checklist standards. Therefore, Turkish MoNE may consider the findings of this evaluation when improving the SC curriculum. Turkish elementary school teachers, who implement SC at the third level, also can benefit from Şentürk's evaluation results when deciding on their curriculum approach, such as curriculum fidelity, adaptation and enactment. However, Şentürk's evaluation did not match with some subdimensions of the checklist, including cost–benefit analysis, formal agreement, fiscal responsibility and disclosure of findings because of the common practice of the evaluation studies in Turkey. Since recent evaluation studies in

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

Turkey have not received funding from any organization or institution, evaluation studies typically exclude signing protocols between the parties; nevertheless, there are few studies that did (Berk, 2012). It is possible to suggest that evaluation studies of researchers should be supported fiscally by MoNE and its institutions or any other private entities and formal agreements should be made in order to conform to evaluation principles and procedure. Meetings, where academics and researchers present their research findings, should be welcomed by those who influence educational policy in Turkey. Thus, the evaluation area in Turkey may transform to comply with the internationally established standards.

This metaevaluation is contextually limited to an evaluation of the MoNE's third-grade SC carried out in the Turkish primary school system. Although dealing with a particular curriculum may limit representativeness and generalizability, it is valuable for gathering in-depth data and reaching more accurate metaevaluation results regarding the evaluation of SC. However, this pioneering study can be regarded as noteworthy since it provides a broad overview and is the first endeavour in investigating how a metaevaluation study is conducted in the context of Turkey. As an implication, a private or government agency can be established to conduct professional metaevaluation studies in Turkey. Besides, given the scarcity of metaevaluation research being conducted in other countries, both the quality and quantity of organizations supporting metaevalutive acts should be boosted, as well.

Another limitation is related to the selection of meta-evaluators. We invited people who are competent in terms of both evaluation theories and practices, but we only worked with those who were accessible. Some certificate programs may be opened for training program evaluators/meta-evaluators according to professionalism standards. This study is also limited in that stakeholder engagement was omitted into the metaevaluation process; thus, it is recommended that stakeholder participation be included in future research to assist in the establishment of new evaluation standards for different needs and to identify additional research questions. The deviations and inaccuracies in future evaluation studies in Turkey could be prevented by applying the evaluation standards as a control mechanism before, during, and after the program evaluation.

NOTE

1 The findings of this study were presented orally on the 12th of October 2019 in the 7th International Congress on Curriculum and Instruction, in Ankara, Turkey.

REFERENCES

Alkin, M.C. (Ed.). (2012). *Evaluation roots: A wider perspective of theorists' views and influences*. Sage.

Association for Evaluation and Accreditation of Teacher Education Programs. (2016). *EPDAD Teacher education undergraduate programs standards*. https://epdad.org.tr/data/genel/pdf/standartlar.pdf

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

Association for Evaluation and Accreditation of Teacher Education Programs. (2021). *EPDAD Teacher education undergraduate programs standards guide*. https://epdad.org.tr/data/genel/pdf/EPDAD_Standartlar_Surum_1.1_Kilavuz.pdf

Aydın-Ceran, S. (2021). Öğretim yöntemlerine dayanan fen eğitimi araştırmalarında güncel eğilimler: İlkokul düzeyinde bir analiz [Current trends in science education research based on teaching methods: A primary school level analysis]. *Journal of Individual Differences in Education*, *3*(2), 113–131. https://doi.org/10.47156/jide.1026165

Berk, Ş. (2012). *Evaluation of the modular system implemented in vocational and technical secondary schools by using Provus' Discrepancy Model*. Doctoral dissertation, Anadolu University. Available at Turkish National Thesis Center.

Berk, Ş. (2018a). *Modüler sistemin değerlendirilmesi: Modüler sistem, program değerlendirme, mesleki ve teknik eğitim* [Evaluation of modular system: Modular system, program evaluation, vocational and technical education]. Pegem Akademi.

Berk, Ş. (2018b). Assessment of public schools' out-of-school time academic support programs with participant-oriented evaluation. *Journal of Education and Learning*, *7*(3), 159–175. https://doi.org/10.5539/jel.v7n3p159

Bobin, K. (2017). *Meta-evaluation: A synthesis of evaluation studies 2005–2016*(Evaluation Series No. 104). The Commonwealth Secretariat. http://meb.ai/JhU2sY

Bourgeois, I., & Whynot, J. (2018). Strategic evaluation utilization in the Canadian federal government. *Canadian Journal of Program Evaluation*, *32*(3). https://doi.org/10.3138/cjpe.43179

Cooksy, L. J., & Caracelli, V.J. (2009). Metaevaluation in practice: Selection and application of criteria. *Journal of MultiDisciplinary Evaluation*, *6*(11), 1–15. https://doi.org/10.56645/jmde.v6i11.211

Daloğlu, A. (1996). *A case study on evaluating the certificate for overseas teachers of English curriculum at Bilkent University*. [Unpublished doctoral dissertation, Middle East Technical University]. Available at Turkish National Thesis Center.

Demirel, Ö. (2017). *Eğitimde program geliştirme: Kuramdan uygulamaya* [Curriculum development: From theory to practice]. Ankara: Pegem A Yayıncılık.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2010). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Pearson.

Gardner, D. (2019). *Beyond the four levels: An evaluation model for growth and sustainability* [Doctoral dissertation, Wayne State University]. Available at ProQuest Dissertations & Theses Global.

Gedik, N. B. (2017). *3. Sınıf fen bilimleri dersi öğretim programının öğretmen görüşlerine dayalı olarak değerlendirilmesi* [Evaluation of elementary school third grade science curriculum based on teachers' views; Unpublished master's thesis, Adıyaman Univesity]. Available at Turkish National Thesis Center.

Gökmenoğlu, T. (2015). The wide angle: Program evaluation studies in Turkey in terms of models and approaches. *International Journal of Curriculum and Instructional Studies*, *4*(7), 55–70.

Güven, G. (2016). *3. sınıf fen bilimleri dersi öğretim programına ilişkin öğretmen görüşleri* [The opinions of teachers about the 3rd class science curriculum; Unpublished master's thesis, Mustafa Kemal University]. Available at Turkish National Thesis Center.

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

Jacob, S., & Desautels, G. (2014). Assessing the quality of Aboriginal program evaluations. *Canadian Journal of Program Evaluation*, *295*(1), 62–86. https://doi.org/10.3138/cjpe.29.1.62

Kürüm-Yapıcıoğlu, D., Atik-Kara, D., & Sever, D. (2016). Türkiye'de program değerlendirme çalışmalarında eğilimler ve sorunlar: Alan uzmanlarının gözüyle [Trends and problems in curriculum evaluation studies in Turkey: The perspective of domain experts]. *Uluslararası Eğitim Programları ve Öğretim Çalışmaları Dergisi*, *6*(12), 91–113.

Kuru, S. (1987). *Mesleki eğitim fakültesi giyim endüstrisi ve giyim eğitimi bölümü "giyim teknikleri ve üretimi" dersinin değerlendirilmesi* [Evaluation of the course "clothing techniques and production" in the department of clothing industry and clothing education, faculty of vocational education; Unpublished master's thesis, Gazi University]. Available at Turkish National Thesis Center.

Leeuw, F.L., & Cooksy, L.J. (2005). Evaluating the performance of development agencies: The role of metaevaluations. In G. K. Pitman, O. N. Feinstein, & G. K. Ingram (Eds.), *Evaluating development effectiveness* (World Bank Series on Evaluation and Development, Vol. 7, pp. 95–108). Transaction.

Lynch, D. C., Greer, A. G., Larson, L. C., Cummings, D. M., Harriett, B. S., Dreyfus, K. S., & Clay, M. C. (2003). Descriptive metaevaluation: Case study of an interdisciplinary curriculum. *Evaluation & the Health Professions*, *26*(4), 447–461. https://doi.org/10.1177/0163278703258099

Ministry of National Education. (2013). *İlköğretim kurumları (ilkokullar ve ortaokullar) fen bilimleri dersi (3, 4, 5, 6, 7 ve 8. sınıflar) öğretim programı* [Science curriculum (primary and middle schools) (grades 3, 4, 5, 6, 7 and 8)].

Morris, M. (Ed.). (2007). *Evaluation ethics for best practice: Cases and commentaries*. Guilford Press.

Organisation for Economic Co-operation and Development. (2011). *Strengthening accountability in aid for trade. The development dimension*. OECD Publishing. https://doi.org/10.1787/9789264123212-en

Özdemir, S. M. (2009). Eğitimde program değerlendirme ve Türkiye'de eğitim programlarını değerlendirme çalışmalarının incelenmesi [Curriculum evaluation and examination of curriculum evaluation studies in Turkey]. *Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi, 6*(2), 126–149.

Patton, M. Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation*, *22*(3), 329–336. https://doi.org/10.1016/s1098-2140(01)00147-3

Patton, M. Q. (2012). Meta-evaluation: Evaluating the evaluation of the Paris declaration. *The Canadian Journal of Program Evaluation*, *27*(3), 147–171.

Patton, M. Q. (2017). *Principles-focused evaluation: The guide*. Guilford Publications.

Patton, M. Q. (2018). Evaluation science. *American Journal of Evaluation*, *39*(2), 183–200. https://doi.org/10.1177/1098214018763121

Perry, K. M. (2008). A reaction to and mental metaevaluation of the experiential learning evaluation project. American Journal of Evaluation, 29(3), 352–357. https://doi.org/10.1177/1098214008321686

Russ-Eft, D., & Preskill, H. (2008). Improving the quality of evaluation participation: A meta-evaluation. *Human Resource Development International*, *11*(1), 35–50. https://doi.org/10.1080/13678860701782311

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yıldız Teknik Üniversitesi IP Address:193.140.7.203

Sanders, J. R. (1999). Metaevaluation of "the effectiveness of comprehensive, case management interventions: Evidence from the National Evaluation of the Comprehensive Child Development Program." *American Journal of Evaluation*, *20*(3), 577–582. https://doi.org/10.1016/s1098-2140(99)00043-0

Scriven, M. (1969). An introduction to meta-evaluation. *Educational Product Report*, *2*(5), 36–38.

Scriven, M. (1975). *Evaluation bias and its control*. The Evaluation Center, Western Michigan University.

Scriven, M. (2009). Meta-evaluation revisited. *Journal of Multidisciplinary Evaluation*, *6*(11), 3–8. https://doi.org/10.56645/jmde.v6i11.220

Şentürk, Ö. (2017). *İlkokul 3. sınıf fen bilimleri dersi öğretim programı'nın değerlendirilmesi* [An evaluation of the third-grade science curriculum in elementary school; Unpublished master's thesis, Marmara University]. Available at Turkish National Thesis Center.

Şentürk, Ö., & Berk, Ş. (2019). İlkokul 3. sınıf fen bilimleri dersi öğretim programının değerlendirilmesi [Evaluation of the 3rd grade science curriculum in primary schools]. *Marmara Üniversitesi Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi*, *49*(49), 144–166.

Snow, D. (2001). Communicating quality. In A. Benson, D. M. Hinn, & C. Lloyd (Eds.), *Visions of quality: How evaluators define, understand and represent program quality* (pp. 29–42). JAI Press.

Stufflebeam, D. L. (1999). *Program evaluations metaevaluation checklist* (Evaluation Checklists Project). National Science Foundation (NSF) & The Evaluation Center, Western Michigan University. http://meb.ai/fI2tIF

Stufflebeam, D. L. (2000). The methodology of metaevaluation as reflected in metaevaluations by the Western Michigan University Evaluation Center. *Journal of Personnel Evaluation in Education*, *14*(1), 95–125. https://doi.org/10.1023/a:1008198315521

Stufflebeam, D. L. (2001). The metaevaluation imperative. *American Journal of Evaluation*, *22*(2), 183–209. https://doi.org/10.1016/s1098-2140(01)00127-8

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

Stufflebeam, D. L., & Coryn, C. L. (2014). *Evaluation theory, models, and applications*. John Wiley & Sons.

Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. Jossey-Bass.

Türk, N. (2019). *Design, implementation and evaluation of science, technology, engineering and mathematics (STEM) curriculum for undergraduate programs of faculty of education*. [Unpublished doctoral dissertation, Gazi University]. Available at Turkish National Thesis Center.

ÜAK Code of Ethics. (2019). *Üniversitelerarası kurul bilimsel araştırma ve yayın etiği yönergesi* [Interuniversity council directive on scientific research and publication ethics]. http://meb.ai/fQT2sN

Wingate, L. A. (2009). *The program evaluation standards applied for metaevaluation purposes: Investigating interrater reliability and implications for use*. [Doctoral dissertation, Western Michigan University]. Available at ProQuest Dissertations & Theses Global.

Yarbrough, D. B. (2017). Developing the program evaluation utility standards: Scholarly foundations and collaborative processes. *Canadian Journal of Program Evaluation*, *31*(3), 284–304. https://doi.org/10.3138/cjpe.349

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). Corwin Press.

Yüksel, İ. (2010). *Türkiye için program değerlendirme standartları oluşturma çalışması* [Development of Turkish program evaluation standards; Unpublished doctoral dissertation, Anadolu University]. Available at Turkish National Thesis Center.

AUTHOR INFORMATION

Esra Kerimoğlu is a PhD student and research assistant in the Department of Curriculum & Instruction at Yildiz Technical University. She holds a Master's degree in Curriculum & Instruction from Marmara University. Her research interests are program evaluation, curriculum studies, learning & teaching processes, teacher education, and language teaching. ORCID: https://orcid.org/0000-0002-7004-3175

M. N. Öykü Ülker is a PhD student in the Department of History of Science and Technology at Istanbul Technical University and a principal at a public middle school in Istanbul. She holds bachelor's degrees in English Language Teaching, Sociology and Social Services, and master's degree in Curriculum & Instruction. She also works as a design and production instructor within The Turkish Technology Team Foundation. Her research interests include program evaluation, program development, second and foreign language teaching, teachers' professional competencies, psychosocial reflections of advances in science and technology, and interdisciplinary studies. ORCID: https://orcid.org/0000-0003-0704-1721

Şaban Berk is an associate professor at the University of Marmara, Department of Educational Sciences, Curriculum & Instruction. His research interests are program evaluation, teacher training, research methodology, and project management. ORCID: https://orcid.org/0000-0002-6821-5249

https://utpjournals.press/doi/pdf/10.3138/cjpe.71619 - Thursday, November 02, 2023 2:36:51 AM - Yildz Teknik Üniversitesi IP Address:193.140.7.203