



## Research Trends of Program Evaluation Studies Conducted between 2010-2019 in Turkey

Muhammed AKINCI<sup>a\*</sup> (ORCID ID - 0000-0002-5001-2080)

Erdoğan KÖSE<sup>b</sup> (ORCID ID - 0000-0003-0426-0267)

<sup>a</sup>Recep Tayyip Erdoğan Üniversitesi, Eğitim Fakültesi, Rize/Türkiye

<sup>b</sup>Akdeniz Üniversitesi, Eğitim Fakültesi, Antalya/Türkiye



### Article Info

DOI: 10.14812/cufej.688142

#### Research Article

#### Article history:

Received 11.02.2020

Revised 25.02.2021

Accepted 03.03.2021

#### Keywords:

Research trends,  
Curriculum,  
Program evaluation,  
Content analysis.

### Abstract

This study aims to determine the trends of the researches on program evaluation between the years 2010-2019 in Turkey. The research is a qualitative research using intrinsic case study design. In the study, 44 research papers conducted between 2010 and 2019 using any program evaluation model were examined. Research findings indicate that in Turkey program evaluation studies have become preferable since 2016, which use a particular program evaluation model as a criterion. Researchers preferred Stufflebeam's CIPP Evaluation Model in most of the studies and most commonly undergraduate and high school programs were evaluated. Methodologically most preferred designs were mixed and non-random sampling and as data collection instruments, interviews, questionnaires, observations, and scales were commonly used. Some results based on the findings of the research show that there are deficiencies regarding the selection of program evaluation models and methodological preferences, and there are inconsistencies between the results and recommendations regarding the quality of the program in Turkey. In order to keep up with international developments in program evaluation, suggestions were made regarding the development and use of approaches, models, and standards that serve the needs of the Turkish education system.

## 2010-2019 Yılları Arasında Türkiye’de Program Değerlendirme Konusunda Yapılan Araştırmaların Eğilimleri

### Makale Bilgisi

DOI: 10.14812/cufej.688142

#### Araştırma Makalesi

#### Makale Geçmişi:

Geliş 11.02.2020

Düzeltilme 25.02.2021

Kabul 03.03.2021

#### Anahtar Kelimeler:

Araştırma eğilimleri,  
Eğitim programı,  
Program değerlendirme,  
İçerik analizi.

### Öz

Bu çalışmada Türkiye’de 2010-2019 yılları arasında eğitimde program değerlendirme konusunda yapılan araştırmaların eğilimlerinin belirlenmesi amaçlanmıştır. Çalışma iç içe geçmiş durum çalışması deseninin kullanıldığı nitel bir araştırmadır. Araştırmada 2010 ile 2019 yılları arasında herhangi bir program değerlendirme modelini kullanılarak gerçekleştirilen 44 çalışma incelenmiştir. Araştırma bulguları Türkiye’de program değerlendirme araştırmalarında 2016 yılı itibari ile belli bir program değerlendirme modelinin ölçüt olarak alan çalışmaların daha fazla tercih edilir hale geldiğini göstermektedir. Yapılan çalışmaların büyük bir çoğunluğunda Stufflebeam’in BGSÜ Program Değerlendirme Modelinin tercih edildiği ve en çok lisans programları ve liselere ait programların değerlendirildiği görülmüştür. Yöntemsel olarak en çok karma yöntem ve seçkisiz olmayan örnekleme yöntemleri, veri toplama aracı olarak ise görüşme, anket, gözlem ve ölçekler tercih edilmiştir. Türkiye’de program değerlendirme modellerinin seçimi ve yöntemsel tercihlere ilişkin eksikler olduğu ve yapılan çalışmalarda programın niteliğine ilişkin sonuçlar ve öneriler arasında tutarsızlıklar olduğu gibi bazı sonuçlara ulaşılmıştır. Program değerlendirmede uluslararası gelişmelere ayak uydurulabilmesi için Türk eğitim sisteminin ihtiyaçlarına uygun yaklaşım, model ve standartların geliştirilmesi ve kullanımına ilişkin önerilerde bulunulmuştur.

\* Author: muhammed.akinci@erdogan.edu.tr

### Introduction

The education system of Turkey has been mentioned with its problems for years and these problems are discussed in different contexts. Failure in international exams, problems in production and economy, brain drain, etc. are possible reasons for this situation (Bakioğlu & Yıldız, 2014; Çalışkan, Karabacak, & Meçik, 2013; Akyüz, 2006; Işıkçok, 2002). The problematic nature of the education system suggests that program development activities may also be problematic. Undoubtedly, program development activities, effective evaluation of these developed programs, and revealing their qualifications directly affect the success of the education system. According to Yüksel and Sağlam (2014), program development is the design, implementation, evaluation, and rearrangement of programs in line with the information obtained. This definition also emphasizes the importance of determining the quality of the programs that is only possible with effective program evaluation activities.

Program evaluation is defined as the evaluation of a planned and implemented program by subjecting it to scientific research processes in line with appropriate program evaluation models and the use of the results in the development of the same program (Yüksel, & Sağlam 2014). Another definition expresses program evaluation as the process of determining whether a program is valid, useful, efficient, effective, etc. with a scientific understanding and a decision-making process about any feature of it (Uşun, 2012). Systematic, valid, and reliable evaluations with specific objectives, place, time, stakeholders are defined as *formal* evaluations (Fitzpatrick, Sanders, & Worthen, 2011). In this respect, formal evaluations that form the basis of formative and summative evaluations could be stated as essential evaluations in which scientific research processes are considered. Moreover, formal program evaluation processes allow for different methodological choices, such as quantitative, qualitative, or mixed (Yüksel & Sağlam, 2014; Uşun, 2012). At this point, methodological preferences may also be related to the approaches adopted. Ornstein and Hunkins (2018) divided program evaluation approaches into scientific (modern) and humanist (post-modern) approaches. While in scientific (modern) approaches, program evaluators take into account clearly stated objectives and precise indicators of whether their students are achieving the intended results of the program, humanists prefer methods such as aesthetics, ethnography, autobiography, phenomenology, critical literacy, etc. instead of statistical methods. House (1990) mentions monolithic, and pluralist approaches in his study. The classification of commonly used program evaluation approaches in Turkey and common models under this classification are as follows (Aygören & Er, 2018; Demirtaş, 2017; Yüksel & Sağlam, 2014; Uşun, 2012; Fitzpatrick, et al., 2011):

- Objective/Program-Oriented Evaluation Approach
  - The Tylerian Evaluation Approach
    - Metfessel-Michael Evaluation Model
    - Provus's Discrepancy Evaluation Model
- Expertise-Oriented Evaluation Approach
  - Eisner's Educational Connoisseurship and Criticism Model
- Consumer-Oriented Evaluation Approach
- System/Decision-Oriented Evaluation Approach
  - Stufflebeam's Context, Input, Process, Product (CIPP) Model
  - UCLA Evaluation Model
- Participant-Oriented Evaluation Approach
  - Stake's Responsive Evaluation Approach
    - Stake's Congruence-Contingency Evaluation Model

### **Objective/Program-Oriented Evaluation Approach**

The Objective/Program-Oriented Evaluation Approach focuses on the achievement of the objectives at the end of the educational processes. In this approach, the main role of the program evaluator is to decide how much and how well the program objectives have been achieved (Fitzpatrick et al., 2011). This approach is interpreted as successful to provide useful information to decision-makers since the implementation stages are also clear in cases where the objectives of the program are clearly stated. However, it is criticized in this aspect since it is thought that focusing only on the objectives raises the risk of neglecting the program features not specified in these objectives (Fitzpatrick, et al., 2011; Stufflebeam, Madaus, & Kellaghan, 2000).

### **Expertise-Oriented Evaluation Approach**

The Expertise-Oriented Evaluation Approach unlike the previous approaches is considered as a humanist and postmodern approach since its focus is not to raise individuals who will contribute to the system economically in program evaluation (Ornstein & Hunkins 2018). This approach places emphasis on field expertise in program evaluation (Yüksel & Sağlam, 2014; Uşun, 2012).

Although the Expertise-Oriented Evaluation Approach is considered valuable in terms of the importance it attaches to individuals and the academic freedom it offers as it is a humanistic approach, it is criticized for being expert-focused and the selection of these experts is difficult in terms of competence and objectivity (Fitzpatrick et al., 2011).

### **Consumer-Oriented Evaluation Approach**

Consumer-oriented evaluations first became important in evaluating educational products from the mid to late 1960s. The most important reason for this is that in those years the governments of the United States of America (USA) started to offer serious funds to develop new educational products (Fitzpatrick et al., 2011). The leading representative of this approach is considered as Michael Scriven. Scriven's approach to evaluation generally aims to identify and rank the costs and effects of alternative programs and educational products that consumers and the wider society can benefit from according to the assessed needs by using product evaluation checklists (Stufflebeam & Coryn 2014).

Consumer-Oriented Evaluation Approach is considered successful in raising awareness about the quality of educational products, providing high-quality educational products, and bringing some standards to the programs (Stufflebeam, et al., 2000). Moreover, high costs and the emergence of unnecessary prescriptiveness are considered as weaknesses of the approach (Fitzpatrick, et al., 2011).

### **System/Decision-Oriented Evaluation Approach**

System/Decision-Oriented Evaluation Approach has emerged as an alternative to Objective/Program-Oriented evaluation approaches and models, which started to be criticized at the end of the 1960s. In this respect, it focuses on the decisions of the managers rather than the objectives of the program. While the systematic and planned perspective of this approach to different dimensions of the program is considered as its positive side, the manager-oriented perspective is sometimes criticized for causing subjective evaluations (Fitzpatrick, et al., 2011; Yüksel & Sağlam, 2014).

### **Participant-Oriented Evaluation Approach**

The birth of the Participant-Oriented Evaluation Approach is based on the reaction that the basic feature of evaluation is limited to the achievement of the objectives or standards in the first years of program evaluation in the USA. Towards the end of the 1960s, some educators, especially Stake stated that program evaluation processes were nothing more than comprehensive reports written after long mechanical studies, and as a result, program evaluators could not fully control the program they were evaluating. These scientists argued that more attention should be given to the individual called “the participant”, in program evaluation, so that the program can be evaluated more effectively with different perspectives (Fitzpatrick, et al., 2011; Yüksel & Sağlam, 2014).

Participant-Oriented Evaluation is a preferred approach as it allows for many different subjective, objective, qualitative and quantitative data collection processes and emphasizes the human factor. However, this approach is criticized due to the difficulty of the analysis of many different types of data and the validity-reliability processes, the lack of objectivity, and high costs (Fitzpatrick, et al., 2011; Yüksel & Sağlam, 2014).

### **Program Evaluation in the World and Turkey**

When the basic approaches adopted in program evaluation are examined, it is observed that a technical and rigid understanding that focuses on the achievement of the objectives was adopted in the years when program evaluation started to become widespread. In the following years, a change and orientation towards an approach that takes the stakeholders of the program more into consideration have emerged (Ornstein & Hunkins 2018; Fitzpatrick, et al., 2011; Stufflebeam, Madaus & Kellaghan, 2000). The biggest problem arising from such consideration of stakeholder participation in program evaluation is shown as the conflict between research and participation processes. While being a scientist requires an independent, impartial, and unbiased approach to research, the inclusion of non-scientists in the research process means dealing with conflicting expectations of these people (Meyer, Funk, & Nguyen, 2016). At this point, the question arises as to whether this problem expressed will cause a standardization problem for the evaluation of the program. There are various attempts of different people and organizations to prevent this problem. Especially Stake, the leading name of Participatory-Oriented Evaluation Approach, and the Joint Committee on Standards for Educational Evaluation (JCSEE) have been working on standards-based program evaluation and program evaluation standards (Stake, 2004; Stufflebeam & Madaus, 1983). In this context, the main purpose of standard development studies, which have been widespread since the 1980s, is to ensure that program evaluation practices are carried out within scientific processes (Fournier, 1994). Particularly, the studies conducted within Western Michigan University, by JCSEE which has an important role in evaluation standards, affected Europe, especially Switzerland and Germany, as of the end of the 1990s and the beginning of the 2000s and related countries translated these standards into their languages in 1997 and 1999 (Beywl, 2000; cited in Yüksel & Sağlam, 2014). This situation expresses the existence of an approach that has evolved from evaluating the program to evaluating the program evaluation processes according to certain standards over time (Astbury, 2016). It is stated that this approach, which is called meta-evaluation and aims to publicly report the strengths and weaknesses of the evaluation, emerged in the late 1960s (Scriven, 2009; Stufflebeam, 2000). Besides, JCSEE's program evaluation standards include a dimension called *Accountability Standards*, the focus of which is meta-evaluation (JCSEE, 2018). This situation is an indication of the importance given to program evaluation studies, whose place, time, stakeholders, models, and approaches used in program evaluation, especially in developed countries, Today, especially in developed countries this situation is an indication of the importance given to program evaluation studies meeting various standards with a certain place, time, stakeholders, models, and approaches.

The position and importance of formal program evaluation studies mentioned above have been expressed in different studies conducted in Turkey. In these studies, it is emphasized that a program evaluation study can only be successful if conducted by appropriate people and the necessary points are taken into consideration in the light of the right approach and models (Yapıcıoğlu, Kara, & Yalçinoğlu, 2016; Yüksel & Sağlam, 2014; Ertürk, 2013; Uşun, 2012). Besides, there are researches in Turkey that examine program development and evaluation studies from various dimensions and aim to identify trends in various aspects (Yetkiner, Erdol, & Ünlü, 2019; Özüdoğru, 2018; Koç, 2015; Kurt & Erdoğan, 2015). These studies show that the number of studies on program evaluation tends to increase as the years progress, and accordingly, program evaluation approaches and models are preferred more than previous years. However, in related studies, the condition of program evaluation in Turkey until 2015 has been revealed. Studies to examine the program evaluation studies conducted in the following years focused on evaluations in a specific area such as primary education, foreign language, a certain teaching level, program type, or only one study type such as Ph.D. dissertations. The increasing trend in approach and model preference expressed in related studies indicates that the program evaluation perspective

has started to change positively by settling on a more formal and scientific basis in Turkey. In this context, it is crucial to conduct new trend research covering the last 10 years, which includes current studies on program evaluation between 2015 and 2019, and that will allow comparing the situation of program evaluation studies before 2015 with the next five years.

### **Purpose**

This study aims to determine the trends of the researches on program evaluation in 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, and 2019 in Turkey. It is observed that the studies under the title of *research trend* in the field of educational sciences in Turkey are mostly thematize and examine studies conducted in a certain field between certain dates according to their historical distribution and subject and method preferences (Ulutaş & Ubuz, 2008; Başaroğlu, Şahin, & Göktaş, 2013; Ozan. & Köse, 2014; Kurt & Erdoğan, 2015). In other words, the concept of the *trend* in related studies refers to the preference for any of the ways to be followed in the solution of a research problem related to a determined subject in any field.

When the previously mentioned structural and methodological qualities of program evaluation studies considered they indicate that some properties such as the type and levels of the program they evaluate, the evaluation model they take as criteria in the evaluation process, the reasons for choosing the relevant model, distribution of preferred models according to program types and levels, methodological preferences such as sampling, data collection, and analysis, results and recommendations regarding the quality of the program, etc. could be examined. In this context, under the general purpose of the study, answers for the following questions were sought regarding the studies on program evaluation, beginning from 2010 till the end of 2019 in Turkey:

1. How is the distribution of the studies according to the years and the databases they are indexed?
2. What are the program evaluation models used in the studies and the reasons for choosing these models?
3. How is the distribution of program evaluation models preferred in the studies according to program types and levels?
4. What are the research methods and designs preferred in the studies?
5. What are the sampling methods preferred in the studies?
6. What are the data collection tools preferred in the studies?
7. What are the achieved results in terms of the quality of the programs evaluated in the studies?
8. What kind of decisions has been suggested about the programs evaluated in the studies?

Examining studies on program evaluation in Turkey, comparing, combining, and interpreting the findings and results, and putting forward suggestions of related studies for the future of the program may be beneficial in recognizing various problems in the process of developing and evaluating the program in the education system and providing contextual information to researchers who will conduct program evaluation studies. This study is believed to be important in this regard.

### **Method**

#### **Research Design**

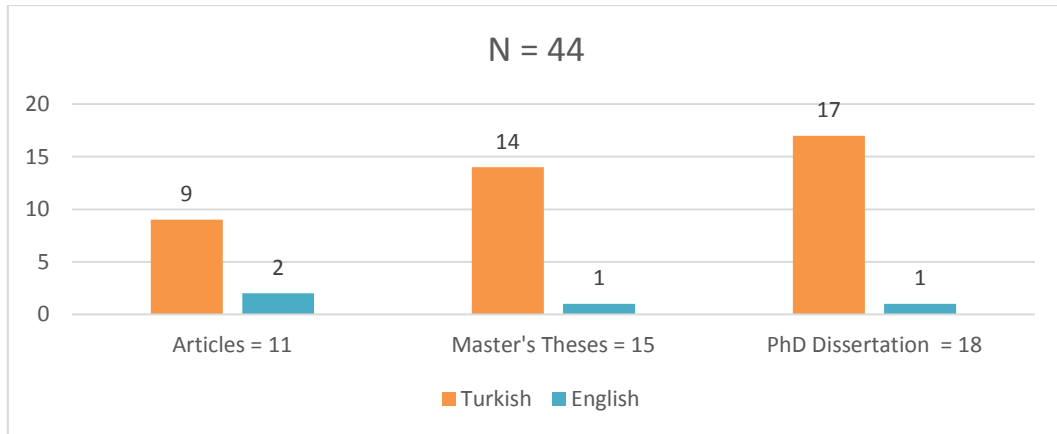
This research is a study that uses qualitative research methods and techniques in methodological design, data collection, and analysis. In this context, the research design is the intrinsic case study. In this design one or more situations, events, settings, programs, or social groups are examined (Creswell, 2007). In the intrinsic case study, a situation is analyzed by dividing it into sub-units (Yıldırım & Şimşek, 2013). Besides, the study is descriptive in terms of expressing the current situation of program evaluation studies. “Descriptive researches are studies that try to investigate the event as it is and determine the existing situation” (Karakaya, 2012, p. 59).

### Examined Research Papers

In the research, 176 studies were accessed from various databases such as Dergipark, National Thesis Center, ULAKBİM, Tr Index, Web of Science (WOS), EBSCO, ERIC, Elsevier, and Google Academic conducted between 2010 and 2019 in Turkish and English languages and 49 of these carried out using any program evaluation model, were taken into account. It was observed that the five published articles belong to master's and doctoral theses, which are among these 41 studies. As a result, 44 studies were included in the study to be examined. The predetermined criteria used to include relevant studies in the research are as follows:

- Studies included in the research consist of articles and postgraduate theses.
- The studies included in the research were carried out in 10 years, beginning from 2010, covering 2019.
- Studies conducted by using any program evaluation model were included in the research.
- Only theses were included in the research in program evaluation studies published as both thesis and article.

Since the above characteristics of the studies included in the study are accepted as criteria, the sampling method used in this study can be expressed as criterion sampling. The main feature of the criterion sampling is that it is valid in all situations that meet a set of pre-determined criteria (Yıldırım & Şimşek, 2013). Information on the numbers, percentage, and publication languages of research papers included in the study is presented in Graphic 1.



**Graphic 1.** Research Papers and Publication Languages of Them

According to Graphic 1, the examined 44 studies consist of 18 Ph.D. dissertations, 15 master's theses, and 11 articles. 40 of these studies were published in Turkish and four were published in English.

### Data Collection and Analysis Procedure

The research data was collected using the "Article Classification Form" developed by Sözbilir and Kutu (2008) and updated by Ozan and Köse (2014). Various additions and updates have been made in the updated version of the related form on the research subject, method, data collection instruments, and sample titles. For the content validity of the form, the researchers consulted three experts in the Program and Instruction field and calculated the Cohen Kappa coefficient as 0.83 for consistency among experts. As it is known, the Kappa coefficient of adaptation varies between -1 and +1, and as it approaches +1, it shows that the random fit decreases in terms of consistency between experts (Fleiss, & Cohen 1973). Although the relevant form is prepared under the title of "Article Classification Form", it takes into consideration the titles such as research method, data collection instruments, sampling, etc., which will allow the examination of research reports such as articles, theses, or full-text presentations.

In this context, documents written between 2010 and 2019 in Turkey consisting of articles master's and Ph.D. dissertations on program evaluation were examined in various dimensions by using the relevant form. Creswell (2012) stated that documents are one of the important data sources, especially for qualitative research. The collection of the relevant studies and the selection of the ones to be used in the study and the analysis of the data were carried out in approximately 3 months. In addition to the dimensions in the form, the types, and levels of the programs in the studies, the program evaluation models used, the results achieved in terms of the quality of the evaluated programs, and the suggested decisions about the evaluated programs were also examined. The citation information of the reviewed studies is given at the end of the study under the title of "Examined Research Papers" in addition to the references. Moreover, the content analysis technique was used in data analysis. Content analysis refers to the organization and conceptualization of data within certain themes in terms of various features (Yıldırım & Şimşek, 2013). In this context, the research data was coded using the NVivo 10 application and classified into themes. Some descriptive statistics of the codes and themes that emerged as a result of the content analysis are shown in Table 1.

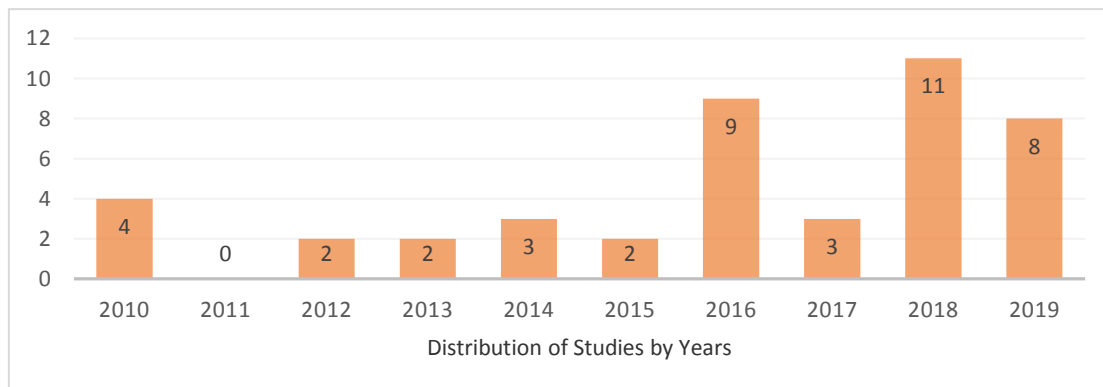
**Table 1.**  
Descriptive statistics of the codes and themes

<b>Emerged Themes</b>	<b>f</b>	<b>%</b>
Distribution of the Studies by Years	9	7,8
Program evaluation models	13	11,2
Preference reason of models	18	15,5
Evaluated program type and level	9	7,8
Research methods and designs	10	8,6
Sampling methods	17	14,7
Data collection tools	8	6,9
Results achieved in terms of the quality of the programs	27	23,3
Suggested decisions regarding evaluated programs	5	4,3
<b>Total</b>	<b>116</b>	<b>100,0%</b>

The frequency and percentage of the total number of codes for each main theme were given in Table 1. Explanations regarding sub-themes and codes under these themes were detailed under the findings title.

### Findings

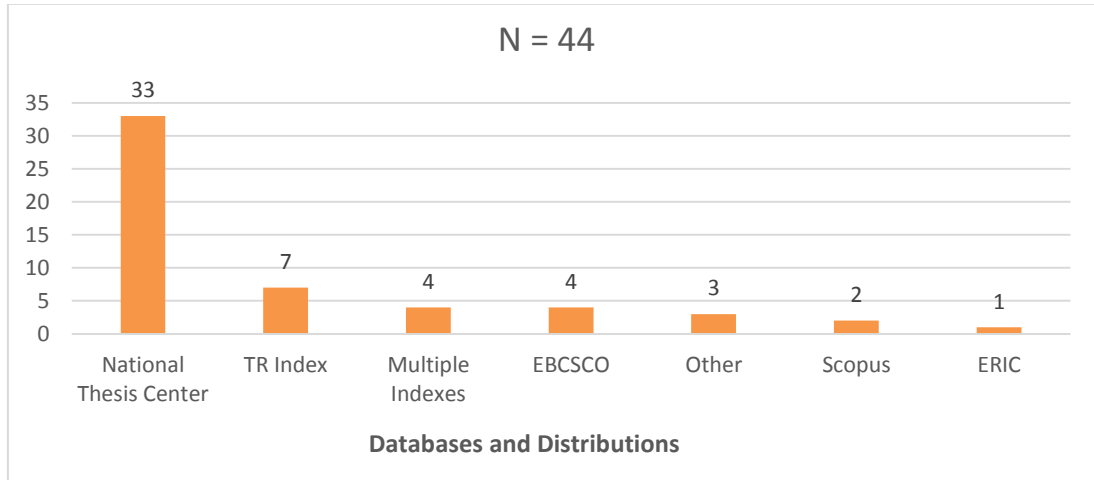
Information on the distribution of program evaluation studies conducted in Turkey using any program evaluation model between 2010 and 2019 was presented in Graphic 2.



**Graphic 2.** Distribution of the Studies Examined by Years

Graphic 2 shows that there were a small increase and decrease in the number of program evaluation studies from 2010 to 2015. In 2011, there was no program evaluation study using any program evaluation model. In program evaluation studies in Turkey, it was observed that studies taking a certain

evaluation model as a criterion as of 2016 have become more preferred. The distribution of the analyzed studies according to the databases in which they were indexed was given in Graphic 3.

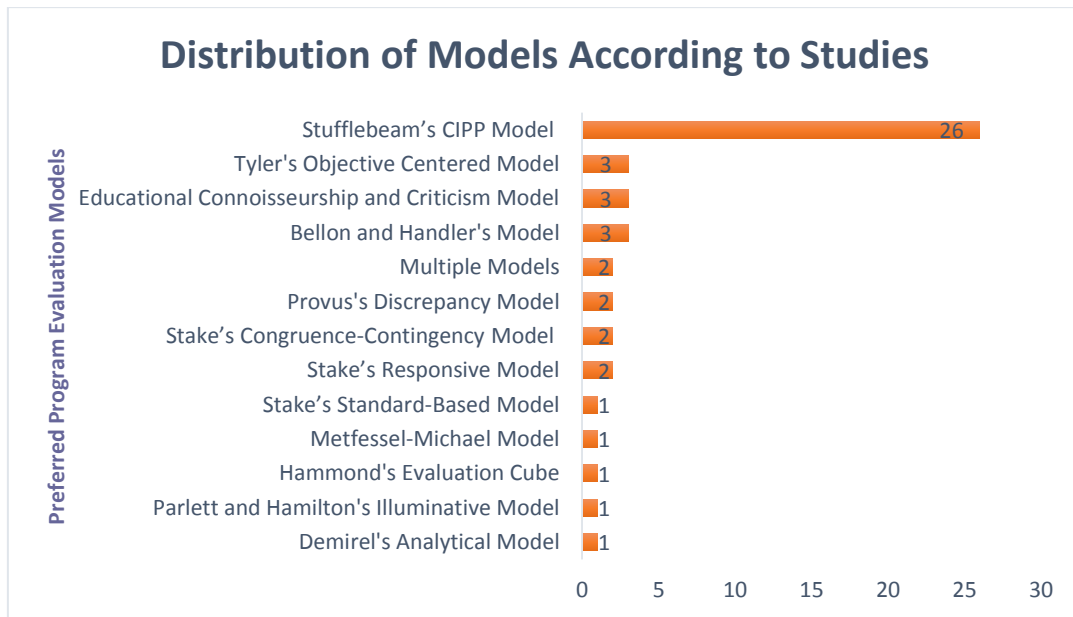


**Graphic 3.** Distribution of Studies according to the Databases in which they were Published

Graphic 3 shows that 33 of the studies conducted under the title of program evaluation were indexed in the National Thesis Center. Besides, seven of the reviewed articles were indexed in TR index, four in EBSCO, two in Scopus, and one in ERIC. Four studies were indexed in more than one database and three studies were indexed in databases not accepted as Higher Education Council Educational Sciences Index. Moreover, four studies in more than one database were indexed simultaneously in EBSCO, Scopus, and ERIC. In other words, the number of studies indexed in international databases is four.

#### The Program Evaluation Models Used in the Studies and the Reasons for the Choice of Related Models

The distribution of program evaluation models used in the studies examined was shown in Graphic 4.



**Graphic 4.** Distribution of Program Evaluation Models Used in Studies



Graphic 4 shows that 12 different evaluation models were used in related studies. No explanation was made regarding the reason for preference of the evaluation model used in 31 of the 44 program evaluation studies. Besides, it was observed that there was a disproportionate distribution in the program evaluation models preferred in the studies. Stufflebeam's CIPP Evaluation Model was preferred in most of the studies. Besides this model, the most preferred models were Tyler's Objective Centered Evaluation Model, Educational Connoisseurship and Criticism Model, and Bellon and Handler's Model. Moreover, multiple program evaluation models were used in two studies. Common reasons for the choice of program evaluation models used in the reviewed studies were shown in Table 2.

**Table 2.**  
Codes Regarding the Preference Reasons of Program Evaluation Models

Preference Reasons	Evaluation Models									
	<i>Stufflebeam's CIPP Model</i>	<i>Demirel's Analytical Model</i>	<i>Parlett and Hamilton's Illuminative Model</i>	<i>Hammond's Evaluation Cube</i>	<i>Metfessel-Michael Model</i>	<i>Bellon and Handler's Model</i>	<i>Provus's Discrepancy Model</i>	<i>Educational Connoisseurship and Criticism Model</i>	<i>Tyler's Objective Centered Model</i>	<i>Stake's Approaches and Models</i>
Description Success			+			+				
Holistic Approach	+									
Multidimensionality	+									
Development Orientation	+									
Providing Systematic Data	+									
Ease of Application	+									
Detailed Evaluation Opportunity		+								
Stakeholder Engagement		+								
Results Reflected in the Program		+								
Control of the Objectives				+						
Compliance with Program Type					+					
Context Evaluation Success			+							
Providing Data Diversity			+							
Comparison Success							+			
No Justification								+	+	+
<b>TOTAL</b>	<b>5</b>	<b>3</b>	<b>3</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>

When Table 2 was examined, it was observed that 14 common reasons for the choice of program evaluation models had been mentioned. No justification for the preference of Educational Connoisseurship and Criticism Model, Tyler's Objective Centered Evaluation Model, and Stake's Approaches and Models was presented. Information on the basic features of the program evaluation models used in the studies and detailed explanations regarding the reasons for being preferred are as follows:

Addressing a wide range of users such as program evaluators, program experts, researchers, developers, politicians, leaders, administrators, committees, and study groups, Stufflebeam's CIPP Evaluation Model aims to evaluate programs in four dimensions: context, input, process, and product (Stufflebeam, & Coryn, 2014). The main reasons for choosing the CIPP Evaluation Model in the studies examined are as follows (Duman, 2019; İlhan, 2018; Orhan, 2016; Aslan, 2015; Yurdakul, Uslu, Çakar, & Yıldız, 2014; Dinçer, 2013; Tekmen, 2012):

- It focuses on both the result and the process and allows for multiple evaluations,
- It focuses on developing the program, not proving the success of it,

- It facilitates the evaluator in determining potential questions that may arise during the evaluation process,
- It allows presenting comprehensive and systematic evaluation data,
- It allows the evaluation of the program in different dimensions.

Analytical Program Evaluation Model developed by Demirel has two data sources: program analysis and stakeholder opinions. In the first dimension, after the analysis of the program in the design, context and needs analysis stages, the analysis continues in the input, process, and product stages. In the second dimension, there are the opinions of stakeholders beginning from the field and program experts to non-governmental organizations (Demirel, 2014). Kuzu (2015) stated in his study that the reason for using the Analytical Program Evaluation Model is that the related model allows the program to take the views of all stakeholders and to examine the program in detail so that the evaluation results reveal the reflections of the course concretely.

Parlett and Hamilton's Illuminative Model, the aim is to evaluate how a new program works, based on the forms of practice in different schools. This model tries to reveal the participation levels of the administrators, students, and teachers in the program and the contribution of resources such as materials and tools to the program (Parlett & Hamilton, 1972). In the study in which the relevant model was used, the reason for the preference of the model was presented as its success in evaluating the context, revealing the strengths and weaknesses of the program, and allowing the variety of data collection instruments (Özüdoğru, 2016).

Hammond's Evaluation Cube is an evaluation model expressed in the form of a three-dimensional cube consisting of behavioral, instructional, and institutional dimensions developed to enable program evaluation at a local level (Hammond, 1967). This model was used only in one study. Hıdıroğlu, Kandemir, and Tuncel (2016) stated that the strong side of the model and the reason for the preference is the consideration of the objectives achieved during the evaluation, as well as those that are not achieved.

In the late 1960s, Metfessel and Michael proposed a program evaluation model, both as an alternative to Tyler's Objective Centered Model and that would take the Objective/Program-Oriented Evaluation Approach further (Metfessel and Michael 1967). The difference of the Metfessel and Michael Evaluation Model from Tyler's model is that it attaches more importance to alternative measurement tools and compares performance with standards instead of objectives (Yüksel & Sağlam, 2014). In the study using Metfessel and Michael Evaluation Model, it was stated that the relevant model was a suitable model for the evaluation of higher education programs (Bayık, 2018).

The Bellon and Handler Evaluation Model has three main elements and four focus areas. The three main elements of the model are situation description, analysis activities, and program improvement. Four focus areas refer to the goals, organizations, operations, and outcomes areas of the program (Mutlu, 2018 cited in Bellon and Handler). Among the examined studies, detailed information about the model was provided in the studies using this model (Mutlu, 2018; Karcı Aktaş, 2018). However, as the reason for the preference of the model, it was stated in only one study that this model allows to present the current situation of the program and to provide effective suggestions for its improvement (Şişman, 2019).

Provus's Discrepancy Model focuses on the objectives in the evaluation of the program. There are studies conducted by Provus to evaluate the programs of different schools in Pittsburgh between 1969 and 1971 (Fitzpatrick et al., 2011). As a result of these studies, it set out from the assumption that the systems of many schools, especially public schools, were established without adequate planning, and defined program evaluation as the process of revealing the difference between goals and specified standards (Provus, 1969). The results of these studies showed that the systems of many schools, especially public schools, were established without adequate planning, thus program evaluation was defined as the process of revealing the difference between goals and established standards (Provus, 1969). In one of the studies using the relevant model, there is no explanation as to why the model was

preferred. Babacan (2016), on the other hand, stated that she chose this model to reveal the difference between the current program and what should be by taking holistic education as a criterion.

The main feature of Eisner's Educational Connoisseurship and Criticism Model is to produce a rich description of educational life as a result of the program (Ornstein & Hunkins, 2018). Eisner expresses educational connoisseurship as the art of appreciating and emphasizes that the expert should have a critical role. It is possible to mention four basic dimensions of Eisner's model as descriptive, interpretative, evaluative, and theming (Eisner, 2002). In the examined studies, issues were explained such as the general characteristics of the model, and the methodological requirements to be considered in different dimensions of the evaluation. However, it was not stated why the use of this model was necessary for the programs evaluated, rather it was emphasized whether the relevant model was used in Turkey before (Kumral, 2010; Çelik, 2018).

Tyler's studies between 1930 and 1945 have a great impact on education and the evaluation of education. Tyler focused on the achievement of objectives at the end of educational activities in The Tylerian Evaluation Approach and stated that the best way to measure this is through achievement tests (Stufflebeam, et al., 2000). In the studies examined, information about the general features of Tyler's model and the basic stages that should be considered in the program evaluation processes were provided. However, it was not explained why this model was preferred in the evaluation of the relevant programs (Kotluk, & Yayla 2016; Aslan, & Çıkar, 2017; Aslan, & Erden, 2018).

Stake's Responsive Evaluation Approach is directly expressed as a model in some studies. Participant-Oriented Program Evaluation includes formal and informal data collected from many different individuals that were used in the evaluation. In this approach, to avoid this confusion of data obtained from many stakeholders and to make the assessment more valid and reliable, Stake has developed a guiding action table consisting of 12 stages in the form of a clock (Stufflebeam, et al.2000). Stake's Congruence-Contingency Evaluation Model, on the other hand, is a participant-oriented model that aims to conceptualize the program evaluation processes to make them understandable (Stake, 1967). Key concepts of the model are "describing" the situation of the program and "making decisions" about it. In the model, program evaluation is conducted in input process and product dimensions (Yüksel, & Sağlam, 2014). Stake states that in the Standard-Based Evaluation Model, standards that will optimize the program should be used in the evaluation process, and these standards should be accurate and comparative standards of different programs (Stake, 2004). In the program evaluation studies using Stake's approaches and models, the features of the relevant approaches and models were explained in detail (Yeşilyurt, 2010; Bayat, 2012; Altındağ, 2017). However, the issue of why these models were preferred or whether they are suitable for the type of program being evaluated has not been discussed.

### Distribution of Program Evaluation Models According to Program Types and Levels

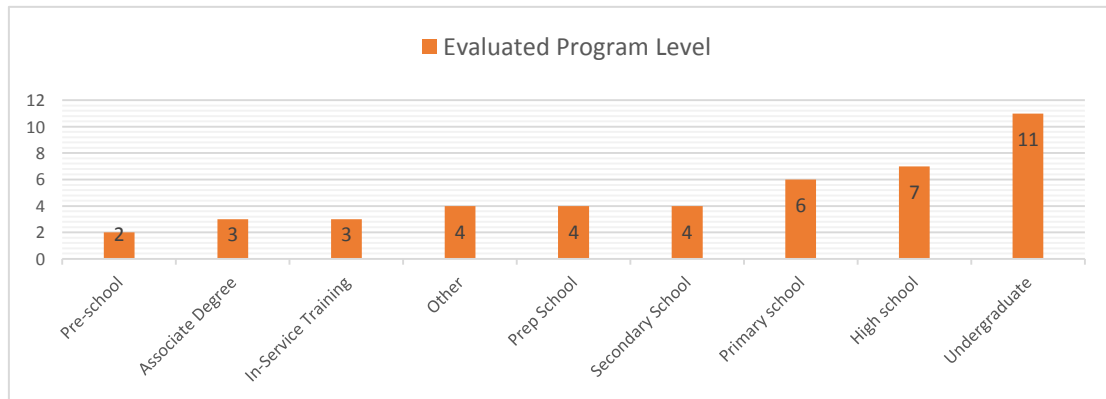
The distribution of the program evaluation models general characteristics of which were given above, according to program types and levels were shown in Table 3 and Graphic 5.

**Table 3.**

Distribution of Program Evaluation Models According to Program Types and Levels

Pre-school	Preschool Program <i>Provus's Discrepancy Model</i>	Preschool Program <i>Stufflebeam's CIPP Evaluation Model</i>		
Primary school	4 <sup>th</sup> Grade Mathematics Program <i>Tyler's Objective Centered Evaluation</i>	4 <sup>th</sup> Grade English Program <i>Stufflebeam's CIPP Evaluation Model</i>	2 <sup>nd</sup> Grade English Program <i>Parlett and Hamilton's Illuminative Model</i>	Primary Reading and Writing Program <i>Stake's Congruence-Contingency Evaluation Model</i>
	4 <sup>th</sup> Grade Turkish Program <i>Stufflebeam's CIPP Evaluation Model</i>	3 <sup>rd</sup> Grade Science Program <i>Stufflebeam's CIPP Evaluation Model</i>		

Secondary School	5 <sup>th</sup> Grade Science Program <i>Tyler's Objective Centered Evaluation</i>	7 <sup>th</sup> Grade Turkish and Mathematics Program <i>Stufflebeam's CIPP Evaluation Model</i>	5 <sup>th</sup> Grade Mathematics Program <i>Stake's Congruence-Contingency Evaluation Model</i>	7 <sup>th</sup> Grade English Program <i>Stufflebeam's CIPP Evaluation Model</i>
High school	9 <sup>th</sup> Grade Physics Program <i>Tyler's Objective Centered Evaluation</i>	High School English Program <i>Educational Connoisseurship and Criticism Model</i>	High School Social Sciences Program <i>Stufflebeam's CIPP Evaluation Model</i>	9 <sup>th</sup> Grade Mathematics Program <i>Stufflebeam's CIPP Evaluation Model</i>
	9 <sup>th</sup> Grade English Program <i>Provus's Discrepancy Model</i>	Abitur Program <i>Educational Connoisseurship and Criticism Model</i>	9th Grade Turkish Literature Program <i>Stufflebeam's CIPP Evaluation Model</i>	
Associate Degree	Hair Care and Beauty Services Program <i>Metfessel-Michael Evaluation Model</i>	Air Petty Officer Vocational School English Program <i>Stufflebeam's CIPP Evaluation Model</i>	Associate Degree Tourism Program <i>Stufflebeam's CIPP Evaluation Model</i>	
Prep school	English Prep School Basic Course Program <i>Bellon ve Handler's Evaluation Model</i>	Prep School English Program <i>Stufflebeam's CIPP Evaluation Model</i>	Prep School English Program <i>Bellon ve Handler's Evaluation Model</i>	Prep School English Program <i>Bellon ve Handler's Evaluation Model</i>
Undergraduate	Teaching Practice Program <i>Stufflebeam's CIPP Evaluation Model</i>	Primary School Undergraduate Program <i>Stufflebeam's CIPP Evaluation Model</i>	Undergraduate Core Program <i>Stufflebeam's CIPP Evaluation Model</i>	Teaching Principles and Methods Program <i>Demirel's Analytical Evaluation Model</i>
	Teaching Practice Program <i>Stufflebeam's CIPP Evaluation Model</i>	Pre-School Undergraduate Program <i>Stufflebeam's CIPP Evaluation Model</i>	Teaching Practice Program <i>Stake's Standard Based ve Stake's Responsive Models</i>	Primary School Undergraduate Program <i>Educational Connoisseurship and Criticism Model</i>
	Teaching Profession Courses Program <i>Stufflebeam's CIPP Evaluation Model</i>	Interpersonal Relations and Therapeutic Communication Program <i>Stufflebeam's CIPP Evaluation Model</i>	Teaching Principles and Methods Program <i>Hammond'un Evaluation Cube</i>	
In-Service Training	Web-Based Content Development Program <i>Stufflebeam's CIPP Evaluation Model</i>	Call Center Professional Competence Development Program <i>Stufflebeam's CIPP Evaluation Model</i>	Intel Teacher Program (ITP) Program <i>Stufflebeam's CIPP Evaluation Model</i>	
Other	Birth Training Program <i>Stake's Responsive and Stufflebeam's CIPP Evaluation Models</i>	Technology Apps in Primary Education e-Certificate Program <i>Stufflebeam's CIPP Evaluation Model</i>	Distance Education Foreign Language Program <i>Stufflebeam's CIPP Evaluation Model</i>	Turkish Program <i>Stufflebeam's CIPP Evaluation Model</i>

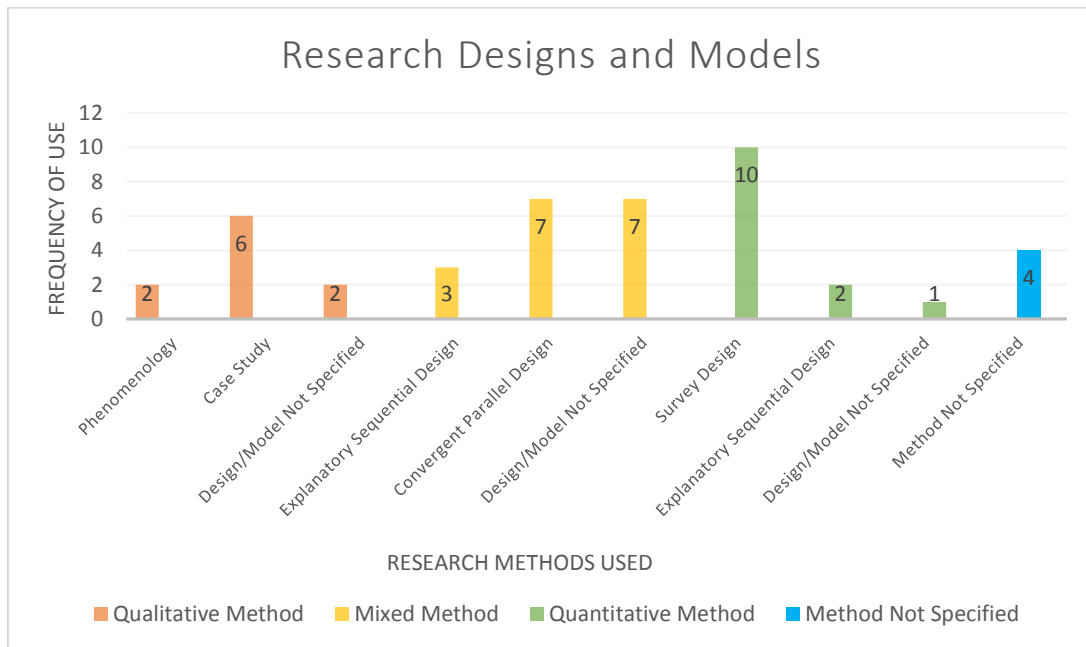


Graphic 5. The Levels of the Programs Evaluated in the Studies

When Table 3 and Graphic 5 were examined, it was observed that the most evaluated program types in evaluation processes were undergraduate and high school programs. The most evaluated undergraduate programs belonged to faculties of education. Besides, there was no study evaluating postgraduate programs, while pre-school programs were the least evaluated programs. When the distribution of evaluation models used according to program levels was examined, it is possible to express model variation in the evaluation of primary and high school programs. Moreover, Bellon and Handler's Evaluation Model was widely used in the evaluation of foreign language preparation programs. However, the only model used in the evaluation of in-service training programs was Stufflebeam's CIPP Evaluation Model. Moreover, various programs such as personal development, certificate programs, and distance education programs that could not be positioned at any level were predominantly evaluated with Stufflebeam's CIPP Evaluation Model, under the title of "other".

**Research Methods, Designs, and Models Used in Studies**

Findings regarding the research methods, designs, and models preferred in the examined studies were shown in Graphic 6.

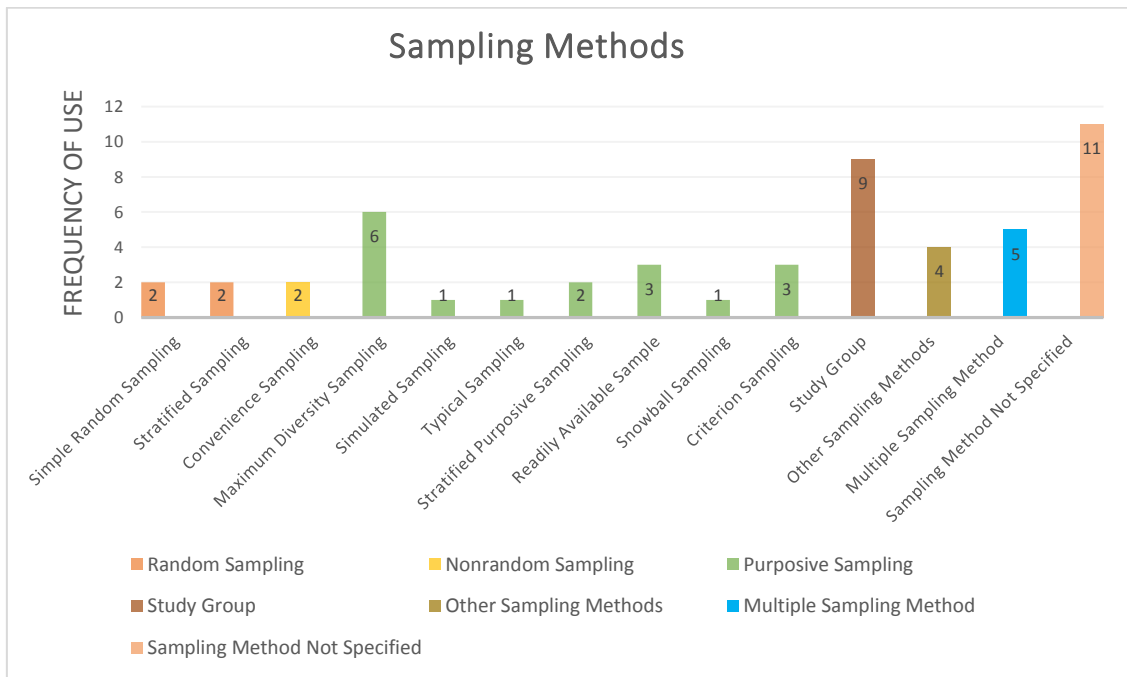


**Graphic 6.** Research Designs/Models Used in the Studies

Graphic 6 shows that researchers mostly preferred mixed methods in the studies, and the most common designs and models were the survey model and the convergent parallel design. Besides, some studies did not specify any method, design, or model, while in others, only qualitative, quantitative, and mixed-method expressions were used.

**Sampling Methods Used in Studies**

Findings regarding the sampling methods of the studies were shown in Graphic 7.

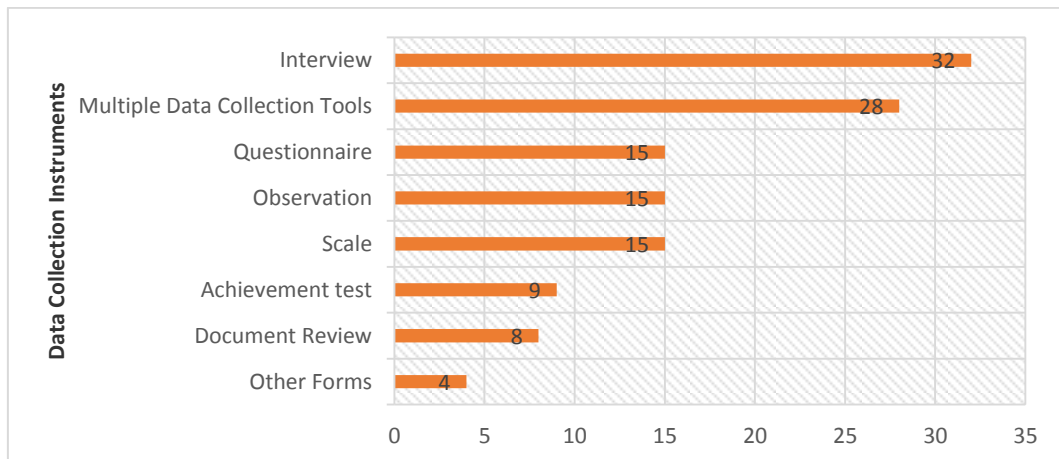


**Graphic 7. Sampling Methods Used in the Studies**

Graphic 7 shows that non-random sampling methods, especially purposive sampling, were commonly preferred in studies. Besides, the most widely used purposive sampling method was maximum diversity sampling. In seven of the examined studies, all participants were tried to be included in the study. Besides, in 11 of the 44 studies examined, no sampling method was mentioned.

#### Data Collection Instruments in Studies

Graphic 8 shows the data collection instruments used in the studies.



**Graphic 8. Data Collection Instruments Used in Studies**

Graph 8 shows that more than one data collection tool was used in 28 of the studies. The most commonly used data collection tools were interviews, questionnaires, observations, and scales.

#### Results of the Studies in Terms of the Quality of the Programs

Descriptive statistics regarding the common results about the programs evaluated as a result of the content analysis were presented in Table 4.

**Table 4.**  
Descriptive Statistics Regarding Results on Behalf of the Quality of Programs in Studies

Common Results Achieved	Positive Results		Negative Results	
	f	%	f	%
Serving the Purpose	15	25,0	18	12,6
Considering Individual Characteristics	2	3,3	4	2,8
Acquiring Affective Skills	6	10,0	2	1,4
Number and Quality of Educators	2	3,3	8	5,6
Physical Conditions	3	5,0	13	9,1
Quality of General and Specific Objectives	4	6,7	6	4,2
Mission Description of Stakeholders			1	0,7
Readiness of Students			3	2,1
Quality of Program Contents	1	1,7	9	6,3
Educator, Administrator and Student Interaction			5	3,5
Institutional Differences			4	2,8
Quality of Equipment and Materials			8	5,6
Students' Development Status	7	11,7	7	4,9
Student Profile			3	2,1
Program Appropriateness for Students	4	6,7		
Quality of Measurement and Evaluation	2	3,3	10	7,0
Quality of Instructional Design	4	6,7		
Autonomy and Flexibility of the Program			2	1,4
Stakeholder Engagement	1	1,7	1	0,7
Appropriateness of the Program Philosophy	2	3,3		
Information Basis About the Program			6	4,2
Classroom Population			5	3,5
Duration of Lessons	2	3,3	8	5,6
Application Association			2	1,4
Applied Training			5	3,5
Program Applicability	2	3,3	7	4,9
Quality of the Methods and Techniques	3	5,0	6	4,2
<b>Total</b>	<b>60</b>	<b>100,0%</b>	<b>143</b>	<b>100,0%</b>

Table 4 provides information on various results of the studies regarding the quality of the programs. The codes determined during the content analysis show that both positive and negative results were achieved in many dimensions of the programs evaluated. This situation is because factors such as the program's serving its purpose, students' development status, the quality of the methods and techniques were evaluated positively in some studies and negatively in others. While the number of coding of negative evaluations regarding the quality of the programs was 143, it was 60 for positive evaluations. The most common positive evaluations about the quality of the programs were, in order, "Serving the Purpose, Students' Development Status and Acquiring Affective Skills". The most common negative evaluations were "Serving the Purpose, Physical Conditions and Quality of Measurement and Evaluation" respectively. Besides, "Quality of Program Contents, Quality of Equipment and Materials, Duration of Lessons and Number and Quality of Educators" were among the common negative results.

#### Decision Suggestions About the Programs in Studies

The studies show that there were various decision suggestions regarding the programs evaluated. In this context, Table 5 shows the descriptive statistics regarding the suggested decisions about the programs evaluated.

**Table 5.**  
Descriptive Statistics on Suggested Decisions About Evaluated Programs

Proposed Decisions	f	%
It could be maintained with a comprehensive update.	15	34,1%
It could be maintained with a partial update.	21	47,7%
No suggestion was made.	5	11,4%
The program failed, must be redeveloped.	2	4,5%
The program is successful, it should be continued as it is.	1	2,3%
<b>Total</b>	<b>44</b>	<b>100,0%</b>

Table 5 contains suggestions for developing or ending the evaluated programs. The most frequently recommended situation about evaluated programs was to maintain with partial updates. This proposal was followed by suggestions such as " It could be maintained with a partial update, the program failed, must be redeveloped and the program is successful, it should be continued as it is". Besides, no suggestions were made for the program evaluated in five studies.

### Conclusion, Discussion, and Suggestions

The research findings show that the studies conducted under the title of program evaluation between 2010 and 2019 in Turkey were carried out largely without taking into account the program evaluation approaches and models. Especially, program evaluation models were used in only 44 of the 176 studies accessed from various databases. This corresponds to 25% of the studies conducted under the title of program evaluation. The criteria for the success or failure of the program in studies conducted without using program evaluation approaches and models are open to discussion. Stufflebeam (1999) stated that such evaluations, which he named *Pseudoevaluations*, are generally evaluations that express positive or negative opinions about the program, regardless of the actual benefit and value of it. Stufflebeam emphasizes that the evaluations, which he named as *Question/Method-Oriented Evaluations*, try to evaluate the program based on some standard questions and methods, and whether they were successful or not is open to discussion. This situation raises the question of whether the studies are successful in program evaluation that aims to evaluate the program according to the views of a certain group, which are frequently preferred under the title of program evaluation in Turkey. Stufflebeam (1999) stated that the most effective evaluations could be carried out through *Improvement and Accountability-Oriented Evaluation Approaches*. The main focus of these approaches in evaluating is whether the program is successful, and they are comprehensive and expensive evaluations that take into account all the questions and criteria required to evaluate the success of the program. Such evaluations are only possible with program evaluation models to be used within the framework of certain approaches. Although the use of models in the studies conducted under the title of program evaluation corresponds to a small percentage, there have been positive developments in the use of models in program evaluation studies in the 10 years. When the 10-year period in which the relevant researches were examined is taken into consideration in two parts as the first and the second 5 years, it is possible to see that the number of program evaluation studies was 11 in the first 5 years, while this number increased to 35 in the second 5-year period. Especially as of 2016, studies in which a certain program evaluation model was used as a criterion became more preferred. Despite the increase in the number of program evaluation studies, only four of the studies were indexed in international databases. This situation makes program evaluation studies questionable in terms of quality.

According to research findings, while there are various evaluations on the programs of formal education institutions such as undergraduate, primary, secondary, and high schools, a limited number of program evaluation studies were conducted for preschool programs. Besides, there is no program evaluation study for graduate programs. Stufflebeam's Context, Input, Process, and Product Program Evaluation Model was preferred in most of the studies (approximately 59%). Studies examining program evaluation studies in Turkey in the past years also express this situation. Kurt and Erdoğan (2015) stated



in their study that Stufflebeam's CIPP Model is the most preferred one in program evaluation studies. Özüdoğru (2018) found a similar finding that the most preferred model in studies evaluating foreign language education programs is the CIPP Model. Stufflebeam's CIPP Evaluation Model is among the most preferred models also in the international literature (Warju, 2016; Zhang, Et al. 2011). There is a perception that the CIPP Model provides ease of application to different types of researchers, and even the dynamics of its construction are related to this (Stufflebeam, & Coryn 2014). Besides, among the models used in the studies, there is no model designed in Turkey other than Demirel's Analytical Evaluation Model, and the related model was preferred in only one study. However, most of the relevant studies do not explicitly state whether the selection of this model and other models is based on decisions that consider the context of the program. For example, a variety of models were used to evaluate primary and high school programs in the studies examined. However, since it is not clearly stated why these models are preferred, it is not possible to understand the actual reason for using these models. Another example of criterion uncertainty in model selection is that Stufflebeam's Program Evaluation Model is the only model used in evaluating in-service training programs. Nevertheless, there are models such as the Kirkpatrick's Model used in the evaluation of in-service training and adult education programs (Smidt, Balandin, Sigafos, & Reed, 2009; Uşun, 2012). At this point, the researcher's preference is important in choosing the relevant approach and model. However, it is necessary for the transparency of the study to decide on the approach and model to be used in the evaluation of any program, following its context, and to report all these processes.

According to the research findings, mostly mixed-method and multiple sampling and data collection processes were preferred in program evaluation studies. Especially in social sciences, studies that use more than one data collection tool called triangulation are considered valuable because they better express the richness and complexity of human character (Cohen, Manion, & Morrison, 2007). In this context, the most preferred data collection tools are qualitative data collection tools such as interviews and observations. This situation shows that program evaluation studies have a wealth of data collected from different samples, and the emphasis is placed on qualitative data aiming to examine human characteristics in-depth. In some of the program studies, on the contrary to the findings stated above, there was no explanation regarding the research and sampling methods or an effort was made to evaluate the program with only one data collection tool. This situation may be related to the case of not choosing the comprehensive studies mentioned above. In fact, it is possible to evaluate the deficiencies in the research design and sample selection as methodological errors. Different studies mention the prevalence of such methodological errors (Toy & Tosunoğlu, 2007; Erdoğan, 2001). Methodological errors and not preferring comprehensive studies in scientific research are undoubtedly the situations that negatively affect the quality of the studies.

Content analysis indicates that there are various results in the studies regarding the programs evaluated and various suggestions presented for the decision-makers about the programs based on these results. The content analysis results show that there are 60 positive and 143 negative evaluations about the programs evaluated. This situation shows that the evaluated programs contain major problems. There are negative results related to these programs such as the inadequacy of physical conditions and the quality of the measurement and evaluation activities. The development status of the students and the programs that provide various affective characteristics are among the positive evaluation results. Whether the programs serve their purpose or not has been the most emphasized issue both positively and negatively. This situation is a crucial issue involved as the standard field called Utility Standards at the top of the list of standards developed by JCSEE (JCSEE, 2018). In this regard, 15 positive and 18 negative results emerged regarding the programs. This is an indicator that the evaluated programs are frequently problematic in serving their purpose. Moreover, there are various suggestions on deciding the future status of the programs based on the research results. These suggestions are mostly for the continuation of the programs with partial updates. However, at this point, it is a debatable issue how the programs that receive a lot of negative criticism compared to their positive aspects can become effective enough with partial updates.

The rapid development and change in the World show its effects also on the development and evaluation processes of the programs as in every field. It is argued that the breakthroughs made by countries such as Brazil, China, India, and South Africa in recent years will also be reflected in the education systems and this might cause a change in the mentality of program evaluation in the following years (Stockmann & Meyer, 2016). At this point, it has become an important issue to meet certain standards in the evaluation as well as the use of approaches and models in program evaluation to keep up with the changes and developments in countries such as America and Europe (Astbury, 2016). The most important example of this is the studies conducted by the Joint Committee on Standards for Educational Evaluation which has managed to draw attention to the program evaluation standards it has developed with more than 400 stakeholders from universities in different countries (JCSEE, 2018). The relevant committee has been developing standards to guarantee the quality of program evaluation studies since 1981 and updates these standards according to current needs (Stufflebeam & Madaus, 1983). The standards developed by the committee have been transformed into checklists to evaluate the quality of program evaluation processes by program experts and scientists such as Daniel Stufflebeam, Robert Stake, Michael Scriven, Michael Quinn Patton, Egon Guba, Ernest House under the name of Checklist Project (WMU, 2020). Research results show that program evaluation studies in Turkey and standardized applications in these studies are becoming more common. There have been various changes in the education system in 2012 such as the transition to the 12-year compulsory education system and the implementation of the FATİH Project in the same year in Turkey (Akıncı, 2017; MEB, 2012). These changes and updates might have been perceived as radical changes in the education system which might cause an interest in program evaluation studies. This may be one of the reasons that the number of program evaluation studies increased approximately three times in the second five years compared to the first five years.

Not using any model designed in Turkey other than Demirel's Analytical Evaluation Model in the program evaluation studies is open to discussion as this is an indication that the efficiency in the production of new approaches and models in program evaluation is very low and the studies produced in Turkey in this context are not respected enough. It would not be wrong to mention similar situations in terms of methodological preferences. Especially in some studies, methodological choices and meaningful data productivity for the program are quite limited. In this context, in correspondence with the results of the research, similar situations are mentioned in terms of methodological preferences in the studies in the style of trend and content analysis in the field of education and program (Kozikoğlu & Senemoğlu, 2015; Ozan & Köse, 2014; Bıkmaz, Aksoy, Tatar, & Altınyüzük, 2013). This situation shows that researchers have not preferred program evaluation studies that require comprehensive data collection and analysis processes over the years. Besides, many studies in the relevant literature in Turkey are not based on any approach or model, although they are under the title of program evaluation. Among the studies that use any program evaluation approach and model, the ones indexed in international databases are limited. Ak and Gülmez (2006) state that in some national and international academic studies, researchers do not prefer comprehensive data collection and analysis processes, and these studies are sometimes used to collect points. The fact that the researchers do not explain the reasons for using the relevant models in most of their studies using any model, the deficiencies in the evaluation of the program, and the inconsistencies regarding the decisions to be taken about the program show that the concept of program evaluation is not adequately understood. Uşun (2012) argues that one of the most important purposes of program evaluation is deciding whether it is qualified enough. This is an indication that not every study conducted under this title, regardless of their number, is a program evaluation research. Akıncı and Köse (2020) obtained similar results in their meta-evaluation study on teacher training programs in Turkey. The fact that the majority of the program evaluation studies examined consist of master's and Ph.D. dissertations shows that researchers tend to prefer program evaluation studies in more comprehensive researches that require time and effort. All these problems may be related to the influence of academic publishing culture in the context of the research approach on methodological tendencies, and preferences. Akcan, Malkoç, and Kızıltan (2018) stated that academic culture is education-oriented rather than scientific productivity in Turkey, and there are serious problems in the research approach. This is an indication that, in program evaluation

studies, researchers prefer to work on more easily accessible sample groups such as undergraduate students of education faculties.

Consequently, there are program evaluation studies with their shortcomings benefiting from program evaluation approaches and models, and a trend towards new studies in this context. Therefore, based on the results of the study, some suggestions that may contribute to future program evaluation studies are as follows:

- In line with the developments in the world, program evaluation approaches and models should be developed under the needs of the Turkish education system. In this context, there should be more effective criteria in the selection of the relevant models in studies, and the preference reasons should be clearly stated.
- Researchers should conduct program evaluation studies according to the program evaluation standards in the international literature or according to the current standards developed under the needs of the Turkish education system.
- The Council of Higher Education and the Ministry of National Education should take more responsibility for program evaluation, and these institutions should carry out comprehensive program evaluation studies.
- Researchers should conduct program evaluation studies on different program types and levels with more effective methodological choices, especially in graduate programs.
- Researchers should offer suggestions for decisions about the future of the program in line with the results of their studies.
- Different researchers should conduct meta-evaluation studies regarding the quality of program evaluation studies in various contexts.

## Türkçe Sürümü

### Giriş

Türkiye’de eğitim sistemi yıllardır sorunları ile anılmakta ve bu sorunlar değişik ortamlarda tartışılmaktadır. Uluslararası sınavlarda başarısızlık, üretim ve ekonomide yaşanan problemler, beyin göçü vb. konular bu duruma gerekçe olarak gösterilebilir (Bakioğlu ve Yıldız, 2014; Çalışkan, Karabacak ve Meçik, 2013; Akyüz, 2006; Işığışık, 2002). Eğitim sisteminin sorunlu olması eğitimde yapılan program geliştirme çalışmalarının da sorunlu olabileceğini akla getirmektedir. Şüphesiz eğitimde program geliştirme çalışmaları ve geliştirilen bu programların etkin bir biçimde değerlendirilmesi ve niteliklerinin ortaya konulması eğitim sisteminin başarısını doğrudan etkilemektedir. Yüksel ve Sağlam (2014)’a göre eğitimde program geliştirme, eğitim programlarının tasarlanması, uygulanması, değerlendirilmesi ve elde edilen bilgiler doğrultusunda yeniden düzenlenmesidir. Bu tanım geliştirilen programların niteliğinin belirlenmesinin önemine de vurgu yapmaktadır. Bu da ancak etkin bir biçimde yürütülen program değerlendirme faaliyetleri ile mümkündür.

Program değerlendirme, planlanan ve uygulanan bir programın, bilimsel araştırma süreçlerine tabi tutularak uygun program değerlendirme modelleri doğrultusunda değerlendirilip sonuçların programın geliştirilmesinde kullanılması olarak tanımlanmaktadır (Yüksel ve Sağlam 2014). Bir diğer tanımda ise program değerlendirme geliştirilen bir programın bilimsel bir anlayışla geçerli, yararlı, verimli, etkili vb. olup olmadığının belirlenmesi ve programın herhangi bir özelliği hakkında karar alma süreci olarak ifade edilmektedir (Uşun, 2012). Belirli bir sistematığı olan, amaçları, yeri, zamanı, muhatapları belli, geçerli ve güvenilir değerlendirmeler ise *formal değerlendirmeler* olarak tanımlanmaktadır. (Fitzpatrick, Sanders ve Worthen, 2011). Sistematik, geçerli ve güvenilir özellikleri ile biçimlendirici ve toplam değerlendirmelerin de alt yapısını oluşturan formal değerlendirmelerin bilimsel araştırma süreçlerinin dikkate alındığı önemli değerlendirmeler olduğu ifade edilebilir. Bununla birlikte formal program değerlendirme süreçleri nicel, nitel ya da karma olmak üzere farklı yöntemsel tercihlere de olanak tanımaktadır (Yüksel ve Sağlam, 2014; Uşun, 2012). Bu noktada yöntemsel tercihlerin, benimsenen yaklaşımlarla ilgili olduğu söylenebilir. Ornstein ve Hunkins (2018) program değerlendirme yaklaşımlarını bilimsel (modern) ve hümanist (post-modern) olarak ikiye ayırmıştır. Bilimsel (modern) yaklaşımlarda programı değerlendirecek kişiler açıkça belirtilen hedefleri ve öğrencilerinin programın amaçlanan sonuçlarına ulaşip ulaşmadığına dair kesin göstergeleri dikkate alırken hümanistler, istatistiksel yöntemler yerine: estetik, etnografi, otobiyografi, fenomenoloji, eleştirel okuryazarlık vb. yöntemleri tercih ederler. House (1990) ise çalışmasında tekilci ve çoğulcu yaklaşımlardan söz etmektedir. Türkiye’de yaygın olarak kullanılan program değerlendirme yaklaşımlarının sınıflandırılması ve bu sınıflandırmanın altında yer alan ortak modeller aşağıdaki gibidir (Aygören ve Er, 2018; Demirtaş, 2017; Yüksel ve Sağlam, 2014; Uşun, 2012; Fitzpatrick, vd., 2011):

- Hedef/Program Odaklı Program Değerlendirme Yaklaşımı
  - Tyler’ın Hedef Odaklı Değerlendirme Yaklaşımı
    - Metfessel-Michael Değerlendirme Modeli
    - Provus’un Çelişki/Fark Program Değerlendirme Modeli
- Uzman Odaklı Program Değerlendirme Yaklaşımı
  - Eisner’in Eğitsel Eleştiri ve Eğitsel Uzmanlık Modelleri
- Tüketici Odaklı Program Değerlendirme Yaklaşımı
- Sistem/Karar Odaklı Program Değerlendirme Yaklaşımı
  - Stufflebeam’in Bağlam, Girdi, Süreç ve Ürün (BGSÜ) Program Değerlendirme Modeli

▪ UCLA Değerlendirme Modeli

- Katılımcı Odaklı Program Değerlendirme Yaklaşımı
  - Stake'in Yanıtlayıcı/İhtiyaca Cevap Verici Program Değerlendirme Yaklaşımı
    - Stake'in Uygunluk ve Olasılık Program Değerlendirme Modeli

**Hedef/Program Odaklı Program Değerlendirme Yaklaşımı**

Hedef Odaklı Program Değerlendirme Yaklaşımı yürütülen eğitim-öğretim faaliyetlerinden sonra hedeflere ne düzeyde ulaşıldığına odaklanır. Bu yaklaşımda program değerlendiricinin temel rolü programın hedeflerinin ne kadarına ulaşıldığı ve ne kadar iyi ulaşıldığına karar vermektir (Fitzpatrick, vd., 2011). Bu yaklaşım programın hedeflerinin açık bir şekilde ifade edildiği durumlarda uygulama aşamaları da belirgin olduğundan karar vericilere kullanışlı bilgiler sunabilmesi adına başarılı olarak görülür. Fakat sadece amaçlara odaklanmanın amaçlarda belirtilmeyen program özelliklerini ihmal etme riskini doğurduğu düşünüldüğünden bu yönü ile eleştiri almaktadır (Fitzpatrick, vd., 2011; Stufflebeam, Madaus ve Kellaghan, 2000).

**Uzman Odaklı Program Değerlendirme Yaklaşımı**

Uzman Odaklı Program Değerlendirme Yaklaşımı kendinden önceki yaklaşımların aksine program değerlendirmede sisteme ekonomik anlamda katkı sağlayacak bireyler yetiştirmek temel odağı olmadığından hümanist ve post modern bir yaklaşım olarak görülmektedir (Ornstein ve Hunkins 2018). Bu yaklaşım program değerlendirmede alan uzmanlıklarına önem vermektedir (Yüksel ve Sağlam, 2014; Uşun, 2012).

Uzman Odaklı Program Değerlendirme Yaklaşımı hümanist bir yaklaşım olmasından kaynaklanan bireylere verdiği önem ve sunduğu akademik özgürlük ortamı adına önemli olarak görülse de odağında uzmanın yer alması ve bu uzmanların seçiminin yeterlilik ve nesnellik adına zor olması sebebiyle eleştiri almaktadır (Fitzpatrick, vd., 2011).

**Tüketici Odaklı Program Değerlendirme Yaklaşımı**

Tüketici odaklı değerlendirmeler ilk olarak 1960'ların ortasından sonlarına kadar eğitsel ürünlerin değerlendirmesinde önemli hale gelmiştir. Bunun en önemli sebebi ise o yıllarda Amerika Birleşik Devletleri (ABD) hükümetlerinin yeni eğitsel ürünler geliştirmek için ciddi fonlar sunmaya başlamasıdır (Fitzpatrick, vd., 2011). Bu yaklaşımın önde gelen temsilcisi Michael Scriven olarak görülmektedir. Scriven'in değerlendirmeye yönelik yaklaşımı, genel olarak, tüketicilerin ve geniş toplumun değerlendirilen ihtiyaçlarına göre yararlanabileceği alternatif programların ve eğitsel ürünlerin maliyet ve etkilerinin belirlenmesini ve derecelendirilmesini amaçlamakta, bunu yaparken de ürün değerlendirme kontrol listelerini kullanmaktadır (Stufflebeam ve Coryn 2014).

Tüketici Odaklı Program Değerlendirme Yaklaşımı eğitsel ürünlerin kalitesi hakkında farkındalığı artırması, nitelikli eğitsel ürünlerin ortaya konulmasını sağlaması ve programlara bazı standartları kazandırması konularında başarılı görülmektedir (Stufflebeam, vd., 2000). Bunun yanında maliyetlerin yüksek oluşu ve bazen gereksiz kuralcılıkların ortaya çıkması yaklaşımın zayıf yönleri olarak görülmektedir (Fitzpatrick, vd., 2011).

**Sistem/Karar Odaklı Program Değerlendirme Yaklaşımı**

Sistem/Karar Odaklı Program Değerlendirme Yaklaşımı 1960'lı yılların sonunda eleştirilmeye başlanan özellikle hedef odaklı program değerlendirme yaklaşım ve modellerine alternatif olarak ortaya çıkmıştır. Bu yönüyle programın hedeflerinden çok yöneticilerin alacağı kararlara odaklanmıştır. Bu yaklaşımın programın farklı boyutlarına sistematik ve planlı yaklaşımı olumlu yanı olarak görülürken yönetici odaklı bakış açısı bazen öznel değerlendirmelere sebep olabileceği hususunda eleştirilmektedir (Fitzpatrick, vd., 2011; Yüksel ve Sağlam, 2014).

### **Katılımcı Odaklı Program Değerlendirme Yaklaşımı**

Katılımcı Odaklı Program Değerlendirme Yaklaşımının doğuşu ABD’de program değerlendirmenin başladığı ilk yıllarda değerlendirmenin temel özelliğinin hedeflere ya da standartlara ne kadar ulaşıldığı ile sınırlı kalmasına olan tepkiye dayanır. Başta Stake olmak üzere 1960’lı yılların sonlarına doğru bazı eğitimciler program değerlendirme süreçlerinin uzun mekanik çalışmalar sonrasında yazılan kapsamlı raporlardan başka bir şey olmadığını ve sonuçta program değerlendiricilerin aslında değerlendirdikleri programa tam olarak hâkim olmadıklarını ifade etmişlerdir. Bu bilim insanları program değerlendirmede birey yani katılımcı ögesine daha fazla önem verilmesi gerektiğini, böylece programın farklı bakış açıları ile daha etkin bir biçimde değerlendirilebileceğini savunmuşlardır (Fitzpatrick, vd., 2011; Yüksel ve Sağlam, 2014).

Katılımcı Odaklı Program Değerlendirme Yaklaşımı öznel, nesnel, nitel ve nicel farklı birçok veri toplama süreçlerine olanak tanınması ve insan faktörüne vurgu yapması sebebiyle tercih edilen bir yaklaşımdır. Fakat bu yaklaşım farklı türden birçok verinin analizinin ve geçerlik-güvenirlik çalışmalarının zorluğu, objektifliği sağlamanın her zaman mümkün olmaması ve yüksek maliyetler sebebiyle eleştirilmektedir (Fitzpatrick, vd., 2011; Yüksel ve Sağlam, 2014).

### **Dünya’da ve Türkiye’de Program Değerlendirme**

Program değerlendirme konusunda benimsenen temel yaklaşımlar incelendiğinde program değerlendirmenin yaygınlaşmaya başladığı yıllarda hedeflere ne kadar ulaşıldığına odaklanan teknik ve katı bir anlayışın benimsendiği görülmektedir. İlerleyen yıllarda ise programın paydaşlarını daha çok gözetilen bir yaklaşıma doğru bir değişim ve yönelimin ortaya çıktığını söylemek mümkündür (Ornstein ve Hunkins 2018; Fitzpatrick, vd., 2011; Stufflebeam, Madaus ve Kellaghan, 2000). Program değerlendirmede paydaş katılımının bu denli dikkate alınmasının ortaya çıkardığı en büyük sorun ise araştırma ve katılım süreçlerinin birbiriyle çatışması olarak gösterilmektedir. Çünkü bilim insanı olmak araştırmaya yönelik bağımsız, tarafsız ve önyargısız bir yaklaşım gerektirirken bilim insanı olmayan kişilerin araştırma sürecine dahil edilmesinin bu kişilerin çelişkili beklentileriyle uğraşmak anlamına geldiği ifade edilmektedir (Meyer, Funk ve Nguyen, 2016). Bu noktada ifade edilen bu sorunun programın değerlendirilmesinde bir standardizasyon sorununu ortaya çıkarıp çıkamayacağı sorusu akla gelmektedir. Bunu önlemek için farklı kişi ve kuruluşların çeşitli çalışmaları söz konusudur. Başta Katılımcı Odaklı Program Değerlendirme Yaklaşımının önde gelen ismi Stake olmak üzere ABD tabanlı Eğitimde Değerlendirme Standartları Ortak Komitesi (JCSEE)’nin standart temelli program değerlendirme ve program değerlendirme standartları konusunda çalışmaları söz konusudur (Stake, 2004; Stufflebeam ve Madaus, 1983). Bu bağlamda 1980’li yıllardan itibaren yaygınlaşan standart geliştirme çalışmalarının temel amacı program değerlendirme uygulamalarının bilimsel süreçlerin dışına çıkmadan yürütülmesini sağlamaktır (Fournier, 1994). Özellikle Western Michigan University bünyesindeki değerlendirme standartları konusunda önemli bir yeri olan JCSEE’nin yürütmüş olduğu çalışmalar 1990’lı yılların sonu ve 2000’li yılların başı itibarıyla başta İsviçre ve Almanya olmak üzere Avrupa’ya da etkilemiş, ilgili ülkeler bu standartları 1997 ve 1999 yıllarında kendi dillerine çevirtmişlerdir (Beywl, 2000; akt. Yüksel ve Sağlam, 2014). Bu durum zaman içerisinde programın değerlendirilmesinden program değerlendirme süreçlerinin belirli standartlara göre değerlendirilmesine doğru evrilen bir yaklaşımın varlığını ifade edilmektedir (Astbury, 2016). Meta-değerlendirme olarak adlandırılan ve temel amacı değerlendirmenin güçlü ve zayıf yönlerini kamuya açık bir şekilde raporlaştırmak olan bu yaklaşımın 1960’lı yılların sonunda ortaya çıktığı bilinmektedir (Scriven, 2009; Stufflebeam, 2000). Ayrıca geliştirdiği standartlar farklı birçok program değerlendirme uzmanı tarafından hazırlanan çeşitli kontrol listelerine dönüştürülmüş JCSEE’nin program değerlendirme standartlarında *Hesap Verebilirlik Standartları* adlı temel odağı meta-değerlendirme olan bir boyuta da yer verilmiştir (JCSEE, 2018). Bu durum günümüzde özellikle gelişmiş ülkelerde program değerlendirmede yeri, zamanı, paydaşları, kullandığı model ve yaklaşımlar belli olan ve çeşitli standartları karşılayan program değerlendirme çalışmalarına verilen önemi göstermektedir.

Türkiye’de yukarıda ifade edilen formal nitelikteki program değerlendirme çalışmalarının yeri ve önemi farklı çalışmalarda ifade edilmiştir. Bu çalışmalarda yapılması planlanan bir program değerlendirme çalışmasında başarının ancak uygun kişilerce doğru yaklaşım ve modeller ışığında gerekli

noktalar dikkate alındığında sağlanabileceği vurgulanmaktadır (Yapıcıoğlu, Kara ve Yalçınoğlu, 2016; Yüksel ve Sağlam, 2014; Ertürk, 2013; Uşun, 2012). Ayrıca Türkiye’de program geliştirme ve değerlendirme çalışmalarını çeşitli boyutlarıyla inceleyen ve eğilimlerini çeşitli yönleriyle belirlemeyi amaçlayan çalışmalar söz konusudur (Yetkiner, Erdol ve Ünlü, 2019; Özüdoğru, 2018; Koç, 2015; Kurt ve Erdoğan, 2015). Bu çalışmalarda program değerlendirme amacıyla yapılan çalışmaların sayılarının yıllar ilerledikçe artış eğiliminde olduğu ve bununla doğru orantılı olarak program değerlendirme yaklaşım ve modellerinin geçmiş yıllara oranla daha fazla tercih edildiği bulgularına ulaşılmıştır. Fakat ilgili çalışmalar incelendiğinde Türkiye’deki program değerlendirme araştırmalarının 2015 yılına kadarki durumunun ortaya koyulduğu görülmektedir. Daha sonraki yıllarda yapılan program değerlendirme araştırmalarının incelenmesine yönelik çalışmalar ise ilköğretim, yabancı dil vb. belirli bir öğretim kademesi, program türü gibi belirli bir alandaki değerlendirmelere ya da doktora tezleri gibi yalnızca bir çalışma türüne odaklanmıştır. İlgili çalışmalarda ifade edilen yaklaşım ve model tercihindeki artış eğilimi Türkiye’de program değerlendirme anlayışının daha formal ve bilimsel bir tabana oturarak olumlu anlamda değişmeye başladığına işaret eder. Bu bağlamda eğitimde program değerlendirme konusunda 2015 ve 2019 yılları arasındaki beş yıllık süreçteki güncel çalışmaların da dâhil edildiği ve 2015 öncesindeki program değerlendirme çalışmalarının durumu ile daha sonraki beş yılı mukayese etmeye de olanak tanıyacak, son 10 yılı kapsayan yeni bir eğilim araştırmasının yapılması önemli görülmüştür.

#### Araştırmanın Amacı

Bu çalışmada Türkiye’de 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018 ve 2019 yıllarında program değerlendirme konusunda yapılan çalışmaların eğilimlerinin belirlenmesi amaçlanmıştır. Türkiye’de eğitim bilimleri alanında yapılan *eğilim araştırması* başlığı altındaki çalışmaların büyük oranda belirli bir alanda, belirli tarihler arasında yapılmış araştırmaları tarihsel dağılımlarına ve konu ve yöntem tercihlerine göre temalaştıran ve inceleyen çalışmalar oldukları görülmektedir (Ulutaş ve Ubuz, 2008; Varışoğlu, Şahin ve Göktaş, 2013; Ozan ve Köse, 2014; Kurt ve Erdoğan, 2015). Bir başka deyişle ilgili çalışmalarda *eğilim* kavramından kasıt herhangi bir alanda belirlenen bir konuya ilişkin araştırma probleminin çözümünde izlenecek yollardan herhangi birine yönelik tercihi ifade etmektedir. Bu durum program değerlendirme çalışmalarının daha önce ifade edilen yapısal ve yönetsel nitelikleri dikkate alındığında ilgili araştırmaların değerlendirdikleri program türü ve kademelerinin, değerlendirme sürecinde ölçüt aldıkları değerlendirme modelinin, ilgili modeli tercih etme sebeplerinin, tercih edilen modellerin program tür ve kademelerine göre dağılımlarının, örneklem, veri toplama ve analizi gibi yönetsel tercihleri, programın niteliğine ilişkin sonuç ve önerileri vb. özelliklerinin incelenebilirliğini göstermektedir. Bu bağlamda çalışmanın genel amacına uygun olarak araştırmada 2010 yılından başlayarak 2019 yılı da dahil olmak üzere Türkiye’de eğitimde program değerlendirme konusunda yapılan çalışmalarla ilgili aşağıdaki sorulara cevap aranmıştır:

1. Çalışmaların yıllara ve indekslendikleri veri tabanlarına göre dağılımı nasıldır?
2. Çalışmalarda kullanılan program değerlendirme modelleri ve bu modellerinin tercihinde öne sürülen gerekçeler nelerdir?
3. Çalışmalarda tercih edilen program değerlendirme modellerinin program türleri ve kademelerine göre dağılımı nasıldır?
4. Çalışmalarda tercih edilen araştırma yöntem ve desenleri nelerdir?
5. Çalışmalarda tercih edilen örnekleme yöntemleri nelerdir?
6. Çalışmalarda tercih edilen veri toplama araçları nelerdir?
7. Çalışmalarda değerlendirilen programların niteliği bakımından ne tür sonuçlara ulaşılmıştır?
8. Çalışmalarda değerlendirilen programlar hakkında ne tür kararlar önerilmiştir?

Türkiye’de program değerlendirme üzerine yapılan çalışmaların incelenmesi, çalışma bulgularının ve sonuçlarının karşılaştırılması, birleştirilmesi, yorumlanması ve ilgili çalışmalarda programın geleceğine ilişkin yapılan önerilerin ortaya koyulması eğitim sistemindeki programın geliştirilmesi ve değerlendirilmesi süreçlerindeki çeşitli sorunların fark edilmesi ve program değerlendirme çalışması

yapacak araştırmacılara bağlamsal bilgi vermesi adına faydalı olabilir. Yapılan çalışmanın bu yönüyle önemli olduğu düşünülmektedir.

## Yöntem

### Araştırma Deseni

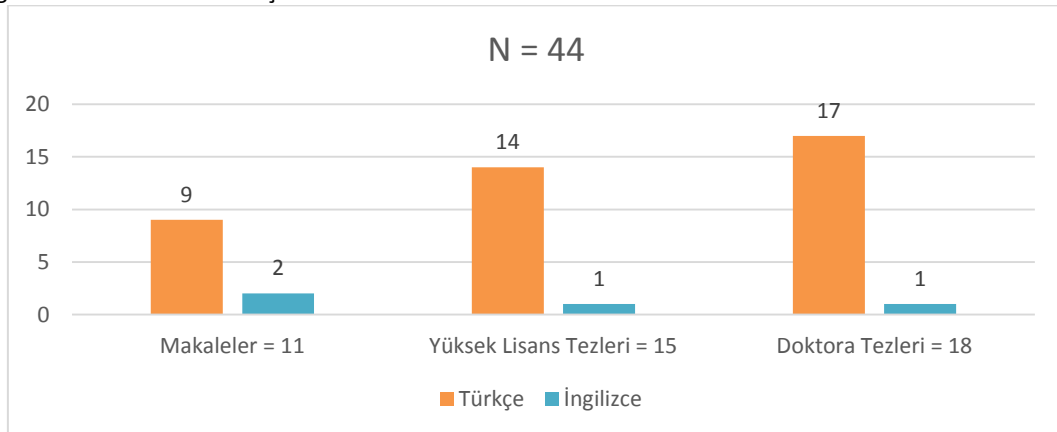
Bu araştırma yöntemsel tasarım, verilerin toplanması ve analizinde nitel araştırma yöntem ve tekniklerini kullanan bir çalışmadır. Bu bağlamda araştırma deseni iç içe geçmiş durum çalışmasıdır. Durum çalışmasında bir ya da daha fazla durum, olay, ortam, program ya da sosyal grubun incelenmesi söz konusudur (Creswell, 2007). İç içe geçmiş durum çalışmasında bir durum kendi içinde alt birimlere ayrılarak incelenir (Yıldırım ve Şimşek, 2013). Ayrıca çalışma Türkiye’de gerçekleştirilen program değerlendirme çalışmalarının mevcut durumunu ifade eden yönüyle betimsel niteliktedir. “Betimsel araştırmalar, olayı olduğu gibi araştırmaya ve var olan durumu belirlemeye çalışan araştırmalardır” (Karakaya, 2012, s. 59).

### Çalışma Kapsamında İncelenen Araştırmalar

Araştırmada Dergipark, Ulusal Tez Merkezi, ULAKBİM, Tr Dizin, Web of Science (WOS), EBSCO, ERIC, Elsevier ve Google Akademik gibi çeşitli veri tabanlarından 2010 ila 2019 yılları arasında program değerlendirme başlığı Türkçe ve İngilizce dillerinde yapılan 176 çalışmaya erişilmiş bunlardan herhangi bir program değerlendirme modelini kullanarak gerçekleştirilen 49’u dikkate alınmıştır. Yayınlanan beş makalenin bu 41 çalışma arasında da yer alan yüksek lisans ve doktora tezlerine ait makaleler olduğu görülmüştür. Sonuç olarak 44 çalışma incelenmek üzere araştırmaya dâhil edilmiştir. İlgili çalışmaların araştırmaya dâhil edilmesinde kullanılan önceden belirlenmiş ölçütler şöyledir:

- Araştırmaya dâhil edilen çalışmalar makale ve lisansüstü tezlerden oluşmaktadır.
- Araştırmaya dâhil edilen çalışmalar 2010 yılından başlayarak 2019 yılını da kapsayan 10 yıllık süreçte gerçekleştirilmiştir.
- Herhangi bir program değerlendirme modeli dikkate alınarak yapılan çalışmalar araştırmaya dâhil edilmiştir.
- Hem tez hem de makale olarak yayınlanan program değerlendirme çalışmalarında yalnızca tezler araştırmaya dâhil edilmiştir.

Çalışmaya dâhil edilen araştırmaların sahip olduğu yukarıdaki özellikler ölçüt olarak kabul edildiğinden bu araştırmada kullanılan örnekleme yöntemi *ölçüt örnekleme* olarak ifade edilebilir. Ölçüt örneklemin özelliği önceden belirlenmiş bir dizi ölçütü karşılayan tüm durumlarda geçerli olmasıdır (Yıldırım ve Şimşek, 2013). Araştırmaya dahil edilen çalışmaların sayısı, yüzde ve yayın dillerine ilişkin bilgiler Grafik 1’de sunulmuştur.



**Grafik 1.** Araştırmaya Dâhil Edilen Çalışmalar ve Yayın dilleri

Grafik 1’e göre incelenen 44 çalışmanın 18’i doktora tezi, 15’i yüksek lisans tezi ve 11’i makalelerden oluşmaktadır. Bu çalışmaların 40’i Türkçe, dördü ise İngilizce olarak yayınlanmıştır.



**Verilerin Toplanması ve Analizi**

Araştırma verileri Sözbil ve Kutu (2008) tarafından geliştirilen ve Ozan ve Köse (2014) tarafından güncellenen “Makale Sınıflama Formu” ile toplanmıştır. İlgili formun güncel halinde araştırma konusu, yöntemi, veri toplama araçları ve örneklem başlıklarında ekleme ve güncellemeler yapılmıştır. Araştırmacılar formun kapsam geçerliği için Eğitim Programları ve Öğretim (EPÖ) alanındaki üç uzmanın görüşüne başvurmuş ve uzmanlar arasındaki tutarlılık adına Cohen Kappa uyum katsayısını formun geneli için 0.83 bulmuşlardır. Bilindiği üzere Kappa uyum katsayısı -1 ile +1 arasında değişir ve +1’e yaklaştıkça uzmanlar arasındaki tutarlılık adına tesadüfî uyumun azaldığını gösterir (Fleiss ve Cohen 1973). Her ne kadar ilgili form “Makale Sınıflama Formu” başlığı altında hazırlansa da makale, tez ya da tam metin bildiriler gibi araştırma raporlarının incelenmesine olanak tanıyacak çalışmanın künyesi, araştırma yöntemi deseni/modeli, veri toplama araçları, örneklem vb. başlıkları dikkate almaktadır. Bu bağlamda ilgili form kullanılarak Türkiye’de 2010 ile 2019 yılları arasında program değerlendirme konusunda yapılan makaleler, yüksek lisans tezleri ve doktora tezlerinden oluşan dokümanlar çeşitli boyutlarda incelenmiştir. Creswell (2012) dokümanların özellikle nitel araştırmalar için önemli veri kaynaklarından olduğunu ifade etmiştir. İlgili çalışmaların toplanması ve çalışmada kullanılacakların seçilmesi ve verilerin analizi yaklaşık 3 aylık bir süreçte gerçekleşmiştir. Ayrıca formda yer alan boyutlara ek olarak yapılan çalışmaların kapsadığı program türleri ve kademeleri, kullanılan program değerlendirme modelleri, değerlendirilen programların niteliği bakımından ulaşılan sonuçlar ve değerlendirilen programlar hakkında önerilen kararlar da incelenmiştir. İncelenen çalışmaların künye bilgileri kaynakçaya ek olarak “İncelenen Çalışmalar” başlığı altında çalışmanın sonunda verilmiştir. Bu bağlamda verilerin analizinde içerik analizi tekniği kullanılmıştır. İçerik analizi verilerin belirli temalar içerisinde çeşitli özellikleri yönünden düzenlenmesi ve kavramsallaştırılmasını ifade etmektedir (Yıldırım ve Şimşek, 2013). Bu bağlamda araştırma verileri NVivo 10 uygulaması kullanılarak kodlanmış ve temalar halinde tasnif edilmiştir. İçerik analizi sonucu ortaya çıkan kod ve temalara ait bazı betimsel istatistikler Tablo 1’de gösterilmiştir.

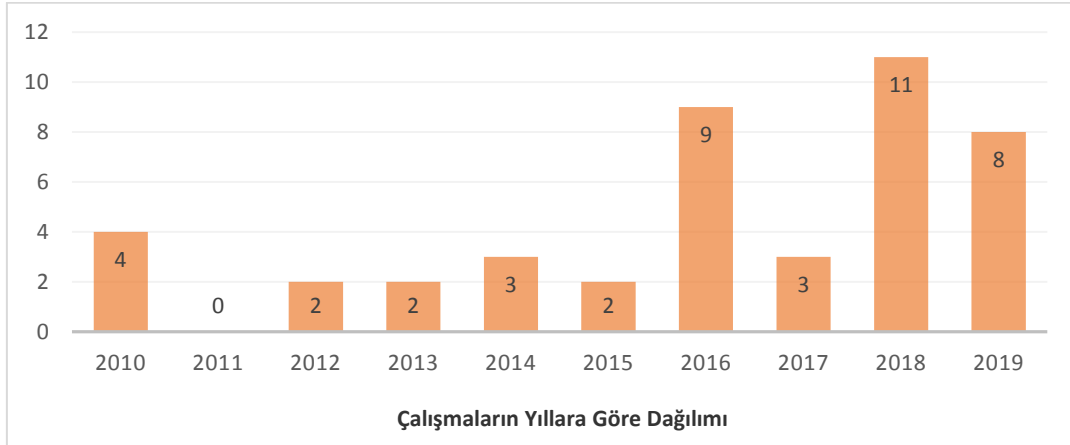
**Tablo 1.**  
İçerik Analizi Sonucu Ortaya Çıkan Kod ve Temalara İlişkin Betimsel İstatistikler

<b>Çıkarılan Temalar</b>	<b>f</b>	<b>%</b>
İncelenen Çalışmaların Yıllara Göre Dağılımı	9	7,8
Kullanılan Program Değerlendirme Modelleri	13	11,2
Kullanılan Modellerin Tercih Sebebi	18	15,5
Değerlendirilen Program Türü ve Kademesi	9	7,8
Kullanılan Araştırma Yöntem ve Desenleri	10	8,6
Kullanılan Örneklem Yöntemleri	17	14,7
Kullanılan Veri Toplama Araçları	8	6,9
Programların Niteliği Bakımından Ulaşılan Sonuçlar	27	23,3
Değerlendirilen Programlar Hakkında Önerilen Kararlar	5	4,3
<b>Toplam</b>	<b>116</b>	<b>100,0%</b>

Tablo 1’de her bir ana temaya ilişkin toplam kod sayısının frekans ve yüzdesi verilmiştir. Bazı temaların altında yer alan alt temalara ve kodlara ilişkin açıklamalar bulgular bölümünde detaylandırılmıştır.

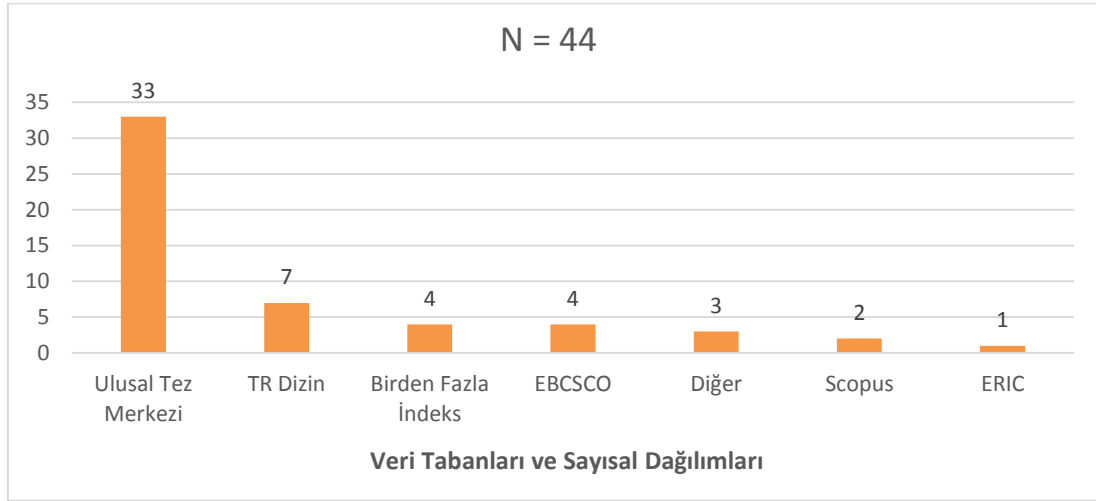
**Bulgular**

Türkiye’de 2010 ile 2019 yılları arasında herhangi bir program değerlendirme modeli kullanılarak yapılan program değerlendirme araştırmalarının yıllara göre dağılımına ilişkin bilgiler Grafik 2’de yer almaktadır.



**Grafik 2.** İncelenen Çalışmaların Yıllara Göre Dağılımı

Grafik 2 incelendiğinde 2010'dan 2015'e kadar yapılan çalışmaların sayısında küçük artış ve azalışların olduğu görülmektedir. 2011 yılında ise herhangi bir program değerlendirme modeli kullanılarak yapılan program değerlendirme çalışması söz konusu değildir. Türkiye'de program değerlendirme araştırmalarında 2016 yılı itibarı ile belli bir program değerlendirme modelinin ölçüt olarak alındığı çalışmaların daha fazla tercih edilir hale geldiği görülmektedir. İncelenen Çalışmaların indekslendikleri veri tabanlarına göre dağılımları ise Grafik 3'te verilmiştir.

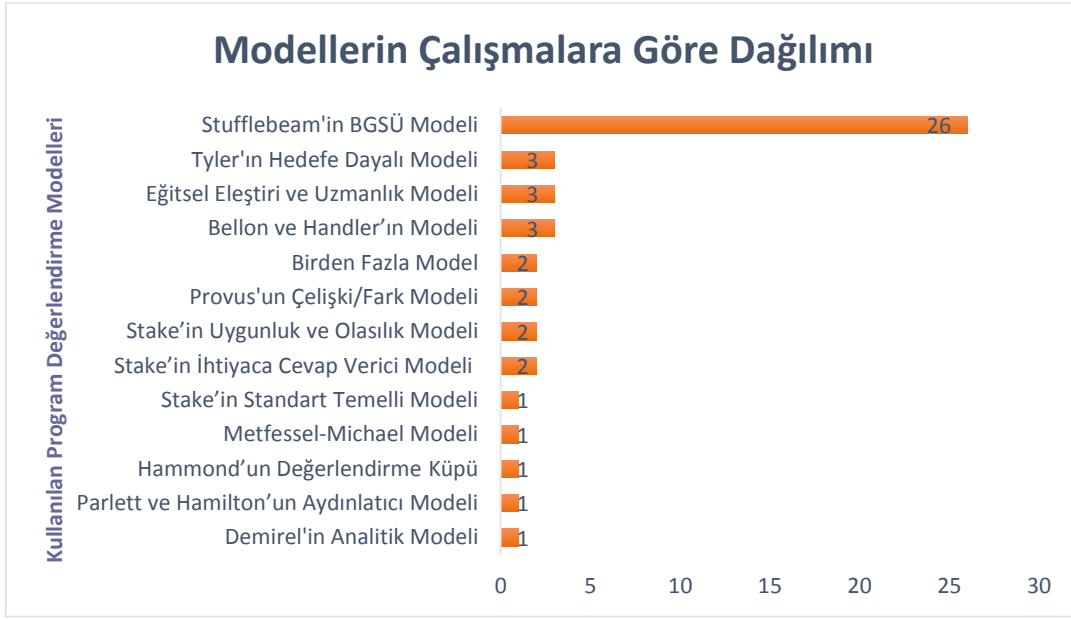


**Grafik 3.** Çalışmaların Yayınlandığı Veri Tabanları ve Sayısal Dağılımları

Grafik 3 incelendiğinde program değerlendirme başlığı altında yürütülen çalışmaların tezlerden oluşan 33'ünün Ulusal Tez Merkezinde indekslendiği görülmektedir. Bununla birlikte incelenen makalelerin yedisi TR dizinde, dördü EBSCO'da, ikisi Scopus'ta ve biri de ERIC'de indekslenmiştir. Dört çalışma birden fazla veri tabanında üç çalışma ise Yükseköğretim Kurulu Eğitim Bilimleri alan indeksi olarak kabul edilmeyen veri tabanlarında indekslenmiştir. Ayrıca birden fazla veri tabanında yer alan dört çalışma EBSCO, Scopus ve ERIC'te aynı anda indekslenmiştir. Başka bir deyişle uluslararası veri tabanlarında indekslenen çalışma sayısı dörttür.

#### **Çalışmalarda Kullanılan Program Değerlendirme Modelleri ve İlgili Modellerin Tercihinde Öne Sürülen Gerekçeler**

İncelenen çalışmalarda kullanılan program değerlendirme modellerinin dağılımları Grafik 4'teki gibidir.



**Grafik 4.** Çalışmalarda Kullanılan Program Değerlendirme Modellerinin Dağılımı

Grafik 4 incelendiğinde ilgili çalışmalarda 12 farklı değerlendirme modelinden yararlanıldığı görülmektedir. Yapılan 44 program değerlendirme araştırmasının 31'inde kullanılan program değerlendirme modelinin tercih sebebine ilişkin herhangi bir açıklama yapılmamıştır. Ayrıca çalışmalarda tercih edilen program değerlendirme modellerinde orantısız bir dağılımın olduğu görülmektedir. Yapılan çalışmaların büyük bir çoğunluğunda Stufflebeam'in BGSÜ Program Değerlendirme Modelinin tercih edildiği görülmektedir. Bu modelin dışında en çok kullanılan modeller ise Tyler'in Hedefe Dayalı Program Değerlendirme Modeli, Eğitsel Eleştiri ve Uzmanlık Modeli ve Bellon ve Handler Değerlendirme Modelidir. Ayrıca iki çalışmada birden fazla değerlendirme modelinden yararlanılmıştır. İncelenen çalışmalarda kullanılan program değerlendirme modellerinin tercihine ilişkin yaygın gerekçeler Tablo 2'de gösterilmiştir.

**Tablo 2.**

Program Değerlendirme Modellerinin Tercih Edilme Sebeplerine İlişkin Kodlar

Tercih Edilme Sebepleri	Değerlendirme Modelleri									
	Stufflebeam'in BGSÜ Modeli	Demirel'in Analitik Modeli	Parlett ve Hamilton'un Aydınlatıcı Modeli	Hammond'un Değerlendirme Küpü	Metfessel-Michael Modeli	Bellon ve Handler'ın Modeli	Provus'un Çelişki/Fark Modeli	Eğitsel Eleştiri ve Uzmanlık Modeli	Tyler'in Hedefe Dayalı Modeli	Stake'in Yaklaşım ve Modelleri
Betimleme Başarısı			+			+				
Bütüncül Yaklaşım	+									
Çok Boyutluluk	+									
Gelişim Odaklılık	+									
Sistematiik Veri Sağlaması	+									
Uygulama Kolaylığı	+									
Detaylı Değerlendirme Olanğı		+								
Paydaş Katılımı		+								
Programa Yansıyan Sonuçlar		+								
Hedeflerin Kontrolü				+						

Program Türüne Uyum					+					
Bağlam Değerlendirme Başarısı			+							
Veri Çeşitliliği Sağlaması			+							
Karşılaştırma Başarısı							+			
Gerekçe Verilmemiş								+	+	+
<b>TOPLAM</b>	<b>5</b>	<b>3</b>	<b>3</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>

Tablo 2 incelendiğinde program değerlendirme modellerinin tercihine ilişkin 14 yaygın gerekçeden söz edildiği görülmektedir. Eğitsel Eleştiri ve Uzmanlık Modeli, Tyler'ın Hedefe Dayalı Modeli ve Stake'in Yaklaşım ve Modellerinin tercihine ilişkin herhangi bir gerekçe sunulmamıştır. Çalışmalarda kullanılan program değerlendirme modellerinin temel özelliklerine ilişkin bilgiler ve tercih edilme sebeplerine ilişkin detaylı açıklamalar şöyledir:

Program değerlendiriciler, program uzmanları, araştırmacılar, geliştiriciler, politikacılar, liderler, yöneticiler, komiteler ve çalışma grupları gibi geniş bir kullanıcı kitlesine hitap eden Stufflebeam'in BGSÜ Program Değerlendirme Modeli programları bağlam, girdi, süreç ve ürün olmak üzere dört boyutta değerlendirmeyi amaçlar (Stufflebeam ve Coryn, 2014). İncelenen araştırmalarda BGSÜ Program Değerlendirme Modelinin tercih edilmesinde gerekçe olarak gösterilen en temel özellikler şöyledir (Duman, 2019; İlhan, 2018; Orhan, 2016; Aslan, 2015; Yurdakul, Uslu, Çakar ve Yıldız, 2014; Dinçer, 2013; Tekmen, 2012):

- Hem sonuca hem de sürece odaklanması ve çoklu değerlendirme yapmaya olanak tanınması,
- Programın başarısını kanıtlamaya değil programı geliştirmeye odaklanması,
- Değerlendirme sürecinde ortaya çıkacak potansiyel soruların belirlenmesinde, değerlendiriciye kolaylıklar sağlaması,
- Kapsamlı ve sistematik değerlendirme verileri sunmaya olanak tanınması,
- Programın farklı boyutlarda değerlendirmeye olanak tanınmasıdır.

Demirel tarafından geliştirilen Analitik Program Değerlendirme Modelinin program analizi ve paydaş görüşleri olmak üzere iki veri kaynağı bulunmaktadır. Birinci boyutta programın tasarımı, bağlam ve ihtiyaç analizi aşamalarında analizi yapıldıktan sonra analize girdi, süreç ve ürün aşamalarında devam edilir. İkinci boyutta ise alan ve program uzmanlarından başlayarak sivil toplum örgütlerine kadar uzanan paydaşların görüşleri söz konusudur (Demirel, 2014). Kuzu (2015) çalışmasında Demirel'in Analitik Program Değerlendirme Modelini kullanma gerekçesini ilgili modelin programın bütün paydaşların görüşlerini almaya ve ders programını detaylı bir biçimde incelemeye olanak tanınması böylece değerlendirme sonuçlarının dersin yansımalarını somut bir şekilde ortaya çıkarması olarak ifade etmiştir.

Parlett ve Hamilton'un Aydınlatıcı Değerlendirme Modelinde amaç yeni bir programın nasıl işlediğini farklı okullardaki uygulama biçimlerinden yola çıkarak değerlendirmektir. Bu model yönetim, öğrenci ve öğretmenlerin programa katılım düzeylerini ve materyal, araç ve gereçler gibi kaynakların programa katkısını ortaya koymaya çalışır (Parlett ve Hamilton, 1972). İlgili modelin kullanıldığı araştırmada modelin tercih sebebi olarak bağlamın değerlendirilmesi boyutunda başarılı olması, programın güçlü ve zayıf yönlerinin ortaya koyulmasına ve veri toplama araçlarının çeşitliliğine olanak tanınması gerekçe olarak sunulmuştur (Özüdoğru, 2016).

Hammond'un Değerlendirme Küpü yerel düzeyde program değerlendirmeye imkân tanıma amacıyla geliştirilmiş davranışlar, öğretim ve kurum boyutlarından oluşan üç boyutlu bir küp şeklinde ifade edilmiş bir değerlendirme modelidir (Hammond, 1967). Bu model sadece bir araştırmada kullanılmıştır. Hıdıroğlu, Kandemir ve Tuncel (2016), modelin güçlü yanı ve tercih sebebi olarak değerlendirme sırasında ulaşılan hedeflerin yanında ulaşılmayanların da dikkate alınmasının önemine dikkat çekmişlerdir.

Metfessel ve Michael 1960'lı yılların sonlarında hem Tyler'ın hedefe dayalı modeline alternatif olarak hem de hedefe dayalı yaklaşımı daha ileri taşıyacak bir program değerlendirme modeli önermişlerdir (Metfessel ve Michael 1967). Metfessel ve Michael Değerlendirme Modelinin Tyler'ın modelinden farkı

alternatif ölçme araçlarına daha fazla önem vermesi ve performansın hedefler yerine belirlenen standartlar ile karşılaştırılmasıdır (Yüksel ve Sağlam, 2014). Metfessel ve Michael Değerlendirme Modelini kullandığı araştırmada, ilgili modelin yükseköğretim programlarının değerlendirilmesinde uygun bir model olduğu ve bu sebeple tercih edildiği ifade edilmiştir (Bayık, 2018).

Bellon ve Handler Değerlendirme Modelinin, üç ana unsuru ve dört odak alanı söz konusudur. Modelin üç ana unsuru durum tanımları, analiz faaliyetleri ve program geliştirme bileşenleridir. Dört odak alanı ise programın hedefler, organizasyonlar, işlemler ve sonuçlar bileşenlerine atıfta bulunur (Bellon ve Handler'den aktaran Mutlu, 2018). İncelenen araştırmalar arasında bu modeli kullanan çalışmalarda model hakkında detaylı bilgi verilmiştir (Mutlu, 2018; Karcı Aktaş, 2018). Fakat modelin tercih gerekçesi olarak yalnızca bir çalışmada bu modelin programın mevcut durumunu ortaya koymaya ve daha iyi hale gelmesinde etkin öneriler sunmaya olanak tanınması gösterilmiştir (Şişman, 2019).

Provus'un Çelişki/Fark Program Değerlendirme Modeli programın değerlendirilmesinde hedeflere odaklanan bir modeldir. Provus'un 1969 ila 1971 yılları arasında Pittsburgh'daki farklı okullarının programlarını değerlendirme amacıyla yaptığı çalışmalar söz konusudur (Fitzpatrick, vd., 2011). Bu çalışmalar sonucunda birçok okulun özellikle devlet okullarının sistemlerinin yeterli planlama yapılmadan kurulduğu varsayımından yola çıkmış ve program değerlendirmeyi hedefler ile belirlenen standartlar arasındaki farkın ortaya çıkarılması süreci olarak tanımlamıştır (Provus, 1969). İlgili modeli kullanan araştırmalardan birinde modelin neden tercih edildiğine ilişkin herhangi bir açıklama yer almamaktadır. Babacan (2016) ise araştırmasında bütüncül eğitimi ölçüt olarak, olması gereken ile mevcut programın farkını ortaya koymak amacıyla Provus'un Çelişki/Fark Program Değerlendirme Modelini tercih ettiğini belirtmiştir.

Eisner'in Eğitsel Eleştiri ve Uzmanlık Modelinin temel özelliği programların bir sonucu olarak eğitim hayatının zengin bir tanımını üretmektir (Ornstein ve Hunkins, 2018). Modelin temel kavramlarını "eğitsel uzmanlık" ve "eğitsel eleştiri" oluşturmaktadır. Eisner eğitsel uzmanlığı takdir etme sanatı olarak ifade eder ve uzmanın eleştiri rolünün olması gerektiğini vurgular. Eisner'in modelinin betimleyici, yorumlayıcı, değerlendirici ve temalaştırıcı olmak üzere dört temel boyutundan söz etmek mümkündür (Eisner, 2002). İncelenen araştırmalarda modelin genel özellikleri, değerlendirmenin hangi boyutlarında ne tür yöntemsel gereklilikleri ortaya koyduğu vb. konulara değinilmiştir. Fakat bu modelin kullanılmasının değerlendirilen programlar için neden gerekli olduğu ifade edilmemiş, daha çok ilgili modelin Türkiye'de daha önce kullanılıp kullanılmadığı üzerinde durulmuştur (Kumral, 2010; Çelik, 2018).

Tyler'in 1930 ila 1945 yılları arasında yaptığı çalışmalarla eğitimde ve eğitimin değerlendirilmesinde büyük etkisi söz konusudur. Tyler, hedefe dayalı program değerlendirme yaklaşımında eğitim-öğretim faaliyetlerinin sonunda hedeflere ne kadar ulaşıldığına odaklanmış ve bunu ölçmenin en iyi yolunun başarı testleri olduğunu ifade etmiştir (Stufflebeam, vd., 2000). İncelenen çalışmalarda Tyler'in modelinin genel özellikleri ve program değerlendirme süreçlerinde dikkate aldığı temel aşamalar hakkında genel bilgiler verilmiştir. Fakat ilgili programların değerlendirilmesinde neden bu modelin tercih edildiği açıklanmamıştır (Kotluk ve Yayla 2016; Aslan ve Çıkar, 2017; Aslan ve Erden, 2018).

Stake'in Yanıtlayıcı/İhtiyaca Cevap Verici Program Değerlendirme Yaklaşımı çeşitli çalışmalarda doğrudan bir model olarak ifade edilmektedir. Katılımcı odaklı program değerlendirme farklı birçok kişiden toplanan ve değerlendirmede kullanılan formal ve informal veriyi içerir. Bu yaklaşımda birçok paydaştan elde edilen bu veri karmaşasının önüne geçmek ve değerlendirmenin daha geçerli ve güvenilir bir yapıda olması için Stake saat şeklinde 12 aşamadan oluşan yönlendirici bir eylem tablosu geliştirmiştir (Stufflebeam, vd. 2000). Stake tarafından geliştirilen Uygunluk-Olasılık Program Değerlendirme Modeli ise program değerlendirme süreçlerini kavramsallaştırarak anlaşılır bir hale getirmeyi amaçlayan katılımcı odaklı bir modeldir (Stake, 1967). Modelin anahtar kavramları programın durumunu "betimleme" ve program hakkında "karar verme" dir. Modelde program değerlendirme girdi süreç ve ürün boyutlarında yapılır (Yüksel ve Sağlam, 2014). Stake, Standart Temelli Değerlendirme Modelinde değerlendirmede karar sürecinde programı en iyi hale getirecek standartlara başvurulması gerektiğini, bu standartların ise belirlenen kesin standartlar ve farklı programlara ait göreceli standartlar olması gerektiğini ifade etmektedir (Stake, 2004). Stake'in yaklaşım ve modellerini kullanan program

değerlendirme çalışmalarında ilgili yaklaşım ve modellerin özellikleri detaylı bir biçimde anlatılmıştır (Yeşilyurt, 2010; Bayat, 2012; Altındağ, 2017). Fakat bu modellerin neden tercih edildiği ya da değerlendirilen program türü için uygun olup olmadığı konularının üzerinde durulmamıştır.

### Program Değerlendirme Modellerinin İncelenen Çalışmalarda Değerlendirilen Program Türleri ve Kademelerine Göre Dağılımı

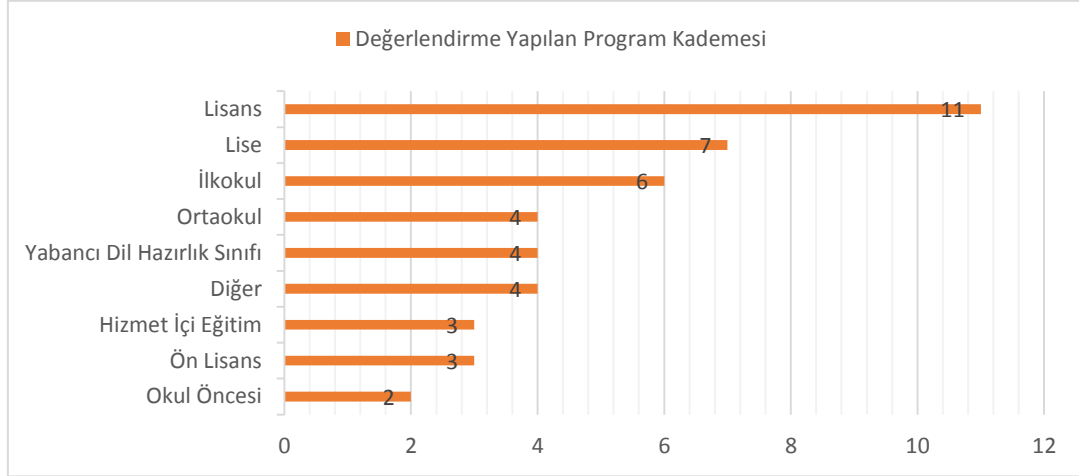
Yukarıda genel özellikleri verilen program değerlendirme modellerinin program türleri ve kademelerine göre dağılımları Tablo 3 ve Grafik 5'te gösterilmiştir.

**Tablo 3.**

*Program Değerlendirme Modellerinin Program Türleri ve Kademelerine Göre Dağılımı*

Okul Öncesi	Okul Öncesi Eğitim Programı <i>Provus'un Çelişki/Fark Değerlendirme Modeli</i>	Okul Öncesi Eğitim Programı <i>Stufflebeam's CIPP Evaluation Model</i>		
İlkokul	4. Sınıf Matematik Öğretim Programı <i>Tyler'in Hedefe Dayalı Değerlendirme Modeli</i>	4. Sınıf İngilizce Dersi Öğretim Programı <i>Stufflebeam's CIPP Evaluation Model</i>	2. Sınıf İngilizce Öğretim Programı <i>Parlett ve Hamilton'un değerlendirme modeli</i>	İlkokuma Yazma Programı <i>Stake'in Uygunluk-Olasılık Modeli</i>
	4. Sınıf Türkçe Öğretim Programı <i>Stufflebeam's CIPP Evaluation Model</i>	3. Sınıf Fen Bilimleri Öğretim Programı <i>Stufflebeam's CIPP Evaluation Model</i>		
Ortaokul	5. Sınıf Fen Bilimleri Öğretim Programı <i>Tyler'in Hedefe Dayalı Değerlendirme Modeli</i>	7. Sınıf Türkçe ve Matematik Programları <i>Stufflebeam's CIPP Evaluation Model</i>	5. Sınıf Matematik Dersi Öğretim Programı <i>Stake'in Uygunluk-Olasılık Modeli</i>	7. Sınıf İngilizce Öğretim Programı <i>Stufflebeam's CIPP Evaluation Model</i>
Lise	9. Sınıf Fizik Öğretim Programı <i>Tyler'in Hedefe Dayalı Değerlendirme Modeli</i>	Ortaöğretim İngilizce Dersi Öğretim Programı <i>Eisner'in Eğitsel Eleştiri Modeli</i>	Sosyal Bilimler Liseleri Eğitim Programı <i>Stufflebeam's CIPP Evaluation Model</i>	9. Sınıf Matematik Dersi Öğretim Programı <i>Stufflebeam's CIPP Evaluation Model</i>
	9. Sınıf İngilizce Dersi Öğretim Programı <i>Provus'un Çelişki/Fark Değerlendirme Modeli</i>	Abitur Programı <i>Eisner'in Eğitsel Eleştiri Modeli</i>	9. sınıf Türk edebiyatı dersi öğretim programı <i>Stufflebeam's CIPP Evaluation Model</i>	
Ön Lisans	Saç Bakımı ve Güzellik Hizmetleri Programı <i>Metfessel-Michael Değerlendirme Modeli</i>	Hava Astsubay MYO İngilizce Öğretim Programı <i>Stufflebeam's CIPP Evaluation Model</i>	Ön lisans Turizm Eğitim Programı <i>Stufflebeam's CIPP Evaluation Model</i>	
Yab. Dil Hazırlık	İngilizce Hazırlık Programındaki Temel Ders <i>Bellon ve Handler's Evaluation Model</i>	Hazırlık Okulu Programı <i>Stufflebeam's CIPP Evaluation Model</i>	Yabancı Diller YO İngilizce Hazırlık Programı <i>Bellon ve Handler's Evaluation Model</i>	İngilizce Hazırlık Öğretim Programı <i>Bellon ve Handler's Evaluation Model</i>
Lisans	Öğretmenlik Uygulaması Dersi <i>Stufflebeam's CIPP Evaluation Model</i>	Sınıf Öğretmenliği Lisans Programı <i>Stufflebeam's CIPP Evaluation Model</i>	Lisans Programları Çekirdek Program <i>Stufflebeam's CIPP Evaluation Model</i>	Öğretim İlke ve Yöntemleri Ders Programı <i>Demirel'in Analitik Değerlendirme Modeli</i>
	Öğretmenlik Uygulaması Dersi <i>Stufflebeam's CIPP Evaluation Model</i>	Okul Öncesi Öğretmenliği Programı <i>Stufflebeam's CIPP Evaluation Model</i>	Öğretmenlik Uygulaması Öğretim Programı <i>Stake'in Standart Temelli ve İhtiyaca Cevap Verici Modeli</i>	Sınıf Öğretmenliği Öğretim Programı <i>Eisner'in Eğitsel Eleştiri Modeli</i>
	Öğretmenlik Meslek Bilgisi Dersleri <i>Stufflebeam's CIPP Evaluation Model</i>	Kişilerarası İlişkiler ve Terapötik İletişim Dersi Programı <i>Stufflebeam's CIPP Evaluation Model</i>	Öğretim İlke ve Yöntemleri Ders Programı <i>Hammond'un Değerlendirme Küpü</i>	

Eğitim	Web Tabanlı İçerik Geliştirme Programı <i>Stufflebeam's CIPP Evaluation Model</i>	Çağrı Merkezi Mesleki Yetkinlik Geliştirme Programı <i>Stufflebeam's CIPP Evaluation Model</i>	Intel Öğretmen Programı (İÖP) <i>Stufflebeam's CIPP Evaluation Model</i>	
	Diğer	Doğum Eğitimi Programı <i>Stake'in İhtiyaca Cevap Verici ve Stufflebeam'in BGSÜ Modeli</i>	İlköğretimde Teknoloji Uygulamaları e-Sertifika Programı <i>Stufflebeam's CIPP Evaluation Model</i>	Uzaktan Eğitim Yab. Dil Dersi Öğretim Programı <i>Stufflebeam's CIPP Evaluation Model</i>
				Türkçe Öğretim Programı <i>Stufflebeam's CIPP Evaluation Model</i>

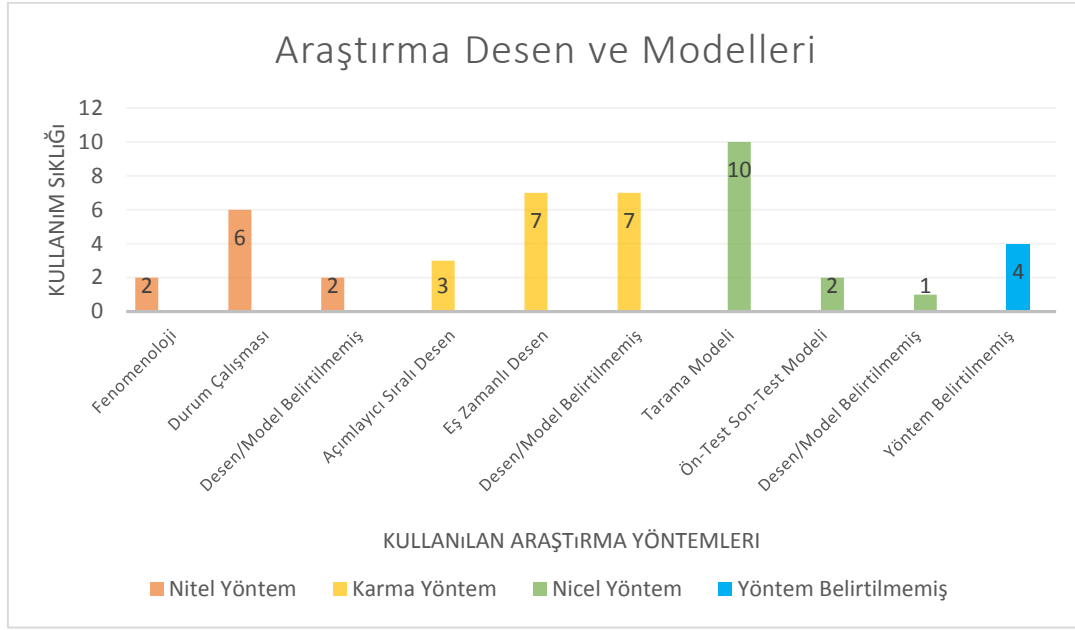


**Grafik 4.** İncelenen Çalışmalarda Değerlendirilen Programların Kademeleri

Tablo 3 ve Grafik 5 incelendiğinde program değerlendirme süreçlerinde en çok tercih edilen program türlerinin lisans programları ve liselere ait programların olduğu görülmektedir. Değerlendirilen lisans programlarının ise büyük oranda eğitim fakültesi bağlamında olduğu söylenebilir. Bunun yanında lisansüstü programlarını değerlendiren herhangi bir çalışmaya rastlanmazken okul öncesi programları iki defa ile en az değerlendirilen programdır. Kullanılan değerlendirme modellerinin program kademelerine göre dağılımı incelendiğinde ilkököl ve lise programlarının değerlendirilmesine model çeşitliliğinden söz etmek mümkündür. Fakat özellikle hizmet içi eğitim programlarının değerlendirilmesinde kullanılan tek modelin Stufflebeam'in BGSÜ Program Değerlendirme Modeli olduğu görülmektedir. Bununla birlikte yabancı dil hazırlık programlarının değerlendirilmesinde Bellon ve Handler's Evaluation Model nin yaygın olarak kullanıldığı görülmektedir. Ayrıca "diğer" başlığı altında verilen çeşitli kişisel gelişim ve sertifika programlarının ve uzaktan eğitim gibi herhangi bir kademeye konumlandırılmayan programların değerlendirilmesinde de ağırlıklı olarak Stufflebeam'in BGSÜ Program Değerlendirme Modeli kullanılmıştır.

#### Çalışmalarda Kullanılan Araştırma Yöntem, Desen ve Modelleri

İncelenen çalışmalarda kullanılan araştırma yöntem, desen ve modellerine ilişkin bulgular Grafik 6'daki gibidir.

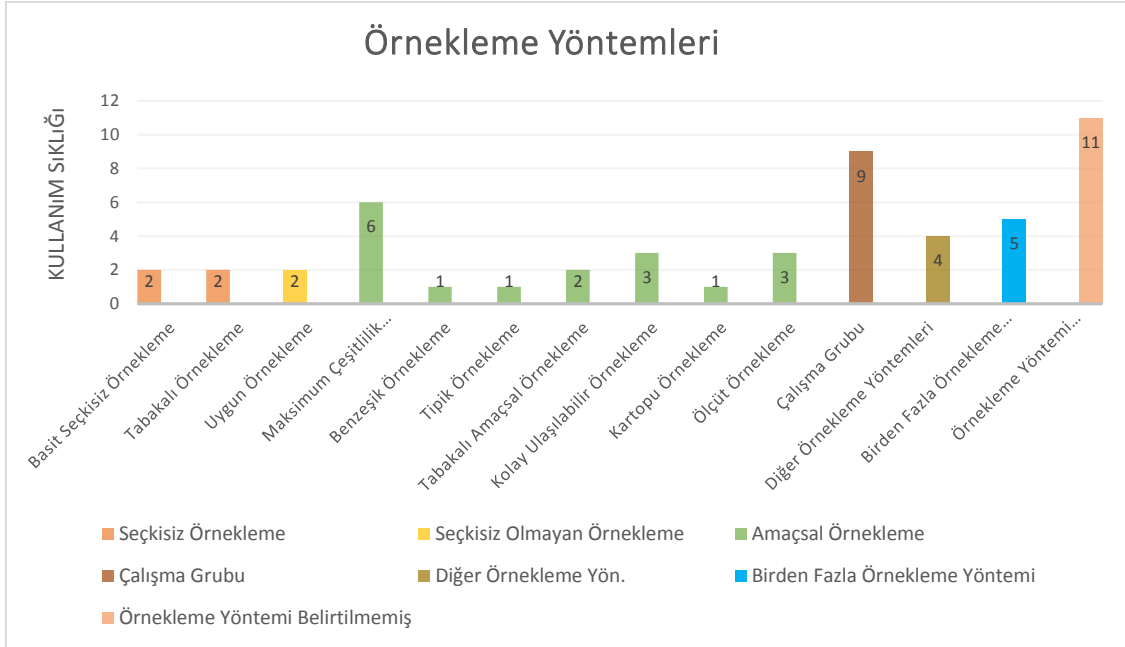


**Grafik 6.** İncelenen Çalışmalarda Kullanılan Araştırma Desen/Modelleri

Grafik 6 incelendiğinde incelenen çalışmalarda en çok karma yöntemlerin kullanıldığı, en yaygın desen ve modellerin ise tarama modeli ve eşzamanlı desen olduğu görülmüştür. Ayrıca bazı çalışmalarda herhangi bir yöntem, desen ya da model belirtilmezken bazılarında sadece nitel, nicel ve karma yöntem ifadesi ile yetinilmiştir.

#### Çalışmalarda Kullanılan Örnekleme Yöntemleri

İncelenen çalışmalarda kullanılan örnekleme yöntemlerine ilişkin bulgular Grafik 7’de gösterilmiştir.



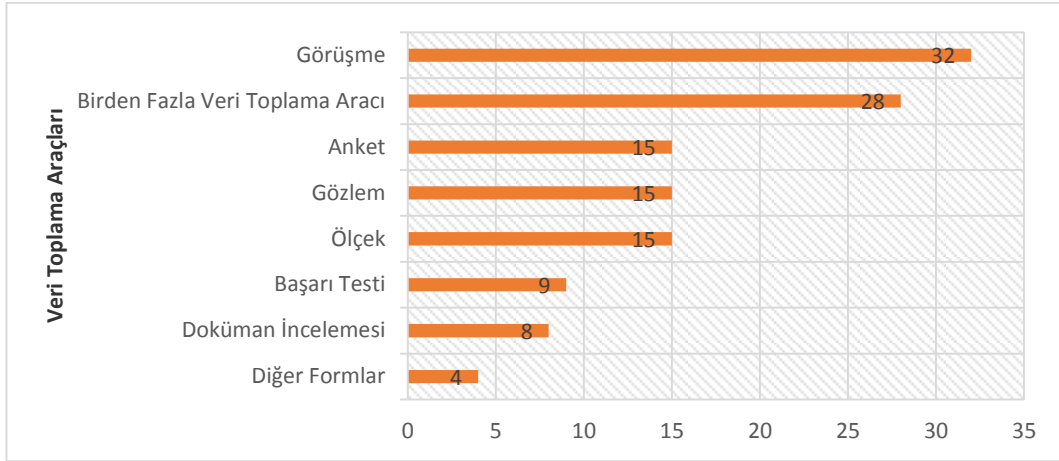
**Grafik 7.** İncelenen Çalışmalarda Kullanılan Örnekleme Yöntemleri



Grafik 7 incelendiğinde araştırmalarda yaygın olarak seçkisiz olmayan örnekleme yöntemlerinin özellikle de amaçsal örneklemenin tercih edildiği görülmektedir. Bunun yanında amaçsal örnekleme yöntemlerinden en yaygın kullanılanın maksimum çeşitlilik örnekleme olduğu söylenebilir. İncelenen çalışmaların yedisinde çalışmaya katılımcıların tamamı dâhil edilmeye çalışılmıştır. Beş çalışmada ise birden fazla örnekleme yöntemi kullanılmıştır. Bunun yanında incelenen 44 çalışmanın 11'inde herhangi bir örnekleme yönteminden söz edilmemiştir.

#### Çalışmalarda Kullanılan Veri Toplama Araçları

İncelenen program değerlendirme araştırmalarında kullanılan veri toplama araçları Grafik 8'de ifade edilmiştir.



**Grafik 8. İncelenen Çalışmalarda Kullanılan Veri Toplama Araçları**

Grafik 8 incelendiğinde çalışmaların 28'inde birden fazla veri toplama aracının kullanıldığı görülmektedir. En yaygın kullanılan veri toplama araçları ise görüşme, anket, gözlem ve ölçeklerdir.

#### Çalışmalarda Değerlendirilen Programların Niteliği Bakımından Ulaşılan Sonuçlar

İncelenen çalışmalara yönelik yapılan içerik analizi sonucunda değerlendirilen programlar hakkında ulaşılan yaygın sonuçlara ilişkin betimsel istatistikler Tablo 4'te verilmiştir.

**Tablo 4.**

Çalışmalarda Programların Niteliği Adına Ulaşılan Sonuçlara İlişkin Betimsel İstatistikler

Ulaşılan Yaygın Sonuçlar	Olumlu Sonuçlar		Olumsuz Sonuçlar	
	f	%	f	%
Amaca Hizmet Etme	15	25,0	18	12,6
Bireysel Özellikleri Dikkate Alma	2	3,3	4	2,8
Duyuşsal Özellikleri Kazandırma	6	10,0	2	1,4
Eğitici Sayı ve Niteliği	2	3,3	8	5,6
Fiziki İmkanlar	3	5,0	13	9,1
Genel ve Özel Hedeflerin Niteliği	4	6,7	6	4,2
Paydaşların Görev Tanımı			1	0,7
Öğrencilerin Hazırbulunuşluk Durumu			3	2,1
Program İçeriklerinin Niteliği	1	1,7	9	6,3
Eğitici, Yönetici ve Öğrenci Etkileşimi			5	3,5
Kurumsal Farklılıklar			4	2,8
Araç-Gereç ve Materyallerin Niteliği			8	5,6
Öğrencilerin Gelişim Durumu	7	11,7	7	4,9
Öğrenci Profili			3	2,1

Programın Öğrenciye Uygunluğu	4	6,7		
Ölçme ve Değerlendirme Niteliği	2	3,3	10	7,0
Öğretim Tasarımının Niteliği	4	6,7		
Programın Özerklik ve Esneklik İmkânı			2	1,4
Paydaş Katılımı	1	1,7	1	0,7
Program Felsefesinin Uygunluğu	2	3,3		
Program Hakkında Bilgi Altyapısı			6	4,2
Sınıf Mevcudu			5	3,5
Öğrenim Süresi	2	3,3	8	5,6
Uygulama Birliği			2	1,4
Uygulama Eğitimi			5	3,5
Programın Uygulanabilirliği	2	3,3	7	4,9
Yöntem ve Tekniklerin Niteliği	3	5,0	6	4,2
<b>Toplam</b>	<b>60</b>	<b>100,0%</b>	<b>143</b>	<b>100,0%</b>

Tablo 4 incelendiğinde değerlendirilen programların niteliğine ilişkin çeşitli sonuçlara ulaşıldığı görülmektedir. İçerik analizi sırasında yapılan kodlamalar değerlendirilen programların birçok boyutunda hem olumlu hem de olumsuz sonuçlara ulaşıldığını göstermektedir. Bu durum programın amacına hizmet etmesi, öğrencilerin gelişim durumu, kullanılan yöntem ve tekniklerin niteliği vb. unsurların bazı çalışmalarda olumlu bazı çalışmalarda ise olumsuz yönde değerlendirilmesinden kaynaklanmaktadır. Programların niteliğine ilişkin olumsuz değerlendirmelerin kodlama sayısı 143 olumlu değerlendirmelerin ise 60'dır. Programların niteliğine ilişkin en yaygın olumlu değerlendirmeler sırasıyla "Amaca Hizmet Etme, Öğrencilerin Gelişim Durumu ve Duyuşsal Özellikleri Kazandırma" hususlarıdır. En yaygın olumsuz değerlendirmeler ise sırasıyla "Amaca Hizmet Etme, Fiziki İmkanlar ve Ölçme ve Değerlendirme Niteliğidir". Ayrıca "eğiticilerin, program içeriklerinin, araç-gereç ve materyallerin niteliği ile öğrenim süresi ve eğitici sayısı" da yaygın olumsuz sonuçlardandır.

#### Çalışmalarda Değerlendirilen Programlar Hakkında Alınması Önerilen Kararlar

Program değerlendirme çalışmaları incelendiğinde değerlendirilen programlara yönelik çeşitli kararların önerildiği görülmektedir. Bu bağlamda değerlendirilen programlar hakkında alınması önerilen kararlara ilişkin betimsel istatistikler Tablo 5'te verilmiştir.

**Tablo 5.**

Değerlendirilen Programlar Hakkında Alınması Önerilen Kararlara İlişkin Betimsel İstatistikler

Önerilen Kararlar	f	%
Kapsamlı güncelleme yapılarak sürdürülebilir.	15	34,1%
Kısmi güncelleme yapılarak sürdürülebilir.	21	47,7%
Herhangi bir öneride bulunulmamış.	5	11,4%
Program başarısız, yeniden hazırlanmalı.	2	4,5%
Program başarılı, olduğu gibi sürdürülmeli.	1	2,3%
<b>Toplam</b>	<b>44</b>	<b>100,0%</b>

Tablo 5 incelendiğinde değerlendirilen programlar hakkında geliştirmeye ya da sonlandırmaya yönelik önerilerin yapıldığı görülmektedir. Değerlendirilen programlar hakkında en sık önerilen durumun programın kısmi güncellemeler sonrasında sürdürülmesidir. Bu öneriyi sırasıyla "Kapsamlı güncelleme yapılarak sürdürülebilir, Program başarısız, yeniden hazırlanmalı ve Program başarılı, olduğu gibi sürdürülmeli" şeklinde öneriler takip etmektedir. Ayrıca beş çalışmada değerlendirilen programa yönelik herhangi bir öneride bulunulmamıştır.

### Sonuç, Tartışma ve Öneriler

Araştırma bulguları incelendiğinde Türkiye’de 2010 ila 2019 yılları arasında program değerlendirme başlığı altında yapılan çalışmaların büyük oranda program değerlendirme yaklaşım ve modellerinden bağımsız olarak yürütüldüğü söylenebilir. Özellikle çeşitli veri tabanlarından ulaşılan 176 çalışmanın yalnızca 44’ünde program değerlendirme modellerinden yararlandığı görülmektedir. Bu da program değerlendirme başlığı altında yapılan çalışmaların %25’ine karşılık gelmektedir. Program değerlendirme yaklaşım ve modellerinden bağımsız yapılan araştırmalarda programın başarısına ya da başarısızlığına hangi ölçüte göre karar verildiği ise tartışmaya açık bir konudur. Stufflebeam (1999) program değerlendirmede *Sözde Değerlendirmeler (Pseudoevaluations)* olarak adlandırdığı bu türden değerlendirmelerin genelde programın asıl yararına ve değerine bakılmaksızın program hakkında olumlu ya da olumsuz görüş belirten değerlendirmeler olduğunu ifade etmiştir. Stufflebeam *Soru ve Yöntem Odaklı Değerlendirmeler* olarak adlandırdığı değerlendirmelerin ise bazı standart soru ve yöntemlerden yola çıkarak programı değerlendirmeye çalıştıklarını ve başarılı olup olmadıklarının tartışmaya açık olduğunu vurgulamaktadır. Bu durum özellikle Türkiye’de program değerlendirme başlığı altında sıkça tercih edilen belirli bir grubun görüşlerine göre değerlendirme yapmayı amaçlayan araştırmaların program değerlendirmede gerçekten başarılı olup olmadıkları sorusunu akla getirmektedir. Stufflebeam (1999) en etkin değerlendirmelerin *Gelişim ve İzlenilebilirlik Odaklı Yaklaşımlar* aracılığıyla yapılabileceğini ifade etmiştir. Bu yaklaşımların değerlendirmede temel odağı programın gerçekten başarılı olup olmadığıdır ve programın başarısını değerlendirmek için gereken tüm soru ve kriterleri göz önünde bulunduran kapsamlı ve pahalı değerlendirmelerdir. Bu türden değerlendirmeler ise ancak belirli yaklaşımlar çerçevesinde kullanılacak program değerlendirme modelleri ile mümkündür. Program değerlendirme başlığı altında yapılan çalışmalarda model kullanımı henüz küçük bir yüzdelik dilime karşılık gelse de 10 yıllık süreçte program değerlendirme araştırmalarında model kullanımının yaygınlaşması konusunda olumlu gelişmeler söz konusudur. İlgili araştırmaların incelendiği 10 yıllık dilim, ilk ve ikinci 5 yıl olarak iki parça halinde değerlendirildiğinde program değerlendirme çalışmalarının sayısı ilk 5 yılda 11 iken, bu sayının ikinci 5 yıllık süreçte 35’e çıktığı görülmektedir. Özellikle 2016 yılı itibari ile belli bir program değerlendirme modelinin ölçüt olarak kullanıldığı çalışmaların daha fazla tercih edilme hale geldiği görülmektedir. Program değerlendirme çalışmalarının sayılarındaki artışa rağmen incelenen araştırmaların yalnızca dördü uluslararası veri tabanlarında indekslenmektedir. Bu durum program değerlendirme çalışmalarını nitelik olarak sorgulanır hale getirmektedir.

Araştırma bulgularına göre örgün eğitim kurumları olan lisans, ilkokul, ortaokul ve liselerde uygulanan öğretim programların değerlendirilmesine ağırlık verilirken okul öncesi programlarına yönelik sınırlı sayıda program değerlendirme çalışması yapılmıştır. Ayrıca lisansüstü programlarına yönelik herhangi bir program değerlendirme çalışmasına rastlanmamıştır. Yapılan çalışmaların büyük bir çoğunluğunda (yaklaşık %59) Stufflebeam’in Bağlam, Girdi, Süreç ve Ürün Program Değerlendirme Modeli kullanılmıştır. Geçmiş yıllarda Türkiye’deki program değerlendirme çalışmalarını inceleyen araştırmalarda da bu durumdan söz edilmektedir. Kurt ve Erdoğan (2015) yaptıkları araştırmada Türkiye’de program değerlendirme konusunda yapılan çalışmalarda en çok tercih edilen modelin BGSÜ Modeli olduğunu ifade etmişlerdir. Özüdoğru (2018) yaptığı çalışmada, yabancı dil eğitimi programlarını değerlendiren çalışmalarda en çok tercih edilen modelin BGSÜ Modeli olduğu şeklinde benzer bir bulguya ulaşmıştır. Uluslararası literatürde de Stufflebeam’in BGSÜ Program Değerlendirme Modeli en çok tercih edilenlerin başında gösterilmektedir (Warju, 2016; Zhang. vd. 2011). Hâlihazırda BGSÜ Modelinin farklı türden araştırmacılara uygulama kolaylığı sağladığı, hatta yapılış dinamiğinin bununla ilgili olduğu bilinmektedir (Stufflebeam ve Coryn 2014). Ayrıca incelenen çalışmalarda kullanılan modeller arasında Demirel’in Analitik Değerlendirme Modeli dışında Türkiye’de tasarlanan herhangi bir model yer almamaktadır. İlgili model ise sadece bir çalışmada kullanılmıştır. Fakat incelenen araştırmalarda bu modelin ve diğer modellerin tercihinin gerçekten de değerlendirilen programın bağlamı düşünülerek alınan kararlara dayanıp dayanmadığı ilgili çalışmaların büyük bir bölümünde net bir biçimde ifade edilmemiştir. Örnek olarak incelenen çalışmalarda ilkokul ve lise programlarının değerlendirilmesine bir model çeşitliliğinin söz etmek mümkünken ilgili çalışmalarda bu modellerin neden tercih edildiği açıkça ifade edilmediği için bu model çeşitliliğinin sebebinin gerçekte ne olduğu net olarak

anlaşılamamaktadır. Başka bir örnekte hizmet içi eğitim programlarının değerlendirilmesinde kullanılan tek modelin Stufflebeam'ın BGSÜ Program Değerlendirme Modeli olduğu görülmektedir. Hâlbuki hizmet içi eğitim ve yetişkin eğitimi programlarının değerlendirilmesinde kullanılan Kirkpatrick Modeli gibi başarılı modellerin varlığı bilinmektedir (Smidt, Balandin, Sigafos ve Reed, 2009; Uşun, 2012). Bu noktada ilgili yaklaşım ve modelin tercihinde araştırmacının takdiri önemlidir. Fakat herhangi bir programın değerlendirilmesinde kullanılacak yaklaşım ve modelin tercihinde, bağlamına uygun bir şekilde karar verilmesi ve bu süreçlerin açık bir biçimde raporlaştırılması çalışmanın şeffaflığı açısından gereklidir.

Araştırma bulgularına göre yapılan program değerlendirme çalışmalarında çoğunlukla karma yöntem ve çoklu örnekleme ve veri toplama süreçleri tercih edilmiştir. Özellikle sosyal bilimlerde üçgenleme de denilen birden fazla veri toplama aracının kullanıldığı çalışmalar insan karakterinin zenginliğini ve karmaşıklığını daha iyi ifade ettiği için değerli görülmektedir (Cohen, Manion ve Morrison, 2007). Bu bağlamda en çok tercih edilen veri toplama araçlarının da görüşme ve gözlem gibi nitel veri toplama araçları olduğu görülmektedir. Bu durum yapılan program değerlendirme çalışmalarının büyük oranda farklı örneklerden toplanan veri zenginliğine sahip olduğunu ve insani özellikleri derinlemesine incelemeyi amaçlayan nitel verilere ağırlık verildiğini göstermektedir. Yapılan program değerlendirme çalışmalarının bir kısmında ise yukarıda belirtilen bulguların aksine araştırma yöntemi ve örnekleme yönteminden bahsedilmemiş ya da sadece bir veri toplama aracıyla program değerlendirme çabasına girilmiştir. Bu da yine yukarıda ifade edilen kapsamlı çalışmaları tercih etmeme durumunu akla getirmektedir. Hatta araştırma deseni ve örneklem seçimindeki eksikliklerin yöntemsel hatalar olduğu söylenebilir. Bu türden yöntemsel hataların yaygın olarak yapıldığı çeşitli çalışmalarda ifade edilmektedir (Toy ve Tosunoğlu, 2007; Erdoğan, 2001). Bilimsel araştırmalarda kapsamlı çalışmaları tercih etmeme durumu ve yapılan yöntemsel hatalar hiç şüphesiz çalışmaların niteliğini olumsuz etkileyen durumlardır.

Yapılan içerik analizlerinde değerlendirilen programlara ilişkin çeşitli sonuçlara ulaşıldığı ve bu sonuçlardan yola çıkarak programlar ile ilgili karar mercilerine çeşitli önerilerde bulunduğu görülmüştür. İçerik analizi sonucunda değerlendirilen programlar hakkında 60 olumlu 143 olumsuz değerlendirmenin yapıldığı görülmüştür. Bu da değerlendirilen programların büyük oranda sorunlar barındırdığını göstermektedir. Bu programlarla ilgili fiziki imkanların yetersizliği ve uygulanan ölçme ve değerlendirme faaliyetlerinin niteliğine ilişkin olumsuz sonuçlar söz konusuken olumlu değerlendirme sonuçları arasında öğrencilerin gelişim durumu ve programların çeşitli duyuşsal özellikleri kazandırması konuları yer almaktadır. Programların amacına hizmet etme durumu hakkında hem olumlu hem de olumsuz anlamda en çok sonuç çıkarılan konu olmuştur. Bu konu JCSEE tarafından geliştirilen standart listesinin en başındaki standart alanı olarak *Amaca Uygunluk Standartlarında* da yer verilen önemli bir konudur (JCSEE, 2018). Bu konuda programlara ilişkin 15 olumlu 18 olumsuz sonuç çıkarılmıştır. Bu da değerlendirilen programların büyük oranda amacına hizmet etme konusunda sorunlu olduğunu göstermektedir. Ayrıca incelenen çalışmalarda araştırma sonuçlarından yola çıkarak değerlendirilen programların gelecekteki durumuna karar verme maksadıyla çeşitli önerilerde bulunulmuştur. Sözü edilen bu öneriler büyük oranda programların kısmi güncellemelerle sürdürülebileceği yönündedir. Fakat bu noktada fiziki imkanların yetersizliği, ölçme ve değerlendirme faaliyetlerinin, eğitimcilerin, program içeriklerinin, araç-gereç ve materyallerin niteliği ile öğrenim süresi ve eğitici sayısındaki sınırlılıklar konusunda olumlu yönlerine kıyasla bu kadar çok olumsuz eleştiri alan programların kısmi güncellemelerle nasıl etkin hale getirilebileceği tartışmaya açık bir konudur.

Dünya'daki hızlı gelişim ve değişim her alanda olduğu gibi programların geliştirilmesi ve değerlendirilmesi süreçleri üzerinde de etkilerini göstermektedir. Brezilya, Çin, Hindistan ve Güney Afrika gibi ülkelerin son yıllarda yapmış olduğu atılımların ilerleyen yıllarda eğitim sistemlerine de yansıtacağı ve bu durumun program değerlendirme anlayışında da değişime sebep olacağı düşünülmektedir (Stockmann ve Meyer, 2016). Bu noktada başta Amerika ve Avrupa gibi gelişmiş ülkelerde bu değişim ve gelişime ayak uydurmak için program değerlendirmede yaklaşım ve modellerin kullanımının yanında değerlendirmede belirli standartların karşılanması da üzerinde durulan bir konu haline gelmiştir (Astbury, 2016). Bunun en önemli örneği farklı ülkelerdeki üniversitelerden 400'den fazla paydaşı ile geliştirdiği program değerlendirme standartları konusunda bu alanda dikkatleri üstüne çekmeyi başarmış

Eğitimde Değerlendirme Standartları Ortak Komitesi'nin yürüttüğü çalışmalarıdır (JCSEE, 2018). İlgili komite 1981'den beri program değerlendirme çalışmalarının niteliğini garanti altına alma maksadıyla standartlar geliştirmekte ve bu standartları güncel ihtiyaçlara göre güncellemektedir (Stufflebeam ve Madaus, 1983). Komitenin geliştirdiği standartlar "Checklist Project (Kontrol Listesi Projesi)" ismiyle daha önce ifade edilen alanda kabul görmüş program değerlendirme yaklaşım ve modellerinin öncü isimlerinden Daniel Stufflebeam, Robert Stake, Michael Scriven gibi bilim insanları ve Michael Quinn Patton, Egon Guba, Ernest R. House gibi program değerlendirme alanında ünlü olarak adlandırılacak diğer bazı bilim insanları tarafından program değerlendirme süreçlerinin niteliğini değerlendirecek kontrol listelerine dönüştürülmüştür (WMU, 2020). Araştırma sonuçları Türkiye'de de program değerlendirme çalışmalarının ve bu çalışmalarda standardize uygulamaların giderek yaygınlaşmakta olduğunu göstermektedir. Türkiye'de 2012 yılında 12 yıllık zorunlu eğitim sistemine geçilmesi ve aynı yıl Fırsatları Artırma ve Teknolojiyi İyileştirme (FATİH) Projesinin uygulanması gibi eğitim sisteminde çeşitli değişiklikler söz konusudur (Akıncı, 2017; MEB, 2012). Bu değişiklik ve güncellemeler eğitim sisteminde köklü değişimler olarak algılanmış ve program değerlendirme çalışmalarına ilgiyi artırmış olabilir. Bu durum araştırmada incelenen 10 yıllık süreçte gerçekleştirilmiş çalışmaların ikinci beş yıllık yarıda ilk beş yıla kıyasla yaklaşık üç kat artmış olmasının arkasında yatan sebeplerden biri olabilir.

İncelenen program değerlendirme çalışmalarında Demirel'in Analitik Değerlendirme Modeli dışında Türkiye'de tasarlanan herhangi bir modelin kullanılmaması program değerlendirmede yeni yaklaşım ve modellerin ortaya çıkarılmasında üretkenliğin oldukça düşük olduğu, üretilen çalışmalara da yeterince itibar gösterilmediği şeklinde yorumlanabilir. Yöntemsel tercihler açısından da benzer durumlardan söz etmek mümkündür. Özellikle bazı çalışmalarda yöntemsel tercihlerin ve program adına anlamlı veri üretkenliğinin oldukça sınırlı olduğu görülmektedir. Bu bağlamda araştırma sonuçlarına paralel olarak Türkiye'de eğitim programları ve öğretim alanında yapılan eğilim ve içerik analizi tarzındaki çalışmalarda yöntemsel tercihler bağlamında benzer durumlardan söz edilmektedir (Kozikoğlu ve Senemoğlu, 2015; Ozan ve Köse, 2014; Bıkmaz, Aksoy, Tatar ve Altınyüzük, 2013). Yıllar içerisinde süregelen bu durum program değerlendirmede çalışma konusu bağlamında kapsamlı veri toplama ve analizi süreçlerini gerektiren araştırmaların tercih edilmediğini göstermektedir. Bunun yanında ilgili literatürde program değerlendirme başlığı altında yapılan ve herhangi bir yaklaşım ve modele dayandırılmayan birçok araştırmayı görmek mümkündür. Herhangi bir program değerlendirme yaklaşım ve modelini tercih eden araştırmalardan uluslararası veri tabanlarında indekslenenler ise sınırlı sayıdadır. Türkiye'de yapılan ulusal ve uluslararası akademik çalışmaların bir kısmında kapsamlı veri toplama ve analizi süreçlerinin tercih edilmediği, yapılan çalışmaların bazen puan toplama aracı olarak kullanıldığı bilinmektedir (Ak ve Gülmez, 2006). Herhangi bir modelden yararlanılarak yapılan çalışmaların büyük çoğunluğunda ilgili modellerin kullanım gerekçesinin yeterince açıklanmaması, programın değerlendirilmesine ilişkin sonuçlar ve program hakkında alınması gereken kararlar hakkındaki tutarsızlıklar ve eksikler program değerlendirme kavramını ifade ettiği olgunun yeterince anlaşılmadığını göstermektedir. Çünkü program değerlendirmenin en önemli amaçlarından biri programın herhangi bir özelliği hakkında karar verme süreci olarak ifade edilmektedir (Uşun, 2012). Bu da sayıları ne kadar fazla olursa olsun bu başlık altında yapılan her çalışmanın program değerlendirme araştırması niteliği taşımadığının bir göstergesidir. Akıncı ve Köse (2020) Türkiye'de öğretmen yetiştiren programlara yönelik yaptıkları meta-değerlendirme çalışmasında benzer sonuçlara ulaşmışlardır. İncelenen program değerlendirme çalışmalarının büyük çoğunluğunun yüksek lisans ve doktora tezlerinden oluşması da zaman ve emek gerektiren program değerlendirme araştırmalarının lisansüstü tezleri gibi daha kapsamlı çalışmalarda tercih edilen araştırma türlerinden olduğunu göstermektedir. Tüm bu sorunlar akademik kültür ve yayın kültürünün araştırmaya olan yaklaşım, bağlamında yöntemsel eğilim ve tercihleri etkilemesiyle ilgili olabilir. Akcan, Malkoç ve Kızıltan (2018) Türkiye'de akademik kültürün bilimsel üretkenlikten ziyade eğitim-öğretim odaklı olduğunu ve araştırmaya olan yaklaşımda ciddi sorunlar olduğunu ifade etmişlerdir. Bu durum, program değerlendirme çalışmalarında ağırlıklı olarak araştırmacıların eğitim fakülteleri bünyesindeki lisans programlarının öğretim elemanları ve lisansüstü öğrencilerinin daha kolay ulaşabileceği örneklem grupları üzerinde çalışmayı tercih ettiğinin bir göstergesi olarak yorumlanabilir.

Sonuç olarak Türkiye’de eksikleri ile birlikte program değerlendirmeye yönelik çalışmalar, bu çalışmalarda program değerlendirme yaklaşım ve modellerinden yararlanma ve bu bağlamda yapılacak yeni araştırmalara yönelim söz konusudur. Bundan dolayı çalışma sonuçlarından yola çıkarak gelecekte yapılacak program değerlendirme araştırmalarına katkı sağlayabileceği düşünülen bazı öneriler şöyledir:

- Dünya’daki gelişmeler doğrultusunda Türk eğitim sisteminin ihtiyaçlarına uygun program değerlendirme yaklaşım ve modelleri geliştirilmeli, bu bağlamda çalışmalarda ilgili modellerin tercihinde daha etkin kriterlere başvurulmalı ve tercih edilme gerekçeleri açık bir biçimde ifade edilmelidir.
- Program değerlendirme çalışmaları uluslararası literatürde yer alan program değerlendirme standartlarının kullanılarak ya da Türk eğitim sisteminin ihtiyaçlarına uygun güncel standartlar geliştirilerek bu standartlar ışığında yürütülmelidir.
- Yükseköğretim Kurulu ve Milli Eğitim Bakanlığı program değerlendirme konusunda daha fazla sorumluluk üstelenmeli ve bu kurumlar tarafından kapsamlı program değerlendirme çalışmaları yürütülmelidir.
- Başta lisansüstü programlar olmak üzere daha etkin yönetsel tercihlerle farklı program tür ve kademelerinde program değerlendirme çalışmaları yapılmalıdır.
- Çalışmalarda ulaşılan sonuçlarla uyumlu, programın geleceğine ilişkin alınacak kararlara yönelik önerilerde bulunulmalıdır.
- Farklı araştırmacılar tarafından çeşitli bağlamlarda yapılmış program değerlendirme çalışmalarının niteliğine ilişkin meta-değerlendirme çalışmaları yürütülmelidir.

## References

- Ak, M. Z., & Gülmez, A. (2006). Türkiye'nin uluslararası yayın performansının analizi. *Akademik İncelemeler Dergisi*, 1(1), 22-49.
- 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.
- Akinci, M. (2017). English teacher candidates' self-efficacy towards FATİH project. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, (30), 543-555
- Akinci, M., & Köse, E., (2020). *Türkiye'de Öğretmen Yetiştiren Programlara İlişkin Bir Meta-değerlendirme Çalışması*. Uluslararası Pegem Eğitim Kongresi (pp.174-175). Diyarbakır, Türkiye
- Akyüz, G. (2006). Türkiye ve Avrupa Birliği ülkelerinde öğretmen ve sınıf niteliklerinin matematik başarısına etkisinin incelenmesi. *İlköğretim Online*, 5(2), 75-86.
- 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.
- Aygören, F., & Er, K. O. (2018). Eğitim programlarını değerlendirmeye ait sınıflamalar. *Turkish Studies Eğitim Bilimleri Dergisi*, 13(11), 269-296.
- Bakioğlu, A., & Yıldız, A. (2014). Finlandiya'nın PISA başarısına etki eden faktörler bağlamında Türkiye'nin durumu, *Eğitim Bilimleri Dergisi* 38(1), 37-53.
- Bıkmaz, F. H., Aksoy, E., Tatar, Ö., & Altinyüzük, C. A. (2013). Eğitim programları ve öğretim alanında yapılan doktora tezlerine ait içerik çözümlemesi (1974-2009). *Eğitim ve Bilim*, 38(168), 288-303.
- Cohen, L., Manion, L., & Morrison, K. (2007). Research methods in education. *New York: Routledge*.
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches*. (2. Ed.). Sage publications.
- Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4. Ed.). USA: Pearson Prentice Hall.
- Çalışkan, Ş., Karabacak, M., & Meçik, O. (2013). Türkiye'de eğitim-ekonomik büyüme ilişkisi: 1923-2011 (Kantitatif bir yaklaşım). *Yönetim Bilimleri Dergisi*, 11(21), 29-48.
- Demirel, Ö. (2014). *Eğitimde program geliştirme kuramdan uygulamaya*. (21. Ed.). Ankara: Pegem Akademi Yayıncılık.
- Demirtaş, Z. (2017). Eğitimde program değerlendirme yaklaşımlarına genel bir bakış. *Sakarya University Journal of Education*, 7(4), 756-768.
- Eisner, E. W. (2002). *The arts and the creation of mind*. London. Yale University Press.
- Erdoğan, İ. (2001). Sosyal bilimlerde pozitivist-ampirik akademik araştırmaların tasarım ve yöntem sorunları. *Anatolia: Turizm Araştırmaları Dergisi*, 12(2), 119-134.
- Ertürk, S. (2013). *Eğitimde "program" geliştirme* (6. Ed.). Ankara: Edge Akademi Yayınları.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Curriculum evaluation: Alternative approaches and practical guidelines*. (4. Ed.). 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.
- Hammond, R. L. (1967). Evaluation at the local level. Adress to the Miller Committee for the National Study of ESEA Title III. ERIC, Retrieved from the database ED016547.
- House, E. R. (1990). Trends in evaluation. *Educational Researcher*, 19(3), 24-28.

- JCSEE. (2018). *Checklist of the program evaluation standards statements*. Retrieved from <https://evaluationstandards.org/program>.
- Karakaya, A. (2012). Bilimsel araştırma yöntemleri. A. Tanrıoğen (Ed.), In *Bilimsel araştırma yöntemleri (3. Baskı)* (pp. 57-83). Ankara: Anı Yayıncılık.
- Koç, E. S. (2015). Türkiye de İlköğretim Programlarının Değerlendirilmesine Yönelik Yapılan Lisansüstü Tezlerin İncelenmesi (2005–2014). *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 16(1), 198-216.
- 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.
- Işığışok, Ö. (2002). Türkiye’de yaşanan son ekonomik krizlerin sosyo-ekonomik sonuçları: Kriz işsizliği ve beyin göçü. *ISGUC The Journal of Industrial Relations and Human Resources*, 4(2). Retrieved from <http://dergipark.org.tr/iscguc/issue/25515/269102>.
- MEB, (2012). *Genelge 2012/20 12 yıllık zorunlu eğitime yönelik uygulamalar*. Retrieved from [www.meb.gov.tr/haberler/2012/12YillikZorunluEgitimeYonelikGenelge.pdf](http://www.meb.gov.tr/haberler/2012/12YillikZorunluEgitimeYonelikGenelge.pdf) adresinden alınmıştır.
- Meyer, W., Funk, E., & Nguyen, P. L. (2016). Participation and Valuation: Ongoing Methodological Challenges. In Stockmann, R., & Meyer, W. (Ed). *The Future of Evaluation* (pp. 294-308). Palgrave Macmillan, London.
- Metfessel, N. S., & Michael, W. B. (1967). A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. *Educational and Psychological Measurement*, 27(4), 931-943.
- Mutlu, G. (2018). Curriculum evaluation study of the main course at a preparatory program: a case study. *Turkish Online Journal of Qualitative Inquiry*, 9(3), 202-239.
- Ornstein, A. C., & Hunkins, F. P. (2018). *Curriculum: Foundations, principles, and issues*. (2. Ed.). Edinburgh: Pearson Education Limited.
- Ozan, C., & Köse, E. (2014). Eğitim programları ve öğretim alanındaki araştırma eğilimleri. *Sakarya University Journal of Education*, 4(1), 116-136.
- Özudoğru, F. (2018). Analysis of curriculum evaluation studies conducted in foreign language education: 2005-2016. *Journal of Language and Linguistic Studies*, 14(2), 113-134.
- Parlett, M., & Hamilton, D. (1972). Evaluation as Illumination: A New Approach to the Study of Innovative Programs. Occasional Paper.
- Provus, M. M. (1969). The Discrepancy Evaluation Model: An Approach to Local Program Improvement and Development.
- Scriven, M. (2009). Meta-evaluation revisited. *Journal of multidisciplinary evaluation*, 6(11), 3-8.
- Smidt, A., Balandin, S., Sigafoos, J., & Reed, V. A. (2009). The Kirkpatrick model: A useful tool for evaluating training outcomes. *Journal of Intellectual and Developmental Disability*, 34(3), 266-274.
- Sözbilir, M., & Kutu, H. (2008). Development and Current Status of Science Education Research in Turkey. *Essays in Education* [Special issue], 1-22.
- Stake, R. E. (1967). *The countenance of educational evaluation*. Department for Exceptional Children, Gifted Children Section.
- Stake, R. E. (2004). *Standards-based & responsive evaluation*, London, New Delhi: Sage Publications.
- Stockmann, R., & Meyer, W. (2016). *The future of evaluation: Global trends, new challenges and shared perspectives*. Palgrave Macmillan, London.



- Stufflebeam, D. L. (1999). *Foundational models for 21st century curriculum 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. Ed.). San Francisco: Josey-Bass.
- 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., Madaus, G. F., & Kellaghan, T. (Eds.). (2000). *Evaluation models: Viewpoints on educational and human services evaluation* (2. Ed.). 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.
- Ulutaş, F., & Ubuz, B. (2008). Matematik eğitiminde araştırmalar ve eğilimler: 2000 ile 2006 yılları arası. *İlköğretim Online*, 7(3), 614-626.
- Uşun, S. (2012). *Eğitimde program değerlendirme: Süreçler yaklaşımlar ve modeller*. Ankara: Anı Yayıncılık.
- Varişoğlu, B., Şahin, A., & Göktaş, Y. (2013). Türkçe eğitimi araştırmalarında eğilimler. *Kuram ve Uygulamada Eğitim Bilimleri*, 13(3), 1767-1781.
- Warju, W. (2016). Educational curriculum evaluation using CIPP model. *Innovation of Vocational Technology Education*, 12(1), 36-42.
- WMU. (2020). *Evaluation Checklists*. Retrieved from <https://wmich.edu/evaluation/checklists>.
- Yapıcıoğlu, K. D., Kara, A. D., & Yalçınoğlu, P. (2016). Türkiye’de program değerlendirme çalışmalarında eğilimler ve sorunlar: Alan uzmanlarının görüşüyle. *Uluslararası Eğitim Programları ve Öğretim Çalışmaları Dergisi*, 12(6), 91-113.
- Yetkiner, A., Erdol, T. A., & Ünlü, Ş. (2019). Content analysis of phd dissertations on curriculum evaluation (1996-2017). *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, 21(1), 247-269.
- Yıldırım, A., & Şimşek, H. (2013). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri* (9. ed.). Ankara: Seçkin Yayıncılık.
- Yüksel, İ., & Sağlam, M. (2014). *Eğitimde program değerlendirme*. (2. Ed.). Pegem Akademi.
- Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misulis, K. (2011). Using the context, input, process, and product evaluation model (CIPP) as a comprehensive framework to guide the planning, implementation, and assessment of service-learning programs. *Journal of Higher Education Outreach and Engagement*, 15(4), 57-84.

## Examined Research Papers

- Akdoğdu, E. (2016). *Sınıf öğretmenliği lisans programının öğretmen adaylarının görüşleri doğrultusunda bağlam, girdi, süreç ve ürün (CIPP) modeli ile değerlendirilmesi* (Master's Thesis, Muğla Sıtkı Koçman University, Department of Educational Sciences, Muğla). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Altındağ, A. (2017). *Ortaokul 5. sınıf matematik dersi öğretim programının Stake'in uygunluk-olasılık modeline göre değerlendirilmesi* (PhD Disstertation, Hacettepe University, Department of Educational Sciences, Ankara). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Aslan, M. (2015). *Eğitim fakültelerindeki öğretmenlik uygulaması dersinin değerlendirilmesi ve öğretim programının hazırlanması* (PhD Disstertation, Anadolu University, Department of Educational Sciences, Eskişehir). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Aslan, M., & Çıkar, İ. (2017). Evaluation of 4 th Grade Mathematics Curriculum By Tyler's Objective Based Evaluation Model. *Necatibey Faculty of Education Electronic Journal of Science & Mathematics Education*, 11(2), 172-196.
- Aslan, M., & Erden, R. Z. (2018). Evaluation of 5<sup>th</sup> grade science curriculum. *Necatibey Faculty of Education Electronic Journal of Science & Mathematics Education*, 12(2), 508-537.
- Aydın, S., Şentürk, Ş., & Duran, V. (2018). Evaluation of pre-school curriculum according to Stufflebeam's CIPP model. *Electronic Turkish Studies*, 13(27), 163-181.
- Babacan, T. (2016). *Dokuzuncu sınıf İngilizce dersi öğretim programının bütüncül eğitime göre değerlendirilmesi* (PhD Disstertation, Eskişehir Osmangazi University, Department of Educational Sciences, Eskişehir). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Bağcıoğlu, G. (2019). *2015 ve 2018 ilkokul dördüncü sınıf Türkçe öğretim programının CIPP program değerlendirme modeline göre değerlendirilmesi* (Master's Thesis, Afyon Kocatepe University, Department of Educational Sciences, Afyonkarahisar). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Balıkçı, Ç. (2019). *3. sınıf fen bilimleri öğretim programının CIPP modeline göre değerlendirilmesi* (Master's Thesis, Uşak University, Department of Educational Sciences, Uşak). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Başaran, S. T., & Ulubey, Ö. (2018). 2013 okul öncesi eğitim programının değerlendirilmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 51(2), 1-38.
- Bayat, S. (2012). *Stake'in uygunluk/olasılık modeline göre ilkokuma yazma programının değerlendirilmesi* (PhD Disstertation, Abant İzzet Baysal University, Department of Educational Sciences, Bolu). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Bayık, S. (2018). *Saç bakımı ve güzellik hizmetleri ön lisans programının Metfessel – Michael modeline göre değerlendirilmesi* (Master's Thesis, Aydın Adnan Menderes University, Department of Educational Sciences, Aydın). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Caner, H. N. (2018). *Öğretmenlik uygulaması dersinin bağlam, girdi, süreç, ürün modeline göre değerlendirilmesi: İngilizce öğretmenliği lisans programı örneği* (Master's Thesis, Akdeniz University, Department of Educational Sciences, Antalya). <https://tez.yok.gov.tr/UlusalTezMerkezi> adresinden alınmıştır.
- Canoğlu, S. N. (2014). *Türkçe ve matematik öğretim programlarının değer tabanlı program değerlendirme modeline göre incelenmesi* (Master's Thesis, Kırıkkale University, Department of Educational Sciences, Kırıkkale). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Cansu, T. (2010). *Anadolu üniversitesi ilköğretimde teknoloji uygulamaları e-sertifika programının öğrenen görüşüne göre bağlam, girdi, süreç ve ürün (CIPP) modeli ile değerlendirilmesi* (Master's

- Thesis, Anadolu University, Department of Distance Education, Eskişehir). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Çelik, K. (2018). *Ortaöğretim İngilizce dersi öğretim programı'nın (2014) Eisner modeline göre değerlendirilmesi* (Master's Thesis, Gazi University, Department of Educational Sciences, Ankara). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Çiftçi, F. (2019). *Ortaöğretim 9. sınıf 2011 Türk edebiyatı dersi öğretim programının CIPP modeline göre değerlendirilmesi* (Master's Thesis, Gazi University, Department of Turkish Language and Literature, Ankara). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Çimili Abat, E. Z. (2016). *9. sınıf matematik dersi öğretim programının bağlam, girdi, süreç, ürün değerlendirme modeline göre değerlendirilmesi* (Master's Thesis, Akdeniz University, Department of Educational Sciences, Antalya). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Diñçer, B. (2013). *7. sınıf İngilizce öğretim programının Stufflebeam'in Bağlam-Girdi-Süreç-Ürün (CIPP) modeline göre değerlendirilmesi* (PhD Disstertation, Aydın Adnan Menderes University, Department of Educational Sciences, Aydın). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Duman, S. N. (2019). *Hizmet öncesi öğretmen eğitiminde öğretmenlik meslek bilgisi derslerinin değerlendirilmesi* (PhD Disstertation, Gazi University, Department of Educational Sciences, Ankara). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Hidroğlu, Ç. N., Kandemir, A., & Tuncel, İ. (2016). Evaluation of teaching principles and methods' course in scope of Hammond's evaluation cube. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 1(37), 47-68.
- İlhan, E. (2018). *Üniversitelerin lisans programlarında uygulanan çekirdek programın değerlendirilmesi* (PhD Disstertation, Gazi University, Department of Educational Sciences, Ankara). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Karcı Aktaş, C. (2018). *Aydın Adnan Menderes Üniversitesi Yabancı Diller Yüksek Okulu İngilizce hazırlık programının Bellon ve Handler modeline göre değerlendirilmesi: Bir durum çalışması* (PhD Disstertation, Aydın Adnan Menderes University, Department of Educational Sciences, Aydın). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Kavgaoğlu, D. (2017). *Çağrı merkezi mesleki yetkinlik geliştirme eğitim programının değerlendirilmesi* (PhD Disstertation, Yıldız Technical University, Department of Educational Sciences, İstanbul). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Kavgaoğlu, D., & Elkin, N. (2019). Hemşirelik Bölümü Kişilerarası İlişkiler ve Terapötik İletişim Dersi Öğretim Programının Değerlendirilmesi. *İstanbul Gelişim Üniversitesi Sağlık Bilimleri Dergisi*, (7), 670-696.
- Kayhan, E., & Gürol, M. (2019). Türkçe Öğretim Programı (2017)'nin Stufflebeam'in (CIPP) Modeline Göre Değerlendirilmesi. *Türkiye Eğitim Dergisi*, 4(1), 48-67.
- Kotluk, N., & Yayla, A. (2016). Ortaöğretim 9. Sınıf Fizik Öğretim Programının Tyler'in Hedefe Dayalı Değerlendirme Modeline Göre Değerlendirilmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*. 16(4), 1832-1852.
- Kumral, O. (2010). *Eğitsel eleştiri modeli ile eğitim fakültesi sınıf öğretmenliği öğretim programının değerlendirilmesi: bir durum çalışması* (PhD Disstertation, Aydın Adnan Menderes University, Department of Educational Sciences, Aydın). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Kurt, A. (2016). *4. sınıf İngilizce dersi öğretim programının bağlam, girdi, süreç, ürün modeline göre değerlendirilmesi* (Master's Thesis, Akdeniz University, Department of Educational Sciences, Antalya). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.

- Kuş, M. (2019). *Türkiye'de uygulanan Abitur programlarının değerlendirilmesi: Bir durum çalışması* (PhD Disstertation, Yıldız Technical University, Department of Educational Sciences, İstanbul). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Kuzu, S. (2015). *Öğretim ilke ve yöntemleri ders programının değerlendirilmesi* (PhD Disstertation, Gaziantep University, Department of Educational Sciences, Gaziantep). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Mutlu, G. (2018). Curriculum evaluation study of the main course at a preparatory program: a case study. *Turkish Online Journal of Qualitative Inquiry*, 9(3), 202-239.
- Orhan, A. (2016). *Uzaktan eğitimle yürütülen yabancı dil dersi öğretim programının bağlam, girdi, süreç ve ürün (CIPP) modeli ile değerlendirilmesi* (Master's Thesis, Düzce University, Department of Educational Sciences, Düzce). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Ödemiş, S. İ. (2018). *Meslek yüksekokulu İngilizce öğretim programının bağlam girdi süreç ürün (CIPP) modeli kullanılarak değerlendirilmesi* (PhD Disstertation, Eskişehir Osmangazi University, Department of Educational Sciences, Eskişehir). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Özaltaş Serçek, G. (2014). *Önlisans turizm eğitim programının CIPP modeline göre değerlendirilmesi* (PhD Disstertation, Dicle University, Department of Educational Sciences, Diyarbakır). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Özdil, N. G. (2016). *Sosyal bilimler lisesi programının değerlendirilmesi* (Master's Thesis, Gazi University, Department of Educational Sciences, Ankara). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Özüdoğru, F. (2016). *İlkokul 2. sınıf İngilizce öğretim programının diller için Avrupa ortak başvuru metni doğrultusunda aydınlatıcı değerlendirme modeli ile değerlendirilmesi* (PhD Disstertation, Anadolu University, Department of Educational Sciences, Eskişehir). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Şeref, M., & Gözütok, F. D. (2018). Bir program değerlendirme örneği: doğum eğitimi. *Ihlara Eğitim Araştırmaları Dergisi*, 3(1), 64-77.
- Şişman, Y. S. (2019). *Üniversite İngilizce hazırlık öğretim programının değerlendirilmesi: Bir durum çalışması* (Master's Thesis, Necmettin Erbakan University, Department of Educational Sciences, Konya). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Tekmen, B. (2012). *Okul öncesi öğretmenliği programının akademisyenlerin bakış açısıyla değerlendirilmesi* (PhD Disstertation, Orta Doğu Technical University, Department of Educational Sciences, Ankara). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Tunç, F. (2010). *Ankara Üniversitesi Hazırlık Okulu programının CİPP modeli ile değerlendirilmesi* (Master's Thesis, Orta Doğu Technical University, Department of Educational Sciences, Ankara). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Uslu, Ö. (2013). *Bilgi ve iletişim teknolojileriyle bütünleşmeyi amaçlayan bir mesleki gelişim programının değerlendirilmesi ve bir model önerisi* (PhD Disstertation, Ege University, Department of Educational Sciences, İzmir). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Yeşilyurt, E. (2010). *Öğretmenlik uygulaması öğretim programının standart temelli ve ihtiyaca cevap verici modeller ışığında değerlendirilmesi* (PhD Disstertation, Fırat University, Department of Educational Sciences, Elazığ). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- Yurdakul, B., Uslu, Ö., Çakar, E., & Yıldız, D. G. (2014). Evaluation of the professional development program on web-based content development. *Educational Sciences: Theory Practice*, 14(4), 1409-1437.