A Meta-Evaluation Research on Teacher Training Programs in Türkiye*

Muhammed Akıncı [i]
Recep Tayyip Erdoğan University

Erdoğan Köse [ii]
Akdeniz University

Abstract

The purpose of this research is to make a meta-evaluation of the program evaluation studies on teacher training programs in Turkey. Meta-evaluation is the process of revealing the deficiencies and errors of the research as the last stage of the program evaluation process. In this context, the steps of the meta-evaluation processes were followed methodically. In the in this research, 9 program evaluation studies conducted between 2010 and 2020 about teacher training programs, using any program evaluation approach and model, were meta-evaluated. The sampling method of the research is criterion sampling.. Research data was collected using the "*Program Evaluation Standards Checklist* developed by the researchers. During the development of the relevant data collection tool, the *Program Evaluation Standards* created by the Joint Committee on Standards for Educational Evaluation (JCSEE) were benefited. The studies included in the research were examined by 6 program experts who formed the meta-evaluation team. Each expert evaluated 3 studies using the checklist. Research data was analyzed using descriptive analysis method. Research findings indicate that the examined program evaluation studies meet the program evaluation standards by 55.67%. From this point, some suggestions that are believed to contribute to future program evaluation and meta-evaluation studies on teacher training were presented.

Key words: Meta-Evaluation, Program Evaluation, Teacher Training, Evaluation Standards, Program Standards.

----

[i] **Muhammed Akıncı,** Research Assist Dr., Curriculum and Instruction, Recep Tayyip Erdoğan University, ORCID: 0000-0002-5001-2080

Correspondence: muhammed.akinci@erdogan.edu.tr

[ii] **Erdoğan Köse,** Prof. Dr., Faculty of Education, Akdeniz University

INTRODUCTION

Evaluation has always been a part of the school. Although evaluation in education is known as a process carried out especially in classrooms, one of its main interlocutors is the program developer. Related program experts make use of evaluation processes to reveal the quality of the program. The effective execution of educational activities in different fields and levels is directly related to the quality of the applied program. Moreover, the quality of the programs depends on the effective execution of the development and evaluation processes. Although there are different approaches, program development is generally expressed as all the processes carried out in the light of scientific processes to make the curriculum effective (Demirel, 2014; Null, 2011; Ornstein & Hunkins, 2018; Ornstein & Levine, 2008). Program evaluation, on the other hand, is defined as determining the extent to which the program meets the features that it should have using the same scientific processes and to make various decisions about the program (Fitzpatrick, Sanders, & Worthen, 2011; Madaus, Stufflebeam, & Scriven, 2000; Mertens, & Wilson, 2019; Stufflebeam, & Coryn, 2014; Uşun, 2012; Yüksel & Sağlam, 2014).

While the evaluations carried out within the school focus more on student success, the evaluation of the program is a comprehensive process that expresses the systematic evaluation of the entire school system (Nevo, 2002). Teachers, on the other hand, are the main practitioners of educational activities in the school systems. In this respect, the role of teachers and teacher training programs in school systems where different programs are developed and evaluated is indisputably very important (Bullough, 1992). This means that the teachers have a key role in the effective implementation and success of the developed programs. The importance of teacher training and quality is emphasized to ensure the continuity of a program developed in various studies (LaChausse, Clark, & Chapple, 2014; Valcke, Rots, Verbeke, & Van Braak, 2007). In this respect, the high quality of teacher-training programs is a great necessity for the teachers, who are the output of these programs, to be able to carry out the current programs effectively.

Perhaps the most well-known purpose of a developed program is to increase the quality of education as much as possible in the field it is developed. This is directly related to the nature of the program. It is important to employ formal processes to make the right decisions about the nature and future of the program. Formal evaluations are valid and reliable evaluations with a certain systematic, purpose, place, time, and address (Fitzpatrick, et al., 2011). Such evaluations, defined as *Improvement/Accountability Oriented Approaches*, are comprehensive and expensive evaluations that consider all the questions and criteria required to evaluate the success of a program (Stufflebeam, 1999). This is an indicator of major financial risks that may arise if the program evaluation processes are not carried out correctly. This causes the quality of the program evaluation processes questionable as well as the quality of the program. The quality of these processes can be revealed through meta-evaluations. Scriven (2009) states that he first invented the concept of meta-evaluation, which he defined as the evaluation of evaluation, shortly before he included it in an article in 1969. Stufflebeam (2000) defines meta-evaluation as the process of identifying, obtaining, and using descriptive and judgmental information about an evaluation, such as its usefulness, feasibility, relevance, and accuracy, to publicly report on the strengths and weaknesses of that evaluation. Meta-evaluation has been used to reveal the quality of evaluations and practices in many different fields such as economy, employment, environment, children's rights, and fight against addiction studies (Alexakis, 2020; Chapman, 2012; Léveillé, & Chamberland, 2010; Pediaditi, Doick, & Moffat, 2010; Windsor, Boyd, & Orleans, 1998). The focus of this study is the quality of various evaluations made in the context of education and program.

The quality of program evaluation research is directly related to whether they are conducted according to certain standards. In this context, studies were carried out by the 1980 Joint Committee on Standards for Educational Evaluation (JCSEE) and in 1981 a document consisting of 30 standards was published to guide the evaluation process of educational programs, projects, and materials and to judge the soundness of such evaluations (Fournier, 1994; Stufflebeam, & Madaus, 1983). In the following years, these standards were developed and updated, and the 2nd and 3rd editions were

published under the title of *Program Evaluation Standards* (JCSEE, 2018). Most of the meta-evaluation studies in the international literature have tried to reveal the quality of the evaluations with the help of these standards since the 1980s. In 2010, the Committee benefited from national and international reviews and field research, involving more than 400 stakeholders from different countries, to develop the most up-to-date version of program evaluation standards (JCSEE, 2018). These standards have been developed within Western Michigan University and have been transformed into different checklists prepared by famous scientists in the field of program evaluation such as Michael Scriven, Daniel Stufflebeam, Robert Stake, Michael Quinn Patton, Egon Guba, Ernest R. House (WMU, 2020).

Program Evaluation Standards have been prepared in a way that considers each stage, from the beginning of evaluation research till the end of the process, to reveal the quality of the research and evaluation. These standards consist of a total of 30 standards under 5 headings as utility, feasibility, propriety, accuracy, and accountability standards. The main features of these standard fields are as follows (JCSEE, 2018):

Utility Standards: Aims to increase the awareness of the program stakeholders about the importance of evaluation processes and products in meeting their needs.

Feasibility Standards: Aims to increase evaluation effectiveness and efficiency.

Propriety Standards: Supports what is appropriate, fair, legal, right, and just in evaluation.

Accuracy Standards: Aims increase reliability and accuracy of descriptions, suggestions, and findings that support comments and judgments, especially about quality of evaluation.

Evaluation Accountability Standards: Encourages adequate documentation for evaluation and a meta-evaluation perspective that focuses on improvement and accountability in evaluation processes and products (p. 1-3).

Although a complete meta-evaluation involving research design, data collection, data analysis, and checking or reconsidering results is rarely done, it is very important to examine even just one of these elements (Scriven, 2009). In this respect, international literature shows that the number of meta-evaluation studies conducted in the context of the program is quite limited. Despite this limitation, it is possible to see that studies using this method, which was put forward in the late 1960s, in the context of program evaluation, were more preferred especially after the 1980s, when the interest in program evaluation standards increased (Ardisson, Smallheer, Moore & Christenbery, 2015; Usmani, Khatoon, Shammot & Zamil, 2012; Scott-Little, Hamann & Jurs, 2002; Finn, Stevens, Stufflebeam & Walberg, 1997; Georghiou, 1995; Odom & Fewell, 1983). Although their number is limited, there are various studies under the heading of meta-evaluation in Turkey. Yasar, Gultekin, Kose, Girmen & Anagun (2005) conducted a meta-evaluation study on the evaluation studies on primary school teacher training programs in Turkey between 1997 and 2004. Yüksel & Akın (2013) conducted a meta-evaluation study on the Student Success Determination Exam. Yağan (2019) on the other hand, conducted a meta-evaluation study on Ph.D. dissertations published on program evaluation in Turkey. Akıncı & Köse (2021) stated that program evaluation studies in Turkey are carried out most frequently at the undergraduate level, in terms of the field, teacher training programs were evaluated most, and these studies have various problems in different dimensions. In this context, the adaptation of the Program Evaluation Standards developed by the Joint Committee on Standards for Educational Evaluation (JCSEE) into Turkish to carry out the program evaluation studies according to the standards accepted in the international literature in the

following years and to conduct meta-evaluation studies with the checklist prepared using these standards might be the solution for this problem.

Purpose of the Research

The purpose of this research is to make a meta-evaluation of the program evaluation studies on teacher training programs in Turkey. Under this general-purpose, answers to the following questions were sought in the study:

- How well do the reviewed studies meet the program evaluation standards?

  o According to standard types,

  o According to standard items,

  o According to program types,

  o According to research types.

The study is important in terms of revealing the quality of program evaluation research, which has become widespread in the national literature in recent years, in the context of teacher training programs. Thus, it is believed that the study will contribute to the conduct of more planned and standardized program evaluation studies.

METHOD

Research Design

This study is a meta-evaluation research. Meta-evaluation is the process of revealing the deficiencies and errors of the research as the last stage of the program evaluation process (Cooksy & Caracelli, 2009; Scriven, 2011). In this context, the steps of the meta-evaluation processes were followed methodically in the research. These steps are as follows (Stufflebeam, 2000):

1. Determine and arrange to interact with the meta evaluation's stakeholders.

2. Establish a qualified meta evaluation team.

3. Define the meta evaluation questions.

4. Agree on standards to judge the evaluation system or particular evaluation.

5. Negotiate the meta evaluation contract.

6. Collect and review pertinent available information.

7. Collect new information as needed, including, for example, on-site interviews, observations, and surveys.

8. Analyze the qualitative and quantitative information and judge the evaluation's adherence to the selected evaluation standards.

9. Prepare and submit the needed reports.

10. Help the client and other stakeholders interpret and apply the findings. (p. 461).

The above stages were used in the processes of establishing the meta-evaluation team, selecting the study to be examined and the standards to be used, collecting, analyzing, and reporting the data.

Examined Research Papers

In the study, program evaluation studies conducted between 2010 and 2020 about teacher training programs were evaluated. Another criterion in the selection of program evaluation studies examined in the research was program evaluation approaches and models. Among the studies under the title of program evaluation, those using any of the program evaluation approaches and models were included in the study. Since the mentioned criteria were used in the study, the sampling method was expressed as criterion sampling. Because the criterion sample is valid in all cases that meet a predetermined set of criteria (Yıldırım & Şimşek, 2013). From databases such as Dergipark, National Thesis Center, ULAKBİM, Tr Index, Web of Science (WOS), EBSCO, ERIC, Elsevier, and Google Academic, 9 program evaluation studies meeting these criteria were reached. The distribution of the examined studies by years is presented in Figure 1.
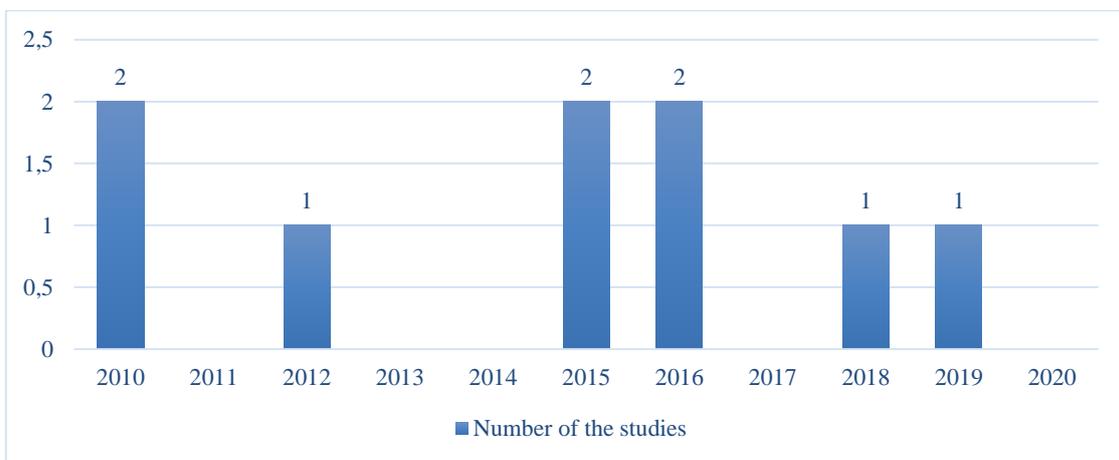
Figure 1. Distribution of the Studies by Years

Figure 1 shows that studies using any program evaluation approach and model in the evaluation of teacher training programs are limited to an average of 1 or 2 studies per year. Even in some years, studies that meet the mentioned criteria about teacher training programs have not been conducted. The distribution of the studies according to the program evaluation models is shown in Figure 2.
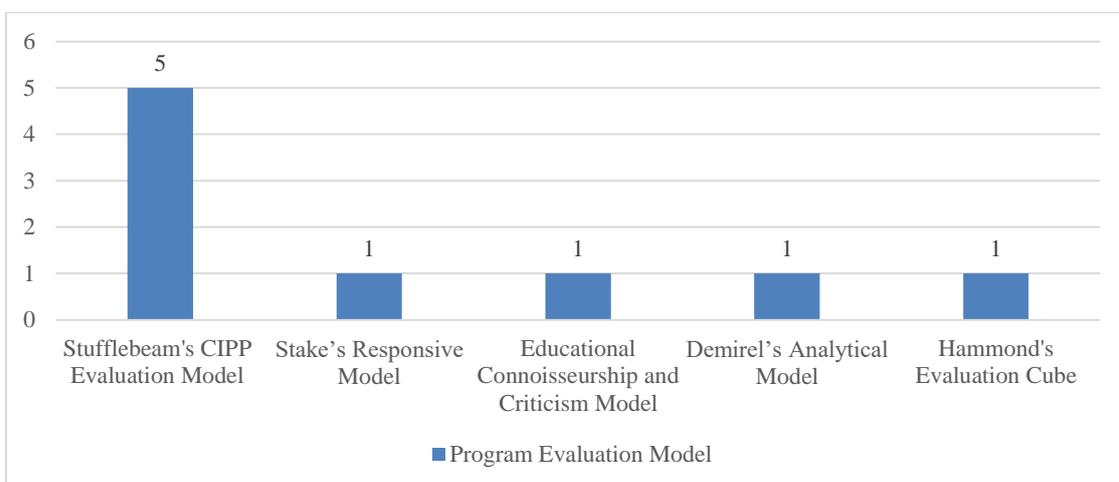
Figure 2. Distribution of Studies According to Program Evaluation Models

Figure 2 indicates that Stufflebeam's CIPP Model was preferred in five of the nine studies meta-evaluated. Apart from this model, four different models were preferred in the evaluation of teacher training programs.

Data Collection and Analysis Procedure

Research data was collected using the "*Program Evaluation Standards Checklist* developed by the researchers. During the development of the relevant data collection tool, the *Program Evaluation Standards* created by the Joint Committee on Standards for Educational Evaluation (JCSEE) were considered. The relevant standards were translated into Turkish, and the opinions of two foreign language education experts for translation and two program experts for conceptual relevance were obtained. The translated standards were transformed into checklist items, and the opinions of two measurement and evaluation experts were obtained on the structure of the items. The checklist was finalized with 30 items. Cohen Kappa coefficient was calculated as 0.81 for consistency among experts who analyzed the items. The Kappa coefficient ranges from -1 to +1, and as it gets closer to +1, it indicates that the random agreement for consistency among experts decreases (Fleiss and Cohen 1973). In addition, the standards expressed as *Evaluation Accountability Standards* were not included in the checklist during the meta-evaluation. This is because the relevant standards are related to whether meta-evaluation has already been done.

The studies included in the research were examined by 6 program experts who formed the meta-evaluation team. Each expert evaluated 3 studies using the checklist. Thus, each research was evaluated twice by different experts for consistency. Experts marked "yes" in the checklist if the study under review meets the relevant standard, and "no" if it does not. If there is not enough information about the relevant standard in the study, the "insufficient information" option was selected. Moreover, the experts in the meta-evaluation team were composed of individuals who have Ph.D. degrees in curriculum and instruction and have studies on program evaluation. Of the relevant experts, 2 are female and 4 are male, and 1 has the title of professor, 2 has associate professor, and 3 of has the title of doctor. In the analysis of the research data, the descriptive analysis method was used, and the findings were presented as descriptive statistics.

FINDINGS

The research findings addressed the extent to which the meta-evaluated studies met the program evaluation standards in terms of the standard types and items, the type of study, and the type of program evaluated. In this context, the descriptive statistics regarding the extent to which the studies examined meet the program evaluation standards are presented in Table 1.

Table 1. Descriptive Statistics Regarding the Level of Meeting the Program Evaluation Standards of the Examined Studies

Studies Reviewed	Research 1				*Total*		Research 2				*Total*	
Reviewing Experts	Expert 1		Expert 2				Expert 1		Expert 2			
	f	%	f	%	f	%	f	%	f	%	f	%
Yes	22	73,3	20	66,7	21	70	18	60	14	46,7	16	53,3
No	3	10	4	13,3	3,5	11,7	5	16,7	10	33,3	7,5	25
Insufficient Info	5	16,7	6	20	5,5	18,3	7	23,3	6	20	6,5	21,7

Studies Reviewed	Research 3				*Total*		Research 4				*Total*	
Reviewing Experts	Expert 1		Expert 2				Expert 3		Expert 4			
	f	%	f	%	f	%	f	%	f	%	f	%
Yes	14	46,7	12	40	13	43,3	19	63,3	18	60	18,5	61,7
No	7	23,3	11	36,7	9	30	3	10	12	40	7,5	25
Insufficient Info	9	30	7	23,3	8	26,7	8	26,7			4	13,3

Studies Reviewed	Research 5				Total		Research 6				Total	
Reviewing Experts	Expert 3		Expert 4				Expert 3		Expert 4			
	f	%	f	%	f	%	f	%	f	%	f	%
Yes	18	60	13	43,3	15,5	51,7	21	70	17	56,7	19	63,3
No	3	10	5	16,7	4	13,3	2	6,7	4	13,3	3	10
Insufficient Info	9	30	12	40	10,5	35	7	23,3	9	30	8	26,7

Studies Reviewed	Research 7				Total		Research 8				Total	
Reviewing Experts	Expert 5		Expert 6				Expert 5		Expert 6			
	f	%	f	%	f	%	f	%	f	%	f	%
Yes	18	60	18	60	18	60	15	50	18	60	16,5	55
No	2	6,7	6	20	4	13,3	6	20	6	20	6	20
Insufficient Info	10	33,3	6	20	8	26,7	9	30	6	20	7,5	25

Studies Reviewed	Research 9				Total		TOTAL		
Reviewing Experts	Expert 5		Expert 6					f	%
	f	%	f	%	f	%			
Yes	12	40	13	43,3	12,5	41,7	Yes	16,7	55,67%
No	7	23,3	11	36,7	9	30	No	5,9	19,67%
Insufficient Info	11	36,7	6	20	8,5	28,3	Insufficient Info	7,4	24,67%
							Total	30	100,00%

Table 1 shows that the meta-evaluated studies met the program evaluation standards by 55.67%. While the related studies do not meet the program evaluation standards by 19.67%, it was observed that insufficient information is given at the rate of 7.4% regarding various standards in some studies. According to the studies examined study 1 meets the relevant standards at the highest level at 70%, while Study 9 meets the lowest level at 41.7%. The findings regarding the extent to which the examined studies meet the relevant standards according to standard types and on the basis of items are presented in Table 2.

Table 2. Descriptive Statistics Regarding the Level of Meeting the Relevant Standards According to Standard Types and Items

Items	Utility Standards																		Total	
	1.		2.		3.		4.		5.		6.		7.		8.		9.			
	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%
Yes	16	88,9	8	44,4	10	55,6	1	5,6	8	44,4	8	44,4	13	72,2	14	77,8	15	83,3	10,3	57,2
No	2	11,1	10	55,6	3	16,7	3	16,7	4	22,2	2	11,1	2	11,1	1	5,6	1	5,6	3,1	17,2
Insufficient Info	-	-	-	-	5	27,8	14	77,8	6	33,3	8	44,4	3	16,7	3	16,7	2	11,1	4,6	25,6

Items	Feasibility Standards								Total	
	10.		11.		12.		13.			
	f	%	f	%	f	%	f	%	f	%
Yes	7	38,9	6	33,3	3	16,7	11	61,1	6,8	37,8
No	7	38,9	12	66,7	1	5,6	-	-	5	27,8
Insufficient Info	4	22,2			14	77,8	7	38,9	6,2	34,4

Items	Propriety Standards																Total	
	14.		15.		16.		17.		18.		19.		20.		21.			
	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%
Yes	15	83,3	2	11,1	1	5,6	16	88,9	15	83,3	15	83,3	1	5,6	9	50	9,3	51,7
No	-	-	4	22,2	2	11,1	-	-	-	-	2	11,1	-	-	-	-	1	5,5
Insufficient Info	3	16,7	12	66,7	15	83,3	2	11,1	3	16,7	1	5,6	17	94,4	9	50	7,7	42,8

Items	Accuracy Standards																		Total	
	22.		23.		24.		25.		26.		27.		28.		29.		30.			
	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%
Yes	7	38,9	16	88,9	17	94,4	10	55,6	13	72,2	15	83,3	5	27,8	13	72,2	11	61,1	11,9	66,1
No	-	-	2	11,1	1	5,6	8	44,4	5	27,8	3	16,7	13	72,2	5	27,8	1	5,6	4,2	23,3
Insufficient Info	11	61,1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	33,3	1,9	10,6

Table 2 shows that the level of meeting the program evaluation standards of the studies examined according to standard types varies between 37.8% and 66.1%. The standard type in which the program evaluation standards are met the highest is Accuracy Standards with 66.1%, and Feasibility Standards are the lowest with 37.8%. In addition, the 4 standards that are met at the highest level in the studies are respectively:

24. Information on evaluation in research supports valid interpretations of program evaluation.

1. The research was conducted by qualified people who proved themselves in evaluation.

17. Evaluation processes in the research were designed and carried out in a way to protect human and legal rights and ensure the dignity of participants and other stakeholders.

23. Information about evaluation in research serves the intended purposes.

The 4 standards that are met at the lowest level in the studies are respectively:

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

16. Agreements made during the evaluation processes in the research were negotiated in a way considering the needs, expectations, and cultural contexts of all stakeholders benefiting from and affected by the program.

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.

15. Agreements made during the evaluation processes of the research were negotiated with all stakeholders benefiting from and affected by the program to clarify their obligations.

The studies examined in the research were divided into 3 groups as studies in the fields of primary education programs, teaching practice, and professional teaching knowledge courses, considering the programs they evaluated. Table 3 shows the findings regarding the extent to which program evaluation studies meet the relevant standards, according to the type of programs evaluated.

Table 3. Descriptive Statistics of Program Evaluation Studies on the Level of Meeting the Standards According to Program Types

Studies Evaluating Teaching Practice Couse Programs													
Studies Reviewed	Study 1				Study 2				Study 3				Total	
Reviewing Experts	Expert 1		Expert 2		Expert 1		Expert 2		Expert 1		Expert 2			
	f	%	f	%	f	%	f	%	f	%	f	%	f	%
Yes	22	73,3	20	66,7	18	60	14	46,7	14	46,7	12	40	16,6	55,4
No	3	10	4	13,3	5	16,7	10	33,3	7	23,3	11	36,7	6,7	22,3
Insufficient Info	5	16,7	6	20	7	23,3	6	20	9	30	7	23,3	6,7	22,3
Studies Evaluating Professional Teaching Knowledge Programs													

Studies Reviewed	Study 4				Study 5				Study 6				Total	
Reviewing Experts	Expert 3		Expert 4		Expert 3		Expert 4		Expert 3		Expert 4			
	f	%	f	%	f	%	f	%	f	%	f	%	f	%
Yes	19	63,3	18	60	18	60	13	43,3	21	70	17	56,7	17,7	59
No	3	10	12	40	3	10	5	16,7	2	6,7	4	13,3	4,8	16
Insufficient Info	8	26,7			9	30	12	40	7	23,3	9	30	7,5	25

Studies Evaluating Primary Education Programs

Studies Reviewed	Study 7				Study 8				Study 9				Total	
Reviewing Experts	Expert 5		Expert 6		Expert 5		Expert 6		Expert 5		Expert 6			
	f	%	f	%	f	%	f	%	f	%	f	%	f	%
Yes	18	60	18	60	15	50	18	60	12	40	13	43,3	15,7	52,3
No	2	6,7	6	20	6	20	6	20	7	23,3	11	36,7	6,3	21
Insufficient Info	10	33,3	6	20	9	30	6	20	11	36,7	6	20	8	26,7

Table 3 indicates that the total scores regarding the level of meeting the program evaluation standards of all studies in terms of the type of program evaluated are between 52.3% and 59%. Studies evaluating the programs of professional teaching knowledge courses are the studies that meet the related standards to 59% at the highest level. The studies that meet the evaluation standards at the lowest level according to the type of program evaluated are the studies that evaluate the primary education programs with 52.3%. Table 4 shows the findings regarding the extent to which program evaluation studies meet the relevant standards, according to research types. In this context, the studies examined are grouped as PhD dissertations master's theses, and articles.

Table 4. Descriptive Statistics on the Level of Meeting the Relevant Standards of Program Evaluation Researches by Research Type

Studies Reviewed	PhD Dissertations													
Reviewing Experts	Study 1		Study 2		Study 4		Study 6		Study 7		Study 8		Total	
	f	%	f	%	f	%	f	%	f	%	f	%	f	%
Yes	21	70	16	53,3	18,5	61,7	19	63,3	18	60	16,5	55	18,2	60,7
No	3,5	11,7	7,5	25	7,5	25	3	10	4	13,3	6	20	5,2	17,3
Insufficient Info	5,5	18,3	6,5	21,7	4	13,3	8	26,7	8	26,7	7,5	25	6,6	22

Studies Reviewed	Master's Theses						Article			
Reviewing Experts	Study 3		Study 9		Total		Study 5		Total	
	f	%	f	%	f	%	f	%	f	%
Yes	13	43,3	12,5	41,7	12,8	42,7	15,5	51,7	15,5	51,7
No	9	30	9	30	9	30	4	13,3	4	13,3
Insufficient Info	8	26,7	8,5	28,3	8,2	27,3	10,5	35	10,5	35

Table 4 shows that the total scores regarding the level of meeting the program evaluation standards of all studies vary between 42.7% and 60.7% according to the research type. According to the study type, the studies that meet the evaluation standards at the highest level are PhD dissertations with 60.7%, and the ones that meet the lowest level are master's theses with 42.7%.

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

The rapid change and development process that takes place in every field in the world also affects the understanding of education. It is thought that the breakthroughs made by developed and developing countries in recent years will also affect the education systems and this effect will also be reflected in the program development and evaluation processes (Stockmann & Meyer, 2016). Today, many countries have focused on the quality of evaluations, which are the decision-making process about the programs, as well as the quality of the education system and programs to keep up with this change (Astbury, 2016). The main purpose of this approach, called meta-evaluation, is to publicly

report the strengths and weaknesses of the evaluation (Scriven, 2009; Stufflebeam, 2000). For this purpose, program evaluation studies carried out on teacher training in Turkey between the years 2010-2020 were meta-evaluated.

The research findings show that the evaluation studies on teacher training programs meet the program evaluation standards adopted in the research at a rate of 55.67%. Yağan (2019) in her study similarly found that various standards were not adequately met in the program evaluation studies she examined. The standard type in which the program evaluation standards are met at the lowest rate is the Feasibility Standards. Evaluation results show that this standard type is met at a very low level of 37.8%. However, Feasibility Standards are about the effectiveness and efficiency of the evaluation (JCSEE, 2018). Effective and efficient program evaluations require the use of methodically practical and economical processes. This situation makes it questionable whether the right choices are made methodologically in the program evaluation studies on teacher training programs. It was stated in different studies that there are problems related to methodological preferences in studies on education programs in Turkey (Kozikoğlu & Senemoğlu, 2015; Ozan & Köse, 2014). On the other hand, the Accuracy Standards are the type of standards that are met at the highest level with 66.1%. Accuracy Standards aim to increase the reliability and accuracy of descriptions, recommendations, and findings that support comments and judgments on evaluation, especially about quality (JCSEE, 2018). Meeting these standards regarding scientific ethics at the highest level might be considered a positive situation. However, it is also important that the researches are to be feasible as much as being ethical, valid, and reliable.

The items that meet the program evaluation standards at the highest level are related to the conduct of studies by experts in the field, within the framework of ethical rules, and in a way that serves its purpose. In addition, the lowest met standards in the study focused on the concepts of cultural values, interests, contexts, and conflicts of interest. Akıncı (2021) reached similar findings in his study and as a possible reason, he showed that cultural and contextual features are not considered sufficiently in the program development processes carried out centrally in Turkey.

When the relevant studies are examined based on program type, it was observed that the studies evaluating the programs of professional teaching knowledge courses meet the program evaluation standards at the highest level, and the studies evaluating the primary education programs at the lowest level. The reason for this might be that the studies evaluating primary education programs try to evaluate all the courses of an undergraduate program at once. In addition, the researches that meet the evaluation standards according to study type at the highest level are doctoral theses, while those that meet the lowest level are master's theses. Stufflebeam (1999) stated that program evaluation studies are comprehensive and require expertise. PhD dissertations might be meeting the relevant standards for this reason. Moreover, evaluation studies on teacher training programs mostly consist of PhD dissertations and master's theses. Akıncı & Köse (2021) reached similar findings of the insufficient number of articles on program evaluation. This situation might be associated with not preferring the studies that require time and effort. Akcan, Malkoç & Kızıltan (2018) stated that there are serious problems in the approach to research in Turkey and that academic culture focuses on education rather than scientific research. Adherence to the CIPP Model of Stufflebeam to a large extent in the studies examined may also be an indication that methods and models that require time and effort are not preferred in program evaluation. Because Stufflebeam's CIPP Model was designed to provide ease of application for different types of researchers in program evaluation (Stufflebeam & Coryn 2014). It was stated in different studies that this model has been widely used in program evaluation in Turkey (Akıncı & Köse, 2021; Kurt & Erdoğan 2015; Özüdogru, 2018).

As a result, the fact that the examined studies meet the program evaluation standards by approximately 55% makes the quality of these studies questionable. There are already studies that draw attention to the problems related to the quality of scientific research conducted in Turkey (Erdoğan, 2001; Ak & Gülmez, 2006 Toy & Tosunoğlu, 2007). When considered in terms of standard types, while focusing on being valid and reliable, there are program evaluations whose feasibility level is decreasing. In addition, the evaluation of an entire undergraduate program as a thesis work or the

fact that the program evaluation is carried out by researchers who have just entered the field at the master's level may prevent meeting the required standards. From this point of view, some suggestions that are thought to contribute to future program evaluation and meta-evaluation studies are as follows:

- The variety of qualitative and quantitative data collection tools should be increased in program evaluation studies, and the focus should be on the feasibility as well as focusing on the methodological validity and reliability of the researches.

- Expanding the use of program evaluation approaches and models may be useful for conducting systematic evaluation research. In this context, program evaluation studies should be conducted in the light of the approach, model, and standards developed and adopted in line with the needs of the Turkish education system.

- It is difficult for program evaluation research to be conducted as a master's thesis because it is extensive and expensive. Therefore, it may be more effective and efficient to conduct related studies at the PhD level or by a group of researchers.

- Different researchers should carry out meta-evaluation studies on the quality of program evaluation studies conducted in various contexts.

REFERENCES

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 (2021). Evaluation of the measurement and evaluation in education curriculum (PhD dissertation, Akdeniz University, Department of Educational Sciences, Antalya). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi.

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.

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

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

Ardisson, M., Smallheer, B., Moore, G., & Christenbery, T. (2015). Meta-Evaluation: Experiences in an Accelerated Graduate Nurse Education Program. *Journal of Professional Nursing*, *31*(6), 508-515.

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

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.

Cooksy, L. J., & Caracelli, V. J. (2009). Meta-evaluation in Practice. *Journal of multidisciplinary evaluation*, *6*(11), 1-15.

Demirel, Ö. (2014). *Eğitimde program geliştirme kuramdan uygulamaya.* (21. Edition). Ankara: Pegem Akademi Yayıncılık.

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

Finn Jr, C. E., Stevens, F. I., Stufflebeam, D. L., & Walberg, H. J. (1997). A meta-evaluation. *International Journal of Educational Research*, *27*(2), 159-174.

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

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, *33*(3), 613-619.

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

Georghiou, L. (1995). Assessing the framework programmes: A Meta-evaluation. *Evaluation*, *1*(2), 171-188.

JCSEE. (2018). Checklist of the program evaluation standards statements. Retrieved from https://evaluationstandards.org/program.

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

Kozikoğlu, I., & Senemoğlu, N. (2015). The Content Analysis of Dissertations Completed in the Field of Curriculum and Instruction (2009-2014). *Education and Science*, *40*(182), 29-41.

Kurt, A., & Erdoğan, M. (2015). Content Analysis and Trends of Curriculum Evaluation Research: 2004-2013. *Education and Science*, *40*(178), 199-224.

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.

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.

Madaus, G. F., Stufflebeam, D. L., & Scriven, M. (2000). Program Evaluation: A Historical Overview. In Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (Ed.) In *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (pp. 3-22). Dordrecht: Kluwer Academic Publishers.

Mertens, D. M., & Wilson, A. T. (2019). *Program evaluation theory and practice*. (2. Edition). Guilford Publications.

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

Null, W. (2011). *Curriculum: From theory to practice*. Rowman & Littlefield Publisher, Inc.

Odom, S. L., & Fewell, R. R. (1983). Program evaluation in early childhood special education: A meta-evaluation. *Educational Evaluation and Policy Analysis*, *5*(4), 445-460.

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

Ornstein, A. C., & Levine, D. U. (2008). *Foundations of education*. (3. Edition). New York: Houghton Mifflin Company.

Ozan, C., & Köse, E. (2014). Eğitim programları ve öğretim alanındaki araştırma eğilimleri. *Sakarya University Journal of Education*, *4*(1), 116-136.

Özüdoğru, F. (2018). Analysis of curriculum evaluation studies conducted in foreign language education: 2005-2016. *Journal of Language and Linguistic Studies, 14*(2), 113-134.

Pediaditi, K., Doick, K. J., & Moffat, A. J. (2010). Monitoring and evaluation practice for brownfield, regeneration to greenspace initiatives: A meta-evaluation of assessment and monitoring tools. *Landscape and Urban Planning*, *97*(1), 22-36.

Scott-Little, C., Hamann, M. S., & Jurs, S. G. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, *23*(4), 387-419.

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.

Stockmann, R., & Meyer, W. (2016). *The future of evaluation: Global trends, new challenges and shared perspectives*. Palgrave Macmillan, London.

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.

Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, and applications*. (2. Edition). San Francisco: Josey-Bass.

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.

Usmani, M. A. W., Khatoon, M. S., Shammot, M. M., & Zamil, A. M. (2012). Meta evaluation of a teachers' evaluation programme using CIPP model. *Archives Des Sciences*, *65*(7), 230-252.

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

Valcke, M., Rots, I., Verbeke, M., & Van Braak, J. (2007). ICT teacher training: Evaluation of the curriculum and training approach in Flanders. *Teaching and teacher Education*, *23*(6), 795-808.

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 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.

Yasar, S., Gultekin, M., Kose, N., Girmen, P., & Anagun, 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.

Yıldırım, A., & Şimşek, H. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri* (9.Edition). Ankara: Seçkin Yayıncılık.

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

Yüksel, İ., & Zeynep, A. (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.