EĞİTİMDE ÖLÇME VE DEĞERLENDİRME DERSİNE YÖNELİK BİR META-DEĞERLENDİRME ÇALIŞMASI*

A META-EVALUATION RESEARCH ON MEASUREMENT AND EVALUATION COURSE

Muhammed Akıncı[1], Erdoğan Köse[2]

ÖZ: Bu çalışmada, *Eğitimde Ölçme ve Değerlendirme Dersi Programının Değerlendirilmesi* başlıklı çalışmanın meta-değerlendirmesinin yapılması amaçlanmıştır. Bu amaç doğrultusunda çalışmanın Eğitimde Değerlendirme Standartları Ortak Komitesi (JCSEE) tarafından belirlenen Program Değerlendirme Standartlarını ne ölçüde karşıladığı ortaya koyulmuştur. Araştırmada yöntemsel olarak meta-değerlendirme prosedürleri benimsenmiş, bu bağlamda gerçekleştirilen işlemlerde JCSEE tarafından belirlenen *Değerlendirmede Hesap Verebilirlik* standartları dikkate alınmıştır. Meta-değerlendirmesi yapılan araştırmanın gerekli standartları taşıyıp taşımadığının kontrolünde ise araştırmacılar tarafından geliştirilen Program Değerlendirme Standartları Kontrol Listesi kullanılmıştır. Bu amaçla bir iç, bir de dış değerlendirici tarafından çalışmanın etkin bir program değerlendirme araştırmasında olması gereken doğru planlama, geçerli ve güvenilir veri toplanma, gerekli analizleri yapma ve raporlaştırma niteliklerini taşıyıp taşımadığına yönelik kontroller yapılmıştır. İç ve dış değerlendiricilerden elde edilen verilere göre bu araştırma JCSEE tarafından belirlenen Program Değerlendirme Standartları'nı büyük oranda (%88,33) karşılamaktadır. Ayrıca her iki değerlendiriciye göre *Amaca Uygunluk Standartları* %88,9, *Fizibilite Standartları* %75, *Doğruluk Standartları* %100 oranında karşılanırken, *Uygunluk Standartları* dış değerlendiriciye göre %87,5, iç değerlendiriciye göre ise %75 oranında karşılanmaktadır. Bunun yanında uzmanlar araştırmada özellikle kültürel değerler, çıkarlar, bağlamlar ve çıkar çatışmaları kavramları hakkında yeterli bilgi sağlanmadığı görüşündedir. Araştırma bulgularından yola çıkarak program değerlendirme ve meta-değerlendirme çalışmalarının daha etkin bir biçimde yürütülmesine yönelik öneriler sunulmuştur.

Anahtar sözcükler: Meta-değerlendirme, ölçme, değerlendirme, program değerlendirme, program standartları

ABSTRACT: This study aimed to make a meta-evaluation of the study titled *Evaluation of Measurement and Evaluation in Education Curriculum*. For this purpose, the study was conducted to determine the extent to which the research complies with the Program Evaluation Standards determined by the Joint Committee on Standards for Educational Evaluation (JCSEE). Methodologically, meta-evaluation procedures were adopted in the research, and the *Evaluation Accountability Standards* determined by JCSEE were considered in the transactions carried out in this context. The Program Evaluation Standards Checklist developed by the researchers was used to control whether the meta-evaluated research met the required standards. For this purpose, controls were made by an internal and an external evaluator to reveal whether the study has the qualities of accurate planning, valid and reliable data collection, performing the necessary analyzes, and reporting, which should be in an effective program evaluation research. The results of the meta-evaluation indicate that the study under examination demonstrates a high degree of alignment with the Program Evaluation Standards used by internal and external evaluators, with a concordance rate of 88.33%. In addition, according to both evaluators, Utility Standards were met by 88.9%, Feasibility Standards by 75%, and Accuracy Standards by 100%, while Propriety Standards were met at a rate of 87.5% according to the external evaluator and 75% according to the internal evaluator. Moreover, the experts think that more information should have been provided in the research, especially about the concepts of cultural values, interests, contexts, and conflicts of interest. Based on the research findings, suggestions were made to conduct program evaluation and meta-evaluation studies more effectively.

Keywords: Meta-Evaluation, measurement, evaluation, program evaluation, program standard

EXTENDED ABSTRACT

Introduction

Measurement and evaluation, essential elements in interpreting and deciding on the quality of education in Turkey, have also become tools used at the point of transition between education levels, admission to employment, and educational institutions (Baykul, 2011). Measurement and evaluation are among the essential teacher qualities established by the Ministry of National Education (MNE), as well as one of the primary concerns that should be emphasized for the education system's performance (MNE, 2017a; MNE, 2017b; MNE, 2017b). Because measurement and evaluation are primarily concerned with student and teacher achievement, systematic evaluation of a school system is only achievable through program evaluation studies (Nevo, 2002). It is necessary to utilize scientific research methods and formal evaluations that use well-structured procedures toward specific goals, locations, timeframes, and intended audiences to ensure the accuracy and practicality of evaluations of the effectiveness and potential of educational programs (Fitzpatrick et al., 2011; Uşun, 2012; Yüksel, & Sağlam, 2014). These program evaluations are comprehensive and costly and are referred to as Improvement/Accountability Oriented Approaches (Stufflebeam, 1999). This issue shows that there may be considerable effort, time, and financial problems if the program evaluation procedures are not carried out accurately. This situation raises concerns about the effectiveness of the program evaluation processes. From this point of view, meta-evaluations reveal the quality of program evaluation processes. Meta-evaluation also reveals the deficiencies and errors of one or more program evaluation studies (Cooksy & Caracelli, 2009; Scriven, 2011). These contributions of meta-evaluation to the program evaluation process and the strong relationship between teacher training, measurement, evaluation, and curriculum evaluation make this study meaningful. In this context, it was deemed necessary to conduct a meta-evaluation of a study aiming to reveal the quality of a technical course, such as measurement and evaluation in education, a compulsory course in teacher training programs.

Method

Methodologically, meta-evaluation procedures were adopted in the research, and the evaluation accountability standards determined by JCSEE were considered in the transactions carried out in this context. The Program Evaluation Standards Checklist developed by the researchers was used to control whether the meta-evaluated research met the required standards. For this purpose, an internal and an external evaluator made controls to reveal whether the study has the qualities of accurate planning, valid and reliable data collection, performing the necessary analyzes, and reporting, which should be in an effective program evaluation research. In addition, what actions were taken for each standard item during the program evaluation process was presented to the evaluators.

Findings

According to data collected from internal and external evaluators, the meta-evaluated study (88.33%) complies with the JCSEE-established Program Evaluation Standards. Furthermore, both evaluators found that the study demonstrated a high degree of conformity with the Utility Standards (88.9%), Feasibility Standards (75%), and Accuracy Standards (100%). The external evaluator determined that the study also met the Propriety Standards at a rate of 87.5%, while the internal evaluator found a rate of 75% conformity with these standards. Moreover, the experts believe that sufficient information should have been provided in the research, especially about the concepts of cultural values, interests, contexts, and conflicts of interest.

Discussion and Conclusion

When the research findings are considered in the context of utility standards, it was observed that the study primarily serves its purpose. However, it was stated that there are problems in terms of quality and quantity in serving the purpose of program evaluation research in Turkey. Akıncı & Köse (2021) stated in their study that both the evaluated programs and program evaluation studies in Turkey have problems in serving their purpose. The small number of program evaluation studies conducted in Turkey has also been highlighted in some other research (Gündüz & Kuzu Demir, 2020; Keskin & Yazar, 2019). This situation may be an indicator that we should conduct more program evaluation studies that evaluate different program types and serve their purpose. At this point, the risks posed by program evaluation research in terms of time, workload, and cost should not be forgotten. The preference for practical, economical, sound

and functional methods to increase the effectiveness and efficiency of the evaluation is related to the feasibility standards (JCSEE, 2018). In this context, the meta-evaluated research has the most problems in the relevant standard field. The basis of this problem may lie in the fact that the methodological choices are inappropriate, especially in terms of practicality. This case is more about research culture. According to Akcan et al. (2018), the research culture in Turkey is education-focused, and there are issues with the research approach in this regard.

When the research findings are examined in terms of accuracy standards, the study, according to internal and external evaluators the study meets this standard 100%. This finding is a positive and expected finding for a study that evaluates the program of a technical course that includes psychometric features such as measurement and evaluation in education. However, this is only sometimes a situation encountered in educational research in Turkey. According to Akıncı and Köse (2021), there are problems in terms of methodological diversity and accuracy in studies conducted under the title of program evaluation in Turkey. On the other hand, Ak and Gülmez (2006) stated that academic studies conducted in Turkey are sometimes used to collect points, and that methodological depth is not always achieved. It may also be possible for this understanding sometimes to lead researchers to methodological errors, and this situation is emphasized in various studies (Toy & Tosunoğlu, 2007; Erdoğan, 2001). When the meta-evaluated study is considered in terms of propriety standards, it is seen that these standards are met to a large extent. However, it has various problems regarding the different cultural contexts mentioned before.

INTRODUCTION

Measurement and evaluation are of great importance in determining the extent to which the objectives have been achieved or revealing the success of the teaching. In addition to being an essential component in the evaluation of education in Turkey, measurement and evaluation have also been utilized as a tool at the point of transition between educational levels, admittance to employment, and educational institutions (Baykul, 2011). However, when measurement and assessment are seen in this light, it is limited to being a standard test practice and grading tool. Several investigations have found data that point to this constraint (Akbaş & Gençtürk, 2013; Berberolu, 2010; Saraç & Uygun, 2020). Instead, we should regard measurement and evaluation as a teaching qualification allowing us to examine student performance from all angles and make the best judgments possible.

Measurement and Evaluation as Teacher Qualification

Measurement is expressing a certain quality of an object with numbers or symbols, and evaluation is deciding the relevant quality based on the measurement results (Mislevy, 2018; Morrow Jr et al., 2000; Scriven, 1981). Based on these definitions, it is expected that teachers who make decisions about student success should be competent in this subject. This qualification is one of the essential competencies of the teaching profession both in universities and in the Ministry of National Education (MNE). It is stated that the first program for measurement and evaluation in education within universities was the doctoral program opened by Hacettepe University in the 1974-1975 academic year (Baykul, 2011). Afterward, undergraduate programs started to teach in this field, but these programs were closed with the regulation made in 1997 by the Council of Higher Education (CHE) (CHE, 1998). Today, the field of measurement and evaluation in education contributes to the educational sciences at the graduate level. In addition, support is received from academics in this field in the conduct of various vocational knowledge courses in teacher training undergraduate programs (CHE, 2007).

Measurement and Evaluation is one of the four essential teacher characteristics under the title of *Professional Skills* among the General Competencies of the Teaching Profession determined by the MNE General Directorate of Teacher Training and Development (MNE, 2017a). In addition, the Teacher Strategy Document emphasized the issue of training teachers who will carry out measurement and evaluation activities according to some standards (MNE, 2017b). Moreover, in the Education Vision 2023 Document, measurement and evaluation draw attention as one of the main issues that should be emphasized for the education system's success (MNE, 2019). These points indicate the importance of measurement and evaluation as a teaching qualification by the institutions responsible for teacher training and employment. This situation also draws attention to the quality of teaching activities expected to provide teachers with measurement and evaluation competencies. The measurement and evaluation process is essential for revealing the quality of instruction. However, as measurement and evaluation are primarily concerned with

student and teacher performance, systematic evaluation of a school system is only achievable through program evaluation studies (Nevo, 2002).

The practitioners of the teaching activity in school systems are the teachers. This situation also draws attention to the quality of teachers and the importance of teacher training programs in this context. In different studies, the importance of programs in teacher training and the role of teachers in ensuring the continuity of these programs were emphasized (Bullough, 1992; LaChausse et al., 2014). In addition, the practitioners of measurement and evaluation activities in schools are also teachers. Moreover, it is stated that measurement and evaluation are of great importance for the principle of continuity in the program development process (Tan, 2019).

In summary, there is a need for successful teachers for the realization of effective teaching, measurement, and evaluation processes, and effective programs for the training of successful teachers and effective teaching, measurement, and evaluation activities within these programs. These explanations indicate the strong relationship between teacher training, curriculum development, measurement, and evaluation concepts. The importance of program evaluation to reveal and increase the success of the developed programs is a separate issue that needs to be discussed.

Program Evaluation and Meta-Evaluation

It is necessary to utilize scientific research methods and formal evaluations that use well-structured procedures toward specific goals, locations, timeframes, and intended audiences to ensure the accuracy and practicality of evaluations of the effectiveness and potential of educational programs (Fitzpatrick et al., 2011; Uşun, 2012; Yüksel, & Sağlam, 2014). These program evaluations are comprehensive and costly and are referred to as Improvement/Accountability Oriented Approaches (Stufflebeam, 1999). Because more importance is given to the human factor in program evaluation studies today compared to previous years, and there are systematic data collection and analysis processes from many program stakeholders in these studies (Ornstein & Hunkins, 2018). However, it is not easy to consider such data collected from different program stakeholders and containing many different opinions as standard evaluations within a particular framework (Fitzpatrick et al., 2011). This situation shows that in addition to the standardization problem expressed, if the program evaluation processes are not carried out correctly, significant risks will arise regarding workload, time, and finances. This issue makes the quality of the program evaluation processes, as well as the quality of the program, questionable. From this point of view, the quality of program evaluation processes is revealed by meta-evaluations. Michael Scriven first introduced the concept of meta-evaluation in 1969 (Scriven, 2009). Meta-evaluation is expressed as the process of revealing the deficiencies and errors of one or more program evaluation studies or the evaluation of evaluation (Cooksy & Caracelli, 2009; Scriven, 2011). In this context, meta-evaluation also reveals whether a formal program evaluation research meets the needs, is carried out in long processes, and requires serious cost, time, and workload. In this respect, it is essential to adopt various standards as criteria to reveal the quality of a program evaluation research.

Program Evaluation Standards

Many meta-evaluation studies conducted since the 1980s have used program evaluation standards and checklists developed by JCSEE to publicly report the merits and drawbacks of program evaluation (Stufflebeam, 2000; WMU, 2020). The relevant Committee includes 12 organizations dealing with evaluation in education (JCSEE, 2018). The second and third editions of the curriculum evaluation standards, which consisted of 30 standard items and the first version published in 1981 by the Committee, were also issued in the following years (Fournier, 1994; Stufflebeam & Madaus, 1983). The Committee also profited from national and international examinations and field research involving over 400 stakeholders from various countries to establish the most recent edition of the program evaluation standards in 2010 (JCSEE, 2018). Since the 1980s, most meta-evaluation studies in domains such as economics, employment, environment, children's rights, addiction treatment, and educational practices have utilized these standards. (Alexakis, 2020; Chapman, 2012; Léveillé & Chamberland, 2010; Windsor et al., 1998). The main features of the Program Evaluation Standards developed by JCSEE (2018) are summarized below:

1. **Utility Standards:** The main focus of this standard area, which consists of eight standard items, is who evaluates the program, to what extent the stakeholders of the program are considered, how

well these stakeholders' needs regarding the program are met, in short, whether the evaluation serves its purpose.

2. **Feasibility Standards:** These standards aim to increase evaluation effectiveness and efficiency. Feasibility standards, which consist of four items, are related to the issue of whether the program evaluation is carried out with adequate, practical, economical, sound, and functional methods.

3. **Propriety Standards:** Propriety standards support the appropriate, fair, legal, truthful, and honest in evaluation. In this context, this standard area, which consists of seven standard items, focuses on issues such as respect for human rights and fulfilling financial obligations in the use of resources in a formal, transparent, and fair manner.

4. **Accuracy Standards:** Accuracy standards aim to increase the trustworthiness and accuracy of program evaluation data, findings, conclusions, and recommendations. This standard area consists of eight standard items and focuses on valid and reliable information obtained through appropriate methodological designs and analysis.

5. **Evaluation Accountability Standards:** Accountability standards encourage sufficient documentation and a meta-evaluation viewpoint emphasizing progress and accountability in evaluation procedures and outputs.

Program evaluation standards developed by JCSEE consist of 30 standards under the five headings shown above. Mainly the focus of *evaluation accountability standards* is meta-evaluation. Although meta-evaluation studies have been increasingly popular worldwide since the 1980s, there aren't many conducted in Turkey. A few meta-evaluation studies have been conducted in Turkey on various educational programs and initiatives. These include Yaşar et al. (2005) on teacher training programs for primary education between 1997 and 2004, Yüksel and Akın (2013) on the Student Success Determination Exam, Kılıç & Aslan (2016) on articles published on adult education in Turkey, Yağan (2019) on Ph.D. dissertations on program evaluation in Turkey, and Akıncı and Köse (2022) on teacher training programs in Turkey.

When these studies are examined in the context of their subjects, it was observed that only a few were conducted on program evaluation studies. In addition, in these studies, the researchers did not meta-evaluate any of their own program evaluation studies. However, as the final step in evaluating a program, meta-evaluation refers to showing the shortcomings and flaws of the research (Cooksy & Caracelli, 2009). In this way, meta-evaluation makes it possible to reveal findings on behalf of accountability about the extent to which the program evaluation research complies with the processes planned and negotiated at the beginning of the study and to what extent it serves its purpose. In this respect, meta-evaluation not only reveals the deficiencies of the program evaluation process but also increases the validity and reliability of the findings. Meta-evaluation provides the internal and external stakeholders of the program with the opportunity to review the program evaluation process while using the standards accepted in the international literature (JCSEE, 2018). Thus, it becomes possible for the program evaluation studies at the local level to gain an international identity. Another contribution of meta-evaluation to program evaluation is that it reveals the deficiencies of the studies and guides similar studies to be carried out in the future for possible problems to be encountered. These contributions of meta-evaluation to the program evaluation process and the strong relationship between teacher training, measurement, evaluation, and curriculum evaluation make this study meaningful. In this context, it was deemed necessary to conduct a meta-evaluation of a study aiming to reveal the quality of a technical course, such as measurement and evaluation in education, a compulsory course in teacher training programs.

Purpose of the Study

This study aims to make a meta-evaluation of the research titled *Evaluation of Measurement and Evaluation in Education Curriculum*. For this purpose, the study was conducted to determine the extent to which the research complies with the Program Evaluation Standards established by the JCSEE. In this regard, the study sought answers to the following questions:

1. To what extent does the research comply with program evaluation standards?
2. To what extent does the research comply with the relevant standards by standard areas?
3. To what extent does the research comply with the relevant standards according to the standard items?

METHOD

Research Design

This study is a meta-evaluation research. Stufflebeam (2000) defines meta-evaluation as the process of identifying, obtaining, and using descriptive and judgmental information about an evaluation's usefulness, feasibility, relevance, and accuracy to report on the strengths and weaknesses of that evaluation publicly. JCSEE (2018) has organized meta-evaluation under the evaluation accountability standards as three standard items: evaluation documentation, internal meta-evaluation, and external meta-evaluation. This research was also conducted considering the related topics. Due to the nature of the study, it does not require ethical permission as it does not involve any experimental procedure or data collection process on living things. However, in the study, due care was taken not to do anything contrary to research and publication ethics.

Meta-Evaluated Study

The meta-evaluated study is the Ph.D. dissertation prepared by the corresponding author of this research. The dissertation prepared by Akıncı (2021) was conducted under the supervision of the study's second author. The study, in which the multi-stage evaluation design and Stake's Responsive Program Evaluation Model were used, was aimed at evaluating the measurement and evaluation course program. For this purpose, the researcher collected data by using data collection tools such as semi-structured interviews, observations, achievement tests, and checklists from 484 pre-service teachers, 16 teachers, eight school administrators, and seven lecturers at faculties of education and theology affiliated with three different state universities and at different levels and types of public schools. The findings of the study indicated that the measurement and evaluation proficiencies of both teachers and pre-service teachers were inadequate. This deficiency was attributed to factors such as the inadequacy of the measurement and evaluation course for diverse teaching disciplines, the brevity of the course, and the absence of opportunities for practical training.

Data Collection Tool

The Program Evaluation Standards Checklist was used to check whether the meta-evaluated research met the required standards. Akıncı and Köse (2020) developed the checklist considering the program evaluation standards determined by JCSEE. Researchers received opinions from different experts on the relevant standards' translation, conceptual relevance, and item structures. In addition, the consistency of the coders between the experts examining the checklist items was checked, and the Cohen Kappa coefficient of the agreement was calculated as 0.81. According to Landis and Koch (1977), an adjustment of 0.61 and above is satisfactory. This case shows that the coder consistency is sufficient for the relevant checklist. At the end of these procedures, the final form of 30 items was given to the checklist by the researchers.

Data Collection and Analysis

The controls were made by an internal and an external evaluator using the Program Evaluation Standards Checklist to reveal whether the study carries the qualities of accurate planning, valid and reliable data collection, and performing the necessary analyzes and reporting, which are required for effective program evaluation research. While the academician selected as the external evaluator has the title of associate professor in the field of curriculum and instruction, the internal evaluator, who works as an assistant professor, is one of the experts in the field of measurement and evaluation, where observations are made in her/his class during the program evaluation process. In this context, the processes corresponding to the three standards related to the JCSEE (2018) accountability standard area in the data collection process are given below, respectively:

1. In response to the *Evaluation Documentation* standard, internal and external evaluators were also given access to the documents related to the research data upon request, provided that the participant information is hidden, as well as the final report version of the study.
2. *Internal Meta-Evaluation* refers to using program evaluation standards by different evaluators or program stakeholders to examine the accountability of evaluation design, procedures used, data collected, and results (JCSEE, 2018). In this context, the internal evaluator, the program's

stakeholder, was first requested to evaluate the study report with the help of the checklist prepared using the relevant standards. In addition, what kind of actions were taken for each standard item during the program evaluation process was presented to the internal evaluator.

3. At this point, the status of the research examined in meeting the program evaluation standards is revealed from the perspective of an external evaluator (JCSEE, 2018). The External Meta-Evaluation process allows an independent evaluator to perform the same procedures as the internal meta-evaluation process. In this context, the same procedures carried out with the internal evaluator were repeated with the external evaluator.

After these processes, the descriptive analysis method was used to analyze the data collected through the checklist from internal and external evaluators. In this context, descriptive statistics and explanations for each standard area and the total level of meeting the standards are presented.

FINDINGS

Research findings are presented in the form of tables and explanations under utility standards, feasibility standards, propriety standards, and accuracy standards. In Table 1, descriptive statistics regarding the level of meeting the utility standards of the meta-evaluated study were given.

Table 1.

Descriptive statistics on the level of meeting the utility standards.

Standard Area	Item	External Evaluator			İnternal Evaluator			Total $\bar{X}$		
		Yes	No	Insufficient Info	Yes	No	Insufficient Info	Yes	No	Insufficient Info
Utility Standards	1	+			+			2		
	2	+			+			2		
	3	+			+			2		
	4			+			+			2
	5	+			+			2		
	6	+			+			2		
	7	+			+			2		
	8	+			+			2		
	9	+			+			2		
	f	8	0	1	8	0	1	16	0	2
	%	88,9	0	11,1	88,9	0	11,1	88,9	0	11,1

Table 1 shows that the meta-evaluated study met the utility standards by 88.9%, according to both internal and external evaluators. Likewise, according to both evaluators, there needs to be more information about the fourth standard item in the relevant study. The mentioned item is presented below.

Item 4. *The personal and cultural values that form the basis of the purposes, processes, and judgments in the research evaluation processes are clearly expressed.*

In Table 2, descriptive statistics regarding the level of meeting the feasibility standards of the meta-evaluated study were presented.

Table 2.

Descriptive statistics on the level of meeting the feasibility standards

Standard Area	Item	External Evaluator			İnternal Evaluator			Total $\bar{X}$		
		Yes	No	Insufficient Info	Yes	No	Insufficient Info	Yes	No	Insufficient Info
Feasibility Standards	10	+			+			2		
	11	+			+			2		
	12		+				+		1	1
	13	+			+			2		
	f	3	0	1	3	0	1	6	1	1
	%	75	0	25	75	0	25	75	12.5	12,5

Table 2 shows that the meta-evaluated study met the program evaluation standards by 75%, according to both evaluators. In addition, according to the external evaluator, the 12th standard item still needed to be met in the study. On the other hand, the internal evaluator thinks that sufficient information should be given about the same item in the study. The above-mentioned standard item is as follows:

> **Item 12.** *The evaluations made in the research recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.*

The descriptive statistics regarding the level of meeting the propriety standards of the meta-evaluated study are shown in Table 3.

Table 3.

Descriptive statistics on the level of meeting the propriety standards

Standard Area	Item	External Evaluator			Internal Evaluator			Total X̄		
		Yes	No	Insufficient Info	Yes	No	Insufficient Info	Yes	No	Insufficient Info
Propriety Standards	14	+			+					
	15	+			+					
	16	+					+			
	17	+			+					
	18	+			+					
	19	+			+					
	20			+			+			
	21	+			+					
	f	7	0	1	6	0	2	6,5	0	1,5
	%	87,5	0	12,5	75	0	25	81,25	0	18,75

Table 3 shows that the propriety standards are met by 87.5% according to the external evaluator and 75% according to the internal evaluator. Both external and internal evaluators stated that sufficient information should have been given about the 20th standard item. In addition, according to the internal evaluator, in the meta-evaluated study, sufficient information needed to be provided about the 16th standard item. The items mentioned are as follows:

> **Item 16.** *Agreements made during the evaluation processes in the research are negotiated in a way that considers the needs, expectations, and cultural contexts of all stakeholders benefiting from and being affected by the program.*

> **Item 20.** *In the evaluation processes of the research, actual or anticipated conflicts of interest that could cast a shadow on the evaluation were clearly and honestly defined and eliminated.*

In Table 4, descriptive statistics regarding the level of meeting the accuracy standards of the meta-evaluated study were presented.

Table 4.

Descriptive statistics on the level of meeting the accuracy standards

Standard Area	Item	External Evaluator			İnternal Evaluator			Total $\overline{X}$		
		Yes	No	Insufficient Info	Yes	No	Insufficient Info	Yes	No	Insufficient Info
Accuracy Standards	22	+			+					
	23	+			+					
	24	+			+					
	25	+			+					
	26	+			+					
	27	+			+					
	28	+			+					
	29	+			+					
	30	+			+					
	f	9	0	0	9	0	0	9	0	0
	%	100	0	0	100	0	0	100	0	0

Table 4 indicates that the meta-evaluated study met the accuracy standards 100% according to external and internal evaluators. Table 5 shows descriptive statistics regarding the level of meeting the program evaluation standards in total.

Table 5.

Descriptive statistics on the level of meeting program evaluation standards

	Standard Sum		
		f	%
External Evaluator	Yes	27	90
	No	1	3.33
	Insufficient Info	2	3.33
İnternal Evaluator	Yes	26	86,66
	No	0	0
	Insufficient Info	4	13,33
Total	**Yes**	**26,5**	**88,33**
	No	**0.5**	**1,66**
	Insufficient Info	**3**	**10**

When Table 5 is examined, while the study meets the program evaluation standards at the rate of 90% according to the external evaluator, this rate is 86.66% according to the internal evaluator. This value is 88.33% in total. According to the evaluators, more information was needed about the study's four standard items. These items emphasize cultural values, interests, contexts, and conflicts of interest. The experts think these concepts should have been considered sufficiently in the study. However, it is still observed that the study largely complies with the program evaluation standards.

CONCLUSION, DISCUSSION, AND RECOMMENDATIONS

The results of the meta-evaluation indicate that the study under examination demonstrates a high degree of alignment with the Program Evaluation Standards, as determined by both internal and external evaluators, with a concordance rate of 88.33%. In addition, according to both evaluators, Utility Standards were met by 88.9%, Feasibility Standards by 75%, and Accuracy Standards by 100%, while Propriety Standards were met at a rate of 87.5% according to the external evaluator and 75% according to the internal evaluator. Moreover, the experts think more information should have been provided in the research, especially about cultural values, interests, contexts, and conflicts of interest. Akıncı and Köse (2020) stated in their study that evaluation studies on teacher training programs in Turkey have similar shortcomings. This issue may occur

because cultural and contextual features are not considered enough in curriculum development studies carried out centrally in Turkey. The problem of developing programs without adequate consideration of social and cultural characteristics is also emphasized in different studies (Aslan, & Uygun, 2019; Kurt, 2016; Paksoy, 2020). However, the concepts of culture and education are so intertwined that this concept has found a place in the definition of education. Education is expressed as enculturation, that is, the society shaping individuals in line with their expectations and wishes (Helvacı, 2008). Another study emphasizes the transfer of cultural heritage to future generations, which is one of the main functions of education (Genç, 2018). In this respect, it is essential to consider cultural and contextual features in curriculum development. In addition, with the decision of the General Council of Higher Education dated 10.08.2020, the authority was transferred to the faculties responsible for training teachers in the determination of the courses, curricula, and credits in the programs, considering various issues in the 2018 curriculum (CHE, 2020). Despite the transfer of authority, no regulation has been found regarding teacher training programs on cultural values and different contexts that will increase compliance with program evaluation standards in the universities where the program evaluation study was conducted. Although this situation allows the relevant institutions to make regulations, it indicates that a centralized approach is still dominant in matters such as cultural values, interests, contexts, and conflicts of interest.

When the research findings are considered in the context of utility standards, it was observed that the relevant study principally serves its purpose. Contrary to these findings, it has been stated that there are problems in terms of quality and quantity in serving the purpose of program evaluation research in Turkey. Akıncı and Köse (2021) stated in their study that both the evaluated programs and program evaluation studies in Turkey have problems in serving their purpose. Such studies are referred to as Pseudoevaluations because they are evaluations that express positive or negative opinions about the program, regardless of its actual benefit and value (Stufflebeam, 1999). Similarly, it has been emphasized in different studies that the number of program evaluation studies conducted in Turkey is limited (Gündüz & Kuzu Demir, 2020; Keskin & Yazar, 2019). This situation shows that more program evaluation studies should be conducted that evaluate different program types and serve their purpose. At this point, the risks posed by program evaluation research in terms of time, workload, and cost should not be forgotten. The preference for practical, economical, helpful, and functional methods to increase the effectiveness and efficiency of the evaluation is related to the feasibility standards (JCSEE, 2018). In this context, the meta-evaluated research has the most problems in the relevant standard field. The basis of this problem may lie in the fact that the methodological choices are inappropriate, especially in terms of practicality. This issue is more about research culture. Akcan et al. (2018) stated that the research culture in Turkey is education-oriented, and there are problems in the approach to research in this respect.

When the research findings are examined in terms of accuracy standards, the study, which was meta-evaluated according to internal and external evaluators, meets this standard 100%. This finding is a positive and expected finding for a study that evaluates the program of a technical course that includes psychometric features such as measurement and evaluation in education. However, this is only sometimes a situation encountered in program development and evaluation studies and other educational research in Turkey. According to Akıncı and Köse (2021), there are problems in terms of methodological diversity and accuracy in studies conducted under the title of program evaluation in Turkey. On the other hand, Ak and Gülmez (2006) stated that academic studies conducted in Turkey are sometimes used to collect points, and that methodological depth is not always achieved. It may also be possible for this understanding sometimes to lead researchers to methodological errors, and this situation is emphasized in various studies (Toy & Tosunoğlu, 2007; Erdoğan, 2001). When the meta-evaluated study is considered in terms of propriety standards, it is seen that these standards are met to a large extent. However, it has various problems regarding the different cultural contexts mentioned before.

As a result, the rapid change in almost every field in the world also reflects in education, program, and evaluation. In this respect, various standards are also used in addition to the approaches or models adopted for many years in program development and evaluation (JCSEE, 2018). In addition to the programs being evaluated, the quality of the evaluation process has become a subject of extensive research (Astbury, 2016). In this study, a meta-evaluation of the related study was carried out after a comprehensive program evaluation research that lasted for about two years with a similar understanding. Although the study primarily meets the program evaluation standards, it has been observed that there needs to be more accountability in matters such as cultural values, interests, contexts, and conflicts of interest based on Turkey's research culture and program development approach. From this point of view, various suggestions that are thought to contribute to program evaluation and meta-evaluation studies are presented:

1. Social and cultural characteristics should be considered more in the program development and evaluation process.
2. Teacher training programs should be arranged to train teachers for different needs and contexts by taking advantage of the transfer of authority to education faculties.
3. Program evaluation studies should be conducted with more effective and comprehensive methodological preferences in different program types and levels.
4. Methodological depth and diversity should be increased in program evaluation studies.
5. Meta-evaluation studies should be conducted on the quality of program evaluation studies carried out by different researchers in various contexts.
6. Program evaluation and meta-evaluation studies should be conducted in light of the approach, model, and standards developed and adopted in line with the structure and needs of the Turkish education system.

Important note: Due to the nature of the study, it does not require ethical permission as it does not involve any experimental procedure or data collection process on living things. However, in the study, due care was taken not to do anything contrary to research and publication ethics.

REFERENCES

Ak, M. Z., & Gülmez, A. (2006). Türkiye'nin uluslararası yayın performansının analizi. *Akademik incelemeler dergisi*, *1*(1), 22-49.

Akbaş, Y., & Gençtürk, E. (2013). Coğrafya öğretmenlerinin alternatif ölçme-değerlendirme teknikleri ile ilgili görüşleri: kullanma düzeyleri, sorunlar ve sınırlılıklar. *Doğu Coğrafya Dergisi*, *18*(30), 331-356.

Akcan, A. T., Malkoç, S., & Kızıltan, Ö. (2018). Akademisyenlere göre akademi ve akademik kültür. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *18*(1), 569-591.

Akıncı, M. (2021). *Eğitimde ölçme ve değerlendirme dersi programının değerlendirilmesi* (PhD dissertation, Akdeniz Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Antalya). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi.

Akıncı, M., & Köse, E. (2022). A meta-evaluation research on teacher training programs in Türkiye. *International Journal of Progressive Education* , *18*(4), 209-222.

Akıncı, M., & Köse, E. (2021). Research trends of program evaluation studies conducted between 2010-2019 in Turkey. *Çukurova University Faculty of Education Journal, 50*(1), 77-120.

Alexakis, D. E. (2020). Meta-evaluation of water quality indices. application into groundwater resources. *Water*, *12*(7), 1-13.

Aslan, M., & Uygun, N. (2019). Okul öncesi eğitim programının Stufflebeam'in bağlam, girdi, süreç ve ürün (BGSÜ) değerlendirme modeline göre değerlendirilmesi. *Eğitim ve Bilim*, *44*(200), 229-251.

Astbury, B. (2016). From evaluation theory to tests of evaluation theory?. In Stockmann, R., & Meyer, W. (Ed). *The Future of Evaluation* (pp. 309–325). London: Palgrave Macmillan.

Baykul, Y. (2011). Ülkemizde ölçme ve değerlendirmenin dünü, bugünü ve yarını. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 2*, 1-32.

Berberoğlu, G. (2010) Ölçme ve değerlendirmenin uygulamaya yansıyan boyutu. *Cito Eğitim: Kuram ve Uygulama*, *10*, 10-14.

Bullough Jr, R. V. (1992). Beginning teacher curriculum decision making, personal teaching metaphors, and teacher education. *Teaching and Teacher Education*, *8*(3), 239–252.

Chapman, L. S. (2012). Meta-evaluation of worksite health promotion economic return studies: 2012 update. *American Journal of Health Promotion*, *26*(4), 1–12.

CHE, (2020). *YÖK'ten eğitim fakültelerine önemli yetki devri kararı.* Retrieved February 4, 2021, from https://www.yok.gov.tr/Sayfalar/Haberler/2020/egitim-fak%C3%BCltelerine-yetki-devri.aspx

CHE, (2007). *Öğretmen yetiştirme ve eğitim fakülteleri (1982-2007).* Retrieved November 26, 2021, from https://www.yok.gov.tr/Documents/Kurumsal/egitim_ogretim_dairesi/Ogretmen-Yetistirme/2.pdf

CHE, (1998). *Türk yükseköğretiminin bugünkü durumu.* Retrieved December 4, 2021, from https://www.yok.gov.tr/Documents/Yayinlar/Yayinlarimiz/Turk-yuksekogretiminin-bugunku-durumu-mart-1998.pdf

Cooksy, L. J., & Caracelli, V. J. (2009). Meta-evaluation in practice. *Journal of Multidisciplinary Evaluation*, *6*(11), 1–15.

Erdoğan, İ. (2001). Sosyal bilimlerde pozitivist-ampirik akademik araştırmaların tasarım ve yöntem sorunları. *Anatolia: Turizm araştırmaları dergisi*, *12*(2), 119-134.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines.* (4th Edition). Pearson Education Inc.

Fournier, D. M. (1994). The program evaluation standards: How to assess evaluations of educational programs. *Journal of Educational Measurement*, *31*(4), 363-367.

Genç, S. Z. (2018). Kültür ve eğitim. In Köse, E., & Genç, S. Z. (Ed.) *Eğitim sosyolojisi* (pp. 115-160). Ankara: Pegem Akademi.

Gündüz, G. F., & Kuzu Demir, E. B. (2020). 2017 bilişim teknolojileri ve yazilim dersi öğretim programının ortaokul 5. sınıf öğrencilerinin görüşlerine göre değerlendirilmesi: Eskişehir ili örneği. *Trakya Eğitim Dergisi*, *10*(3), 1024-1041.

Helvacı, A. (2008). Eğitimle ilgili temel kavramlar. In Figen Ereş (Ed.). *Eğitim bilimine giriş* (pp. 1-19). Ankara: Maya Akademi.

JCSEE. (2018). *Checklist of the program evaluation standards statements*. Retrieved January 5, 2022, from https://evaluationstandards.org/program

Keskin, İ., & Yazar, T. (2019). Ortaöğretim matematik dersi öğretim programının öğretmen görüşlerine göre değerlendirilmesi. *Türkiye Sosyal Araştırmalar Dergisi*, *23*, 1-28.

Kılıç, Ç., & Arslan, S. (2016). Türkiye'de yayınlanan yetişkin eğitimine ilişkin makalelerin meta değerlendirmesi. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, *18*(2), 865-899.

Kurt, A. (2016). *4. sınıf İngilizce dersi öğretim programının bağlam, girdi, süreç, ürün modeline göre değerlendirilmesi* (Masters Thesis, Akdeniz Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Antalya). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi.

LaChausse, R. G., Clark, K. R., & Chapple, S. (2014). Beyond teacher training: The critical role of professional development in maintaining curriculum fidelity. *Journal of Adolescent Health*, *54*(3), 53-58.

Landis, J, R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

Léveillé, S., & Chamberland, C. (2010). Toward a general model for child welfare and protection services: A meta-evaluation of international experiences regarding the adoption of the Framework for the Assessment of Children in Need and Their Families (FACNF). *Children and Youth Services Review*, *32*(7), 929-944.

Mislevy, R. J. (2018). On measurement in educational assessment. In Secolsky, C., & Denison, D. B. (Ed.) *Handbook on measurement, assessment, and evaluation in higher education* (2nd Edition) (pp. 37-64). Routledge.

MNE, (2019). *2023 eğitim vizyonu.* Retrieved January 5, 2022, from http://2023vizyonu.MNE.gov.tr/

MNE, (2017a). *Öğretmenlik mesleği genel yeterlikleri*. Öğretmen Yetiştirme ve Geliştirme Genel Müdürlüğü.

MNE, (2017b). *Öğretmen Strateji Belgesi 2017-2023*. Öğretmen Yetiştirme ve Geliştirme Genel Müdürlüğü.

Morrow Jr, J. R., Jackson, A. W., Disch, J. G., & Mood, D.P. (2000). *Measurement and Evaluation in Human Performance* (2nd Edition). Human kinetics.

Nevo, D. (Ed.). (2002). *School-based evaluation: An international perspective*. Oxford: Elsevier Science ltd.

Ornstein, A. C., & Hunkins, F. P. (2018). *Curriculum: Foundations, principles, and issues*. (2nd Edition). Pearson Education Limited.

Paksoy, E. N. (2020). *Okul öncesi eğitim programının Stufflebeam (CIPP) program değerlendirme modeli ile incelenmesi* (Master's thesis, Kilis 7 Aralık Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Kilis). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi.

Saraç, E., & Uygun, N. (2020). Sınıf öğretmeni adaylarının ölçme ve değerlendirme tekniklerine yönelik tercihlerinin ders planları ile incelenmesi. *Uluslararası Eğitim Araştırmacıları Dergisi*, *3*(2), 199-212.

Scriven, M. (1981). *Evaluation thesaurus* (3rd Edition). Edge Press.

Scriven, M. (2009). Meta-evaluation revisited. *Journal of multidisciplinary evaluation*, *6*(11), 3–8.

Scriven, M. (2011). Evaluating evaluations: A meta-evaluation checklist. *Consultado em*, *9*, 1-7.

Stufflebeam, D. L., & Madaus, G. F. (1983). The standards for evaluation of educational programs, projects, and materials. In Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (Ed.) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (pp. 395-404). Dordrecht: Kluwer Academic Publishers.

Stufflebeam, D. L. (1999). *Foundational models for 21st century program evaluation*, The Evaluation Center Occasional Papers Series, Western Michigan University.

Stufflebeam, D. L. (2000). The methodology of metaevaluation. In Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (Ed.) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (pp. 457–471). Dordrecht: Kluwer Academic Publishers.

Toy, B. Y., & Tosunoğlu, N. G. (2007). Sosyal bilimler alanındaki araştırmalarda bilimsel araştırma süreci, istatistiksel teknikler ve süreci, istatistiksel teknikler ve yapılan hatalar yapılan hatalar. *Journal of commerce*, *1*(1), 1-20.

Uşun, S. (2012). *Eğitimde program değerlendirme: Süreçler yaklaşımlar ve modeller.* Anı Yayıncılık.

Windsor, R. A., Boyd, N. R., & Orleans, C. T. (1998). A meta-evaluation of smoking cessation intervention research among pregnant women: improving the science and art. *Health education research*, *13*(3), 419–438.

WMU. (2020). *Evaluation checklists.* Retrieved January 5, 2022, from https://wmich.edu/evaluation/checklists

Yağan, S. A. (2019). Program değerlendirme alanında yayınlanmış doktora tezlerinin meta değerlendirmesi (2015-2019). *Electronic Journal of Education Sciences*, *8*(16), 188-208.

Yaşar, S., Gültekin, M., Köse, N., Girmen, P., & Anagün, S. (2005). The meta-evaluation of teacher training programs for elementary education in Turkey. In Cooper (Ed.) Teacher Education: *Local and Global:* Proceedings of the 33rd Annual Australian Teacher Education Association Conference (pp. 498-504). Australia: Griffith University.

Yüksel, İ., & Akın, Z. (2013). Öğrenci başarısını belirleme sınavının (ÖBBS) öğrenci değerlendirme standartları kapsamında değerlendirilmesi: Bir meta değerlendirme araştırması. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, *32*(1), 473-495.

Yüksel, İ., & Sağlam, M. (2014). *Eğitimde program değerlendirme*. (2nd Edition). Ankara: Pegem Akademi Yayıncılık.