International Journal of Curriculum and Instructional Studies

***13*(2), 2023, 325-356**

www.ijocis.com

A Systematic Meta-Evaluation of Curriculum Evaluation Research Studies Conducted between 2004-2022 on K12 School Curricula in Türkiye

Derya Şahin İpek, Eötvös Loránd University, derya@student.elte.hu, 0009-0005-9929-1884

Elif Tuğçe Gezer, University of Luxembourg, eliftugce.gezer@uni.lu, 0000-0001-5158-6919

Funda Kelahmetoğlu Tunçer, Middle East Technical University, fundak@metu.edu.tr, 0000-0001-9578-1855

Hasan Şerif Baltacı, Başkent University, hsbaltaci@baskent.edu.tr, 0000-0003-3199-7838

Ahmet Ok, Middle East Technical University, as@metu.edu.tr, 0000-0002-5941-4158

Keywords

Curriculum evaluation
Metaevaluation
Systematic review
K12
School curricula

Article Info:

Received : 18-06-2023
Accepted : 08-12-2023
Published : 30-12-2023

Abstract

This study aimed to carry out a meta-evaluation of the curriculum evaluation research studies conducted in K12 schooling levels between 2004 and 2022 in Turkey based on the Joint Committee's standards for educational evaluation to present an overview of the evaluation studies and enhance understanding of evaluation. Through a systematic review, 42 studies were included in the meta-evaluation. The data were collected through a descriptive matrix and a meta-evaluation checklist and were analyzed through document analysis. The findings revealed that most of the eligible articles were published between 2018-2022. Most studies focused on primary and secondary education, and only a few of them dwelled upon preschool education. Among the studies, Mathematics and English are the most evaluated courses. Moreover, data were often gathered from one or two data sources, such as teachers and students, through qualitative research designs and without referring to any evaluation models. Besides, curriculum evaluation research studies made recommendations regarding practice and further research and mainly focused on including stakeholders, in-service training, and curriculum development and evaluation. As for the data gathered through the meta-evaluation checklist, the results showed that most articles demonstrated more than half of the indicators stated in the items of the checklist. While utility seemed to be a successful area for the articles evaluated, feasibility and accuracy criteria were not fully met, and most articles did not refer the propriety criteria. In line with the findings, the evaluation studies need to be planned in a way to provide a more complete and deeper understanding of the evaluated curricula by considering its complexities.

DOI:10.31704/ijocis.2023.014

To cite this article: Şahin-İpek, D., Gezer, E. T., Kelahmetoğlu-Tunçer, F., Baltacı, H. Ş. & Ok, A. (2023). A systematic meta-evaluation of curriculum evaluation research studies conducted between 2004-2022 on K12 school curricula in Türkiye. *International Journal of Curriculum and Instructional Studies, 13*(2), 325-356. https://doi.org/10.31704/ijocis.2023.014

Introduction

The validity and reliability of the decisions based on curriculum evaluation studies depend on the evaluation process (Sağlam & Yüksel, 2007); therefore, evaluation should be carried out cautiously and thoroughly. Scriven (1969) and Stufflebeam (2000) emphasize the importance of the quality of evaluation as they state that producing inaccurate or biased reports may lead to erroneous results. Stufflebeam (2000) further accentuates that without being subjected to a meta-evaluation, evaluation reports might cause the audience to make ill-structured decisions. Originated by Scriven (1969), meta-evaluation refers to the task of gathering and utilizing descriptive information to guide and report on the strengths and weaknesses or combining several studies' findings, analyzing them, and deriving conclusions (Sağlam & Yüksel, 2007; Stufflebeam, 2000). It examines whether an evaluation adequately reveals an object's merit, provides valuable guidance for decision-making, and is ethical and practical in terms of resource usage (Stufflebeam, 1978). Besides, it identifies the quality of the processes and findings through a systematic review (Cook & Gruder, 1978; Greene et al., 1992; Scriven, 1991).

Stufflebeam (2000) states that meta-evaluations are of public concern due to their significance in analyzing assessment systems, new curricula, equipment, or technologies. Therefore, they should be based on professional standards. Thus, he defined meta-evaluation as gathering and utilizing information regarding an evaluation's utility, feasibility, propriety, and accuracy to lead an evaluation and inform people about the strengths and shortcomings of an evaluation study (Fitzpatrick et al., 2004; Stufflebeam, 2001; Stufflebeam, 2004).

Considering the purpose and importance of curriculum evaluation, there was a growing focus on drawing theoretical frameworks and setting standards for evaluation studies. Besides, formal curriculum evaluation started to appear as a maturing field, and its development accelerated from the 1960s to the present (Fitzpatrick et al., 2004). The same trend applies to Türkiye since there was a growing interest in conducting curriculum evaluation studies. Accordingly, since the 1950s, curriculum development and evaluation studies have begun to be conducted more systematically and scientifically as a reflection of the recognition of curriculum development as a field of study worldwide (Özdemir, 2009).

In Türkiye, the curriculum evaluation needs to be examined with the curriculum development studies to get meaningful insights since they have mostly intertwining processes. On the official basis, the first documented curriculum evaluation study conducted in Türkiye was on the 1948 primary school curriculum (Turgut, 1983), and this was followed by other curriculum development and evaluation efforts on the 1968 curriculum (Korkmaz, 2020) in the 1960s; with the cooperation between MoNE and TUBİTAK on modern science and mathematics curriculum for selected schools between 1972 and 1975 (Demirbaş & Yağbasan, 2005; Özdemir, 2009). Besides growing efforts, some issues might be seen such as focusing on specific subjects, specific school levels, and being far from following incremental policies.

The National Education Development Project was launched in 1990 to enhance the quality of the education system. One of the aims of this project was to improve the quality of the curricula and educational materials (MoNE, 1995). To assure representativeness, from 1992 to 1997, a pilot study for Curriculum Laboratory Schools (CLS) was implemented through chosen 208 schools from the seven regions of the country. These schools served as the test sites for

the new curricula and educational materials, allowing for a comprehensive evaluation of their efficacy and the identification of areas that required improvement.

On the other hand, the 2005 primary school curricula were developed based on the two-year studies of field trial and revision as an educational reform (MoNE, 2006). Even though the curriculum development process was not initiated based on systematic curriculum evaluation studies, MoNE made evaluations by comparing the existing curricula and the new curriculum in Turkish, social studies, science and technology, and mathematics (Özdemir, 2009). A thorough evaluation of the 2005 curriculum was conducted by a group of academics (ERG, 2005). This evaluation report provides a comprehensive comparison of the aforementioned curricula through external criteria, such as the former or foreign curricula, and internal criteria, such as the examination of the objectives, sample activities, explanations, and assessment procedures. Following the 2005 curriculum, MoNE initiated studies to improve and revise the curriculum in 2008 and 2009; however, these efforts remained limited to minor revisions rather than leading to fundamental changes (Koç, 2016).

More recently, in the 2012-2013 academic year, the structural changes regarding the extension of compulsory education from 8 to 12 years and its division into three stages (4+4+4) led to an updated curriculum to be applied gradually (MoNE, 2012). In the 2016-2017 academic year, MoNE initiated curriculum revision studies for 51 subjects to meet the needs of individuals and society in line with the changing educational approaches (MoNE, 2017). Even though the draft curriculum was shared to get feedback from the public, how the evaluations were held and to what extent the feedback was reflected were not explained. Following, in the 2019-2020 academic year, MoNE (2020) conducted a curriculum evaluation study on the implementation process of an updated curriculum for 31 subjects both in primary and secondary levels in 2018-2019 based on the teachers' opinions. This evaluation study was limited to teachers' opinions and produced results for subjects separately on the level of revision of acquisitions.

Despite the growing number of curriculum evaluation studies in Türkiye, several problems exist. To begin with, it was seen that evaluation studies do not focus on all components of the curriculum (Ünal et al., 2004; Yaşar, 1998), which limits the reliability of the findings. Moreover, Kürüm-Yapıcıoğlu et al. (2016) mentioned that the stakeholders included in the evaluation studies conducted in Türkiye were generally teachers and students, which hinders the multiple perspectives represented in these studies. Besides, Yaşar (1998) stated that the implementation of findings in practice was inhibited as the majority of the evaluation studies were conducted for academic purposes and had limited cooperation with the Ministry of National Education (MoNE). Similarly, Kürüm-Yapıcıoğlu et al. (2016) argued that implementing the findings into practice was an essential concern in evaluation studies. However, they were not sufficiently used in practice, and the same curricula were re-evaluated without making efforts for refinement based on the previous studies (Gökmenoğlu, 2014; Kürüm-Yapıcıoğlu et al., 2016). Another problem with curriculum evaluation studies in Türkiye was the communication of the evaluation findings. Stufflebeam (2000) stated that meta-evaluations are of public concern due to their significance in evaluating several aspects of education systems. Thus, it is crucial to share the evaluation findings with stakeholders and audiences. Although there have been several evaluation studies, it is hard to claim that curriculum evaluation is carried out fully transparently.

Considering the importance attributed to curriculum evaluation and the number of studies that have been conducted, along with the problems identified about prior studies, a systematic meta-evaluation of curriculum evaluation research studies in Türkiye in light of the program evaluation standards proposed by the Joint Committee on Standards for Educational Evaluation may provide a holistic understanding of trends, procedures, and strengths and weaknesses of these studies. Although there were some studies presenting general trends of included curriculum evaluation studies regarding their focus, it was seen that the focus in some of these studies was on certain dimensions of curriculum (e.g., Kazu & Aslan, 2012); specific courses (e.g., Ertekin & Bozkurt, 2020; Kablan, 2011) and general trends in methodological and content related features (e.g., Akşan & Baki, 2017; Koç, 2016). Moreover, it was seen that previous systematic reviews descriptively presented the general trends in curriculum evaluation, and a comprehensive systematic meta-evaluation study has not been found through the scanning of the databases included in the present study.

Thus, in this study, the standards proposed by the Joint Committee on Standards for Educational Evaluation were shortened and adapted to the Turkish context by the authors. It is believed that such a study will be significant in guiding the refinement of the previous evaluation studies and shed light on future curriculum development and evaluation initiatives. In this regard, this study aimed to carry out a meta-evaluation of the curriculum evaluation research studies conducted in all subjects in K12 schooling levels between 2004 and 2022 in Türkiye based on the four standards (i.e., utility, feasibility, propriety, and accuracy) proposed by the Joint Committee to present a general picture of the previous curriculum evaluation research studies and to enhance evaluators' understanding of quality and sound evaluation (Fitzpatrick et al., 2004; Stufflebeam, 2001; Stufflebeam, 2004).

This study is limited to the curriculum evaluation research studies conducted in K12 schooling levels between 2004 and 2021 because it is claimed that starting from 2004, the curricula developed for K12 schooling levels have begun to reflect a more constructivist, progressivist, and student-centered understanding (Bulut, 2006; Gözütok et al.), and this change also encouraged an increase in the curriculum evaluation research studies.

Concerning the purpose of the study, the main research questions guiding this study are:

1. What are the general trends in curriculum evaluation studies conducted in K12 schooling levels between 2004 and 2022 in Türkiye in terms of year of publication, course, schooling level, evaluation approach/model, research design, type, and the number of data sources, and scope of the evaluation study?

2. In what aspects do the curriculum evaluation research studies make recommendations regarding various components of school curricula?

3. To what extent do these evaluation studies conform to utility, feasibility, propriety, and accuracy standards set for curriculum evaluation of programs by the Joint Committee?

Method

Research Design

Within the scope of the current study, a systematic meta-evaluation of curriculum evaluation research studies was utilized to provide a holistic understanding of trends, procedures, and

strengths and weaknesses of these studies. Through focusing on relevant questions and employing sound criteria for inclusion and exclusion, a systematic review helps researchers synthesize existing evidence to come up with answers and statements of conclusion (Harris et al., 2014). After the systematic review was carried out, a meta-evaluation was performed by evaluating the curriculum evaluation research studies conducted in K12 based on the standards created by the Joint Committee on Standards for Educational Evaluation (Stufflebeam, 1981). According to Scriven (1969), meta-evaluation refers to the evaluation of evaluations performed to inform the evaluation and reveal the strengths and weaknesses of the evaluations under examination by investigating the standards of utility, feasibility, propriety, and accuracy (Stufflebeam, 2000).

Data Collection Instruments

Two data collection instruments were utilized throughout the study. First, the researchers prepared a descriptive matrix table to situate the general trends in the curriculum evaluation research studies and the aspects on which the recommendations made in these studies provide insights into curriculum development and curriculum evaluation. This included "year of publication, evaluation approach/model, research design, school level, course, data source, and the categories of recommendations" domains, and it was used to record the essential information about articles included in this study. Thus, it was employed to answer the first and second research questions of the study.

Second, the researchers prepared a checklist named "The Meta-Evaluation Checklist" (see Appendix-I) for the third question to evaluate how these studies comply with the four standards proposed by the Joint Committee (i.e., utility, feasibility, propriety, and accuracy). The primary concern was to prepare an instrument that is concise, understandable, applicable, relevant to the Turkish context, and fit for the purpose (i.e., evaluating the curriculum evaluation research studies). The checklist consisted of four sections, each corresponding to the standard areas proposed by The Joint Committee. Each standard included a number of items. The researchers prepared the items by considering the definitions of the standard provided by Stufflebeam (1981) and examining Stufflebeam's (1999) checklist and Yüksel's (2010) adaptation of this checklist to Turkish. After the instrument was developed, expert opinion was taken from a professor from a public university in Ankara, who specialized in curriculum studies, in relation to the general format of the instrument and the clarity and the appropriateness of the items to each standard. The aim was to ensure the face and content validity of the instrument (Fraenkel et al., 2012). Accordingly, revisions were made, and the instrument was translated into Turkish. Then, both versions were examined by the same expert to establish language uniformity.

After revising the instruments based on the recommendations (e.g., shortening some statements, revising some words considering contextual understanding, and clarifying some concepts such as stakeholders), the final versions of the instruments were formed. The original checklist by the Joint Committee focused more on the clients' needs and requirements because it was designed in the American context, where institutions or individuals employ curriculum specialists to evaluate their programs. However, in the Turkish context, all the studies evaluated were in line with academic endeavors and were not conducted with a business motive. Instead, they are evaluation studies initiated by academics to contribute to the field of curriculum. That

is why items like U1*[1] and F1*[2] were omitted from the checklist with the rationale that they did not apply to the Turkish curriculum evaluation studies' context. Also, some items were rephrased to fit in the context better. For example, in item A1*[3], there is an emphasis on the clients. The final version of the instrument consisted of 29 items in four domains which correspond to the utility, feasibility, propriety, and accuracy standards (see Appendix-I). Using this instrument, curriculum evaluation research studies were evaluated based on each standard and overall quality.

Selection of the Curriculum Evaluation Research Studies

To select the studies for the current meta-evaluation, a systematic literature search was carried out in the databases of DergiPark, WoS, Scopus, and EBSCOhost, which cover English and Turkish literature comprehensively and were last accessed in February 2022. Programme evaluation studies conducted in K12 settings in Türkiye between 2004 and 2022 were searched. The search terms in both Turkish (DergiPark) and English (Web of Science, Scopus, EBSCOhost, DergiPark) for the databases are presented in Table 1.

Table 1

The Search Terms for the Databases (DergiPark, Web of Science, Scopus, EBSCOhost)

Databases	*Searching keywords*
Web of Science, Scopus, EBSCOhost, DergiPark	(program OR curriculum) AND (evaluation OR assessment OR examination OR analysis)
DergiPark	In Turkish: (Program VEYA öğretim programı VEYA eğitim programı VEYA müfredat) VE (değerlendirme VEYA değerlendirilmesi VEYA inceleme VEYA incelenmesi VEYA analizi)

During the article selection process, Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Liberati et al., 2009) protocol was used. The initial search carried out by focusing on the title, keywords, and abstract of each record yielded a total of 1120 studies in the aforementioned databases. After the studies not conducted in Türkiye and those not within K12 context were eliminated, 263 articles were left to be assessed for eligibility. Upon examining the full texts of the articles based on the inclusion-exclusion criteria and removing the duplicates, 42 evaluation research studies were included for the meta-evaluation, as can be seen in Figure 1.

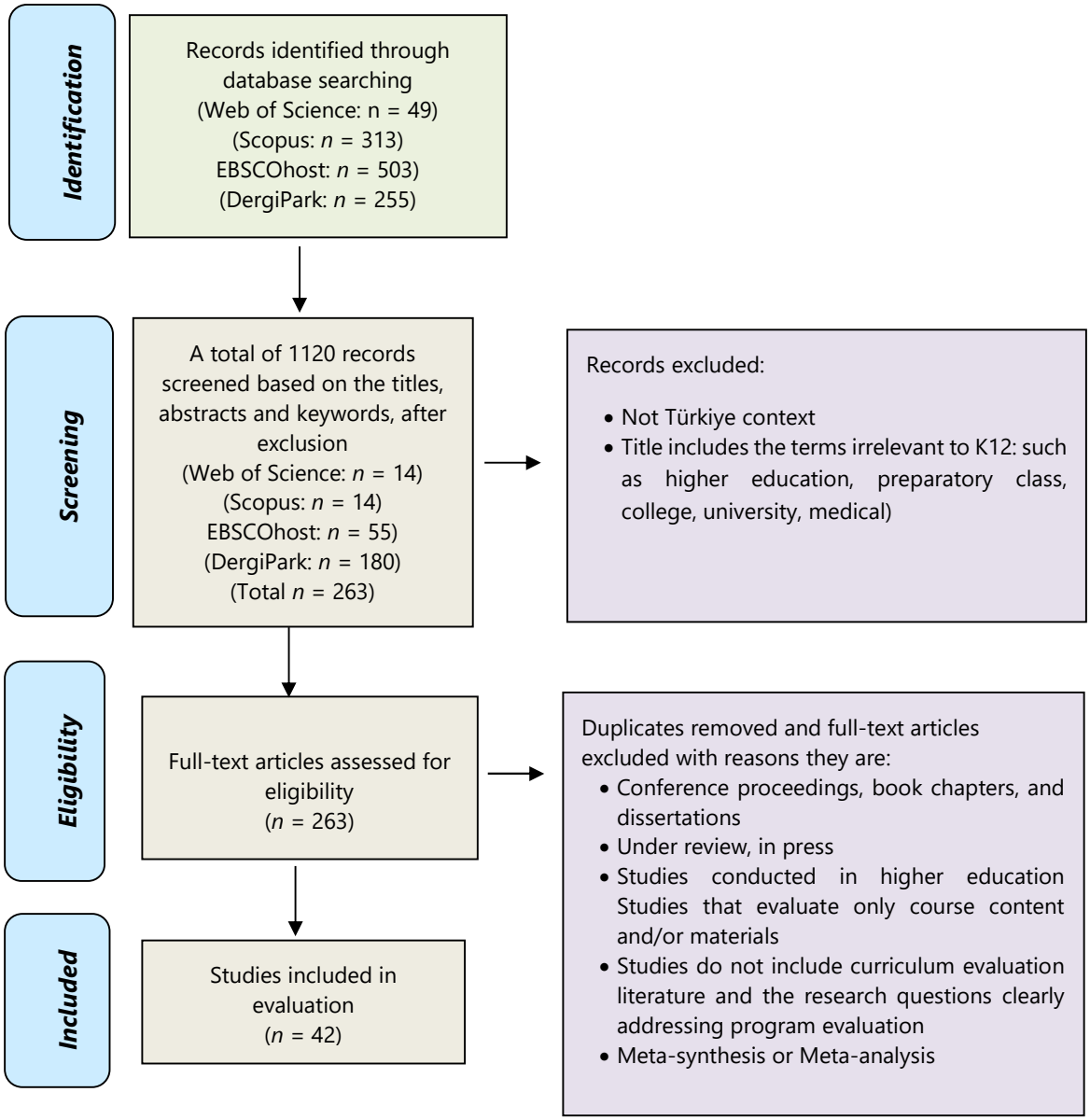
[1] U1. Clearly identify the evaluation client

[2] F1. Appoint competent staff and train them as needed

[3] A1. Collect descriptions of the intended program from various written sources and from the client and other key stakeholders

Figure 1

PRISMA Flow Diagram



Data Analysis

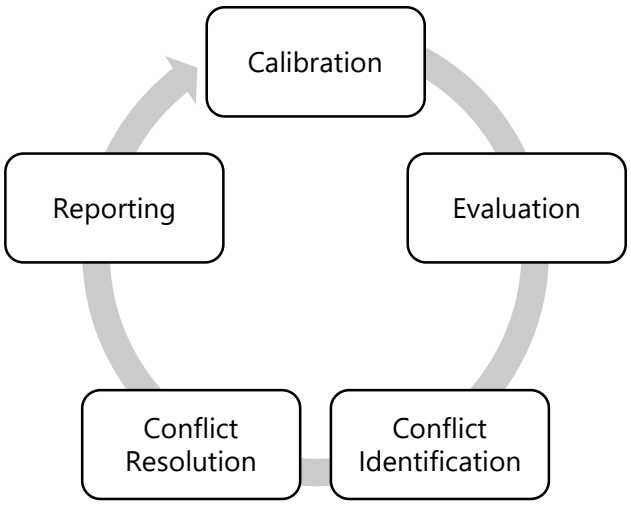
The current study employed document analysis and descriptive analysis to analyze the data. Document analysis requires a systematic process for evaluating or reviewing documents (Bowen, 2009). To answer the first and the second research questions, the researchers used document analysis. O'Leary (2014) highlighted that it requires creating a pool of texts to explore and considering how they will be accessed at the beginning. Within the data analysis process for the first research question, the selected articles were shared among the researchers after a session was held to discuss how descriptive information (i.e., year of publication, evaluation approach/model, research design, schooling level, course, data source, and the categories of recommendations) was to be recorded to the descriptive matrix table for each

article. Then they were presented in figures and tables by reporting frequencies and percentages.

In order to answer the second research question, the recommendations made in selected curriculum evaluation research studies were reviewed using content analysis. First, the researchers coded the recommendations separately based on recommendation areas. Then they cross-checked them to finalize the emerging codes. After that, for each recommendation category, the number of articles that made that particular recommendation was reported, and the ratio of each category to all recommendations made in the selected articles was provided. Finally, regarding the third research question, selected studies were evaluated based on the Meta-Evaluation Checklist. The researchers endeavored to situate the studies in line with the program evaluation standards mentioned to evaluate the quality of the curriculum evaluation.

Figure 2

The Data Analysis Cycle for the Metaevaluation



As can be seen in Figure 2, the data analysis for the third research question consisted of five consecutive stages. Firstly, before data analysis, the researchers evaluated three studies by using the Meta-Evaluation Checklist, got involved in a calibration session to eliminate inconsistencies and conflicts of any kind to agree upon the criteria to be used while judging the evaluation studies (Stuffleam, 2000). The calibration of the raters generally follows a three-step process in which researchers firstly come up with a manual or a checklist that is well-established and discussed. Secondly, the researchers randomly review some samples from the data, enabling them to practice. After reaching a common understanding, the researchers are able to start to assign small portions of the data for rating (Syed & Nelson, 2015). In the current study, after a shared understanding of the evaluation instrument was reached among the researchers and the sample studies were discussed, the actual evaluations were carried out, and the researchers evaluated a certain number of studies individually. During the process, the researchers consulted each other when potential conflicts and contradictory statements were observed and solved the agreement issues. Upon completing the evaluation process successfully, the findings were reported based on the standard areas, and an overall score was assigned to each study.

Trustworthiness

Regardless of whether being insider or external, evaluators have a lot to ensure the credibility of their findings, mainly by conducting their studies openly and consistently with professional integrity (Stufflebeam, 1974). The current study involved multiple external evaluators who do not have any conflict of interest with the authors of the selected studies to ensure analyst triangulation so that the threats to credibility could be minimized. Furthermore, the use of data collection instruments was standardized through calibration sessions. Two researchers checked the clarity and appropriateness of the language in the instruments as language experts to avoid any confusion in terms of the constructs. In addition, to ensure transparency and transferability of the study, each step was clearly defined, and data collection and analysis procedures were made explicit by providing a thick description of the meta-evaluation and stating inclusion/exclusion criteria (Korstjens & Moser, 2018).

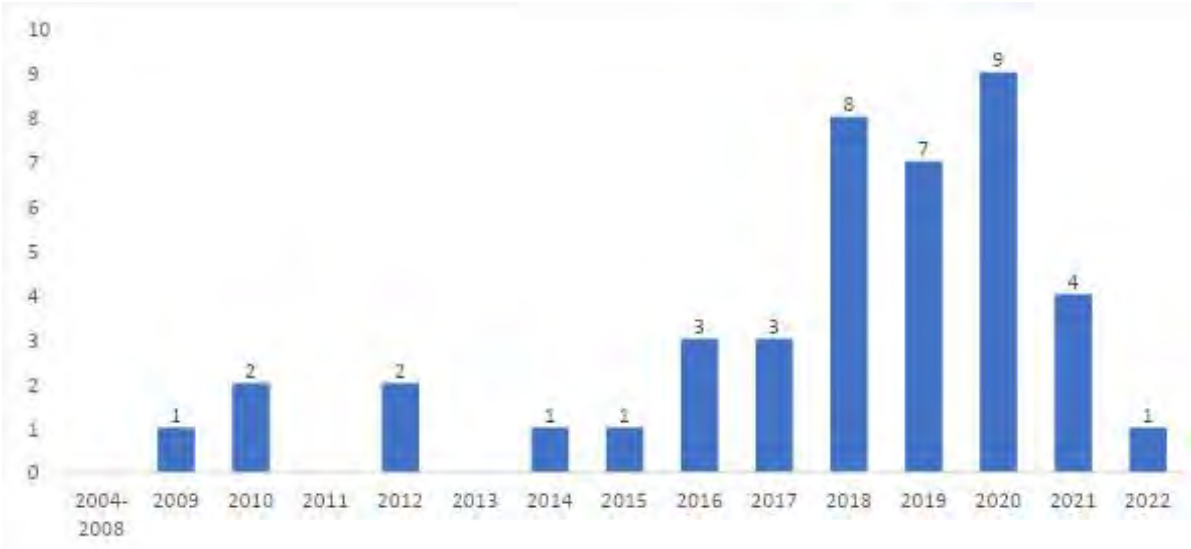
Findings

The General Trends in Curriculum Evaluation Studies

The year of publication of the studies is analyzed, and the related findings are presented in Figure 3 below. As seen in the bar graph, the earliest study was published in 2009, and the most recent study was published in 2022. Even if the search of the meta-evaluation study included research studies published since 2004, no article published earlier than 2009 was eligible regarding the selection criteria. While only 7 of the articles were published between 2004-2015, 6 of them were published in 2016 and 2017, with a marked increase of 29 of the articles published between 2018-2022*. There seems to be a sharp increase in 2018-2021 and a sharp decrease in 2021.

Figure 3

The Distribution of Studies Based on The Year of Publication



Note. *Databases were last accessed on 28.02.2022

This meta-evaluation is limited to K12 curriculum evaluation studies. However, the schooling levels in the evaluation studies provide valuable data in terms of trends. Most of the studies (n=24) focus on primary education (1st-8th grades). A considerable number of the studies (n=14) evaluate high school (9th-12th grades) curricula, but only a few studies focus (n=4) on the preschool level.

The distribution of the courses evaluated is presented in Figure 4, which shows a common trend in specific courses. Mathematics and English are the most commonly evaluated courses, whereas there is an even distribution in other courses, which are Class Guidance, Geography, Guidance, Health Education, Human Rights, Civics and Democracy, Information Technology and Software, Physics, Pre-School, Social Studies. Science courses in general and Biology, Chemistry, Physics, and Citizenship and Democracy Education courses are evaluated two or three times in the studies included in the meta-evaluation.

Table 2

The Distribution of the Courses Evaluated

Schooling Level	*Courses*	*f*
Preschool (36-72 months)	Preschool	3
Primary School (Grades 1-8)	Mathematics	1
	English	4
	Turkish	2
	Class Guidance	2
	Science	2
	Citizenship and Democracy Education	2
	Social Studies	1
	Information Technology	1
	Technology and Design	1
	Mathematics and Science	1
	Human Rights, Civics, and Democracy	1
Secondary School (Grades 9-12)	English	3
	Chemistry	2
	Biology	2
	Turkish Revolution History	
	Information Technology and Vocational Development Modular Private High School Curriculum	
	Mathematics	1
	Geography	1
	Health Education	1

The evaluation approach and the model were not reported in 19 of the evaluation research articles, while 23 were conducted based on an evaluation approach/model (See Table 2). Considering the most used models, the CIPP and Eisner's model were used in six of the studies; and the other models were used in two studies. Accordingly, the most used approaches were decision-oriented and expert-oriented, followed by objective-oriented and participant-oriented approaches. Moreover, most of the authors of the evaluated articles neither provided a clear rationale for utilizing a certain evaluation model or approach in their articles nor did they recommend alternative models to be adopted in the evaluation.

Table 3

Curriculum Evaluation Studies by Evaluation Approach/Model

*Evaluation Approach**	*Evaluation Model*	f
Objective Oriented	Tyler's Objective-Based Evaluation Model	2
	Provus's Discrepancy Model	2
Participant Oriented	Stake's Congruence-Contingency Model	2
	Stake's Responsive Evaluation Model	2
Expert Oriented	Eisner's Educational Connoisseurship and Criticism Model	6
Decision Oriented	Stufflebeam's Context, Input, Process, Product-CIPP Model	6
Multiple Models		2
Other/Not Reported		20**

Note. *The evaluation approaches were classified based on Fitzpatrick et al. (2004).
**One of the articles utilized an evaluation model developed by its researcher.

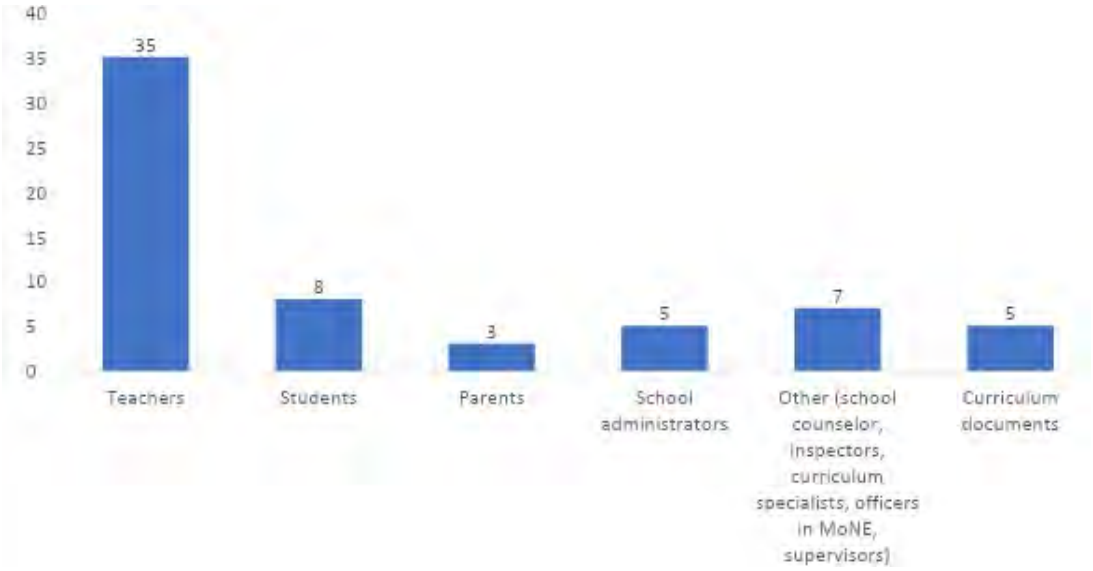
The number of curriculum evaluation research studies by research design is presented in Table 3. Among the studies, 20 of them used qualitative (e.g., case study, phenomenology, content analysis), 11 of them used quantitative (e.g., survey design, relational survey model), and lastly, 11 of the studies used mixed-method design (e.g., sequential explanatory design and convergent parallel mixed-method design).

Based on the document analysis, it is found that a variety of sources were used to collect data in relation to the evaluation of the curricula under investigation. While some curriculum evaluation research studies focused on analyzing the curriculum documents and course materials, others collected data from various individuals to evaluate the programs. The individuals used as data sources can be summarized as students, teachers, parents, school administrators, and others (school counselors, inspectors, curriculum specialists, officers in MoNE, and supervisors) (see Figure 4).

In terms of the type of data sources used, teachers in 83.33% of the studies (n = 35), students in 19.04% of the studies (n = 8), curriculum documents in 11.90% of the studies (n = 5), administrators in 11.90% of the studies (n = 5), parents in 7.14% of the studies (n = 3), and other (school counselor, inspectors, curriculum specialists, officers in MoNE, supervisors) in 16.66% of the studies (n = 7) were utilized as their data sources. It is seen that more than half of the studies included teachers (n = 22), curriculum documents (n = 4), or students and teachers (n = 3) as their only data sources, and the rest of the data sources were utilized by a limited number of studies (e.g., "students, teachers, and administrators" were used as data sources in 2 research studies). The curriculum evaluation studies are categorized into 5 groups based on the number of data sources they utilized. It is seen that 69.0% of the studies (n = 29) used one; 14.3% of the studies (n = 6) used two; 7.1% of the studies (n = 3) used three; 4.8% of the studies (n = 2) used four data sources; and 2.4% of the studies (n = 1) used five data sources.

Figure 4

Curriculum Evaluation Studies by Data Sources



The Categories of the Recommendations Deduced from K12 School Curricula Evaluation Studies

Data analysis revealed that out of 42 articles, 39 recommended practice and further research that would provide insights into curriculum development and evaluation. These recommendations were grouped under 13 categories: informing and including stakeholders, in-service training, pre-service teacher training, implementation, course hours, classroom arrangement/size, physical conditions, course materials, content and objectives, testing and evaluation, curriculum development and evaluation, policy, and methodology. Table 4 presents the categories and the number of articles falling into these categories. Besides, it shows each category's ratio (percentage) to all recommendations made in the selected articles. On the other hand, there were no recommendations in three of the articles.

Table 4

The Foci of Recommendations in the Curriculum Evaluation Research Studies

Foci of Recommendations	*f*	*%*
Informing and including stakeholders	12	28.60
Inservice training	28	66.70
Pre-service teacher training	5	10.90
Implementation	14	33.30
Course hours	4	9.50
Classroom arrangement/size	3	7.10
Physical conditions	5	11.90
Textbook/materials/resources	17	40.50
Course/curriculum content/objectives	17	40.50
Testing and evaluation	6	14.30
Curriculum development/evaluation	19	45.20
Policy	6	14.30
Methodology	20	47.60
No recommendation	3	7.10

Note. The frequencies represent the number of articles that made the related recommendations, and the percentages represent the ratio of each category to all recommendations made in the selected articles.

The content analysis revealed that in-service teacher training throughout different phases, such as curriculum development, evaluation, and implementation, was the most salient category of recommendation accentuated by the majority of the articles. To that end, Demir et al. (2017) underscore that "teachers should be guided by organizing in-service training activities on the basic philosophy and objectives of the curriculum, learning-teaching approach, measurement and evaluation approach, how to implement curriculum, and the issues to be considered in practice, in order for the curriculum to achieve its purpose by applying it effectively." (p. 177). Secondly, the articles underline the significance of the processes that cater to curriculum development and evaluation. İzci and Eroğlu (2018) suggest that reviewing the development, implementation, and evaluation processes through a scientific approach is of paramount importance by ensuring the stakeholders' active participation and paying attention to their opinions and conducting pilot studies upon which deficiencies are to be eliminated. Although there were recommendations regarding curriculum development, only two articles (Gökalp & Köksaldi, 2019; Sağlık & Aldan-Karademir, 2019) mention the necessity for conducting a needs assessment before a development or revision process. Besides, Aslan and Uygun (2019) state that the curriculum should be developed considering the needs of different regions/conditions and be practiced considering the importance of family involvement. More significantly, it is underlined that curricula should demonstrate a range of flexibility, pave the way for localization, and have the potential to complement one another.

In addition to curriculum development and evaluation practices, curriculum content, objectives, materials, and resources were other significant dimensions. Considering that textbooks and materials function as the implemented curriculum in the classroom, the studies suggest that they can be updated by considering songs, games, and different educational activities suitable for students' ages, interests, and levels. In addition, teachers can be encouraged to develop course materials suitable for the activities they will implement in the lesson (Kandemir & Tok, 2015). Finally, it is well known that implementation of a curriculum is as critical as development since, without implementation, the goals of the curriculum cannot be realized. Underlining this notion, Gömleksiz and Akyıldız (2012) suggest that teaching activities should be organized by taking the student to the center, different intelligence areas of the students should be taken into account, and methods and techniques such as cooperative learning, drama, and case studies should be included. In addition, the use of computer-aided materials that stimulate both visual and auditory senses to appeal to the different learning styles of the students is important.

The evaluated articles mostly provided recommendations on including all stakeholders, such as students, teachers, parents, administrators, and counselors, not only in the curriculum development but in evaluation processes, as well as informing them about the programs and their outcomes. Erdoğan and Gürol (2016) highlight that "...parents, teachers, administrative and school partnership needed to be supported. All these stakeholders share equally valued roles in education. Parents should be part of a school's learning community because school is not a closed or self-sufficient system." (p. 130). Regarding the recommendations that cater to research and particularly the methods utilized in the evaluation studies, Gelen and Alış (2018) suggest increasing the number and variety of data collection instruments used in curriculum evaluation studies, getting the opinions of various stakeholders, and evaluating all aspects of the program rather than only one or a few aspects.

On the other hand, there were only a few recommendations regarding pre-service teacher training, course hours, classroom arrangement and size, physical conditions, testing and evaluation, and policy. It was suggested that pre-service teacher training programs should be revised to provide more opportunities for practice and courses to enhance pre-service teachers' curriculum literacy and awareness of the changes in the curricula. Besides, some studies recommended increasing the course hours, decreasing the number of students in the class, and using student-centered approaches for classroom management and instruction. For effective implementation of the curricula, there are suggestions for improving the physical conditions of the school by establishing laboratories, providing necessary materials, revising the schools' physical infrastructure, and eliminating hardware and technological shortages. Finally, there are recommendations regarding the use of achievement tests to evaluate the effectiveness of the curriculum and the attainment of learning objectives, the use of formative assessment in addition to summative assessment, and the utilization of different assessment materials.

It is of paramount importance that curriculum evaluation studies aim to evaluate the effectiveness of educational programs and provide recommendations for improvement. However, it is not uncommon for the recommendations presented in these studies to be at odds with the findings generated from the data they generated. This can occur due to various reasons, such as limitations in the research design, insufficient analysis, or the presence of biases. To ensure that the recommendations are in accordance with the evidence, it is essential for researchers to critically examine their findings and ensure that their recommendations are grounded in the data. By doing so, the evaluation studies can provide reliable and valid insights that can improve programs.

Meta-evaluation of Evaluation Studies in terms of Program Evaluation Standards

Shortlisted articles were evaluated based on the checklist, including the utility, feasibility, propriety, and accuracy standards, respectively. Under the utility (U, 7 items) standard, there were nine items as indicators of the quality of the evaluation study with regard to answering the needs and stating the limitations of the study. Upon the evaluation of each article based on utility, all articles showed more than half of the indicators embodying the criteria. Of all 42 articles, 23 articles showed all nine indicators, 11 articles showed eight, four articles showed seven, two articles showed six, and only one article showed five out of nine indicators on utility. For items U2, U3.1, and U5.2, every article was appropriate to the standard stated. Accordingly, results showed that all studies were conducted by trustworthy evaluators (U2), the scope of the studies was compatible with the needs of the stakeholders (U3.1), and they all stated a clear purpose for the studies conducted (U5.2). Overall evaluation of the studies regarding the utility standard showed that evaluation articles comply with the boundaries highlighted by the matrix.

The second standard according to which the articles were evaluated was feasibility (F, 4 items). The standard included four items regarding the time spent on the study, resources employed, the procedures and the objectivity of the reports provided. Of all 42 articles, 23 of them showed all four, 15 of the articles showed three, and four of the articles showed two indicators ensuring feasibility. For the item F2.2, all the evaluated articles were appropriate to the standard stated, meaning that all the articles seemed to be objectively reported without any interference or misapplication by interest groups.

The following standard in the matrix was propriety (P, 7 items). Articles were evaluated to identify their appropriateness to the proper procedures in evaluation, such as reporting approvals, limitations, confidentiality precautions, strengths, and weaknesses of the evaluation. Overall examination of the results showed that most articles failed to comply with the necessary regulations of propriety. Of all 42 articles, only two of the articles showed all seven indicators, six of the articles showed six, 16 of the articles showed five, 14 of the articles showed four, three of the articles showed three, and one article showed two of the indicators of propriety. Most of the articles didn't comply with three of the items of propriety, namely P1.1, P1.2, and P3.2. Accordingly, the results showed that most articles did not report the necessary approvals (P1.1) and the limitations of the study (P3.2). Also, more than half of the articles evaluated lacked details about informing the stakeholders about the procedures and the results of the evaluation study (P1.2).

The final standard included in the checklist was accuracy (A, 9 items). The items focused on the evaluation of the correct application of procedures of an evaluation study. The results showing the articles' positions on accuracy were presented in this part. There were nine items under the accuracy standard. Only nine of the articles showed all the indicators of nine items on accuracy, while 21 of the articles showed eight. Meanwhile, eight articles showed seven; two articles showed six; another two articles showed five; and one article showed four indicators of accuracy. Item A11, which focused on reporting the findings objectively, was checked for all the articles. However, for item A4.2, evaluating the employment of various data sources of information, the results showed that most articles failed to include multiple data sources.

When it comes to overall results on the evaluation of the articles using the checklist, which included four standards and 29 items in total, only one article evaluated showed all 29 indicators from the checklist, and one article showed 14, which is the lowest. Most articles showed more than half of the indicators, but instead of a quantitative picture, a detailed analysis of the results shows that most articles fail to meet the crucial standards of an evaluation study.

Discussion and Conclusion

This study aimed to present an overview of the curriculum evaluation research studies by carrying out a meta-evaluation of the studies conducted in K12 schooling levels between 2004 and 2022 in Türkiye based on the Joint Committee's evaluation standards. The findings regarding the trends in curriculum evaluation studies revealed that most articles were published between 2018-2021. Since this study applied some inclusion-exclusion criteria regarding the quality of the method/preferred design ,including evaluation literature and has a holistic approach to the evaluated program, it was seen that a higher number of studies matched the quality criteria in recent years.

Most studies focused on primary education, followed by the studies dealing with secondary education, and only a few dwelled upon preschool education. Among the studies, Mathematics and English were the most commonly evaluated courses, whereas there was an even distribution in other courses. This may be a result of curriculum development studies held by MoNE being intensified at the primary school level (2005, 2017, and 2020 curriculum development studies). Tan-Şişman et al. (2019) suggested another reason as there are fewer

graduate programs for secondary school than primary school level for some fields such as mathematics and science. Moreover, in nearly half of the studies, the evaluation approach and the model were not reported, which also supports the related literature indicating limited usage of program evaluation approaches and models in the studies (e.g., Aslan & Sağlam, 2017; Gökmenoğlu, 2014; Kurt & Erdoğan, 2015; Tan-Şişman et al., 2019). This finding shows a need for evaluation studies to be conducted on a theoretical background regarding the curriculum evaluation approach and model.

Although a variety of sources are used to collect data in relation to the evaluation of a curriculum, it is found that data were often gathered from one or two particular data sources, such as teachers and students. Hence, only in a few of the studies, three or more data sources were utilized. This result was parallel with the previous results indicating that stakeholders other than students and teachers were less included in the evaluation studies (e.g., Aslan & Sağlam, 2017; Tan-Şişman et al., 2019). In a similar vein, the findings also echoed those of Özdemir (2009), Süer (2022), and Yapıcıoğlu et al. (2016), who underscored the limited involvement of stakeholders and noted that this weakness hinders reaching multiple perspectives since they solely deal with teachers and students. This may reduce the discussions to a narrower perspective and constrain seeing a fuller picture, which is a limitation as it was pointed out by Connelly (2013) that curriculum "is a complex system involving teachers, students, curricular content, social settings, and all manner of impinging matters ranging from the local to the international." (p. ix).

Moreover, the findings of this study revealed that most curriculum evaluation research studies made suggestions about practice and further research. Among these suggestions, some aspects were mentioned more commonly than others. On the other hand, few studies made suggestions about pre-service teacher training, course hours, classroom arrangement and size, physical conditions, testing and evaluation, and policy. All in all, the lack of suggestions in these domains may have resulted from diverse reasons such as limited scope, data limitations, lack of expertise, political constraints, and lack of communication. While these factors may have limited the ability of researchers to make suggestions about certain aspects of the program, it is important for program evaluation studies to consider a wide range of factors that have a profound impact on program effectiveness. Ultimately, program evaluation studies should aim to provide a comprehensive assessment of the program, including suggestions for areas of improvement, to ensure that the program meets its intended goals and objectives.

Secondly, content analysis revealed a need for more comprehensive and in-depth curriculum evaluation research studies suggested by the evaluated studies so that these studies' findings might be utilized in curriculum development and design processes, as stated by Yapıcıoğlu et al. (2016). At the same time, the evaluated articles proposed that more studies in this field are more than essential to increase the effectiveness of these studies at diverse levels, which is also pointed out by Kurt and Erdoğan (2015). In general, the studies did not provide recommendations for improving the curriculum evaluation process itself. The reasons for this could be multifaceted and require further investigation in future studies. Moreover, the findings yielded that whereas there should be an attempt to increase the number of evaluation studies, researchers should also diversify their methods, contexts, data collection instruments, and samples to reach a more holistic picture of curricula and contribute to the reliability and

validity of those studies (Kurt & Erdoğan, 2015; Süer, 2022). Furthermore, based on the aspects on which the evaluated studies made recommendations, our findings uncovered that needs assessment and pilot studies should function as an integral part of building unity and coherence in curriculum implementation and development (Yüksel, 2000). According to the findings, another aspect on which the recommendations were made was concerning the pre-service teacher education to increase teacher candidates' proficiency in various domains, such as teaching methodology, curriculum literacy, assessment, and addressing diverse needs of students. This finding was supported by Ünver (2021), who argued the importance of providing pre-service teachers with opportunities to design ,develop, implement, and evaluate programs through a holistic approach and to enhance their attitudes towards curriculum studies. Finally, as Akşan and Baki (2017) suggest the findings based on the recommendation made by the evaluation studies also unearthed an array of problems in the curriculum implementation due to crowded classrooms, insufficient class time, and lack of quality materials.

Results from the evaluation of each standard in the matrix suggest that there are various parameters influencing evaluation studies' overall achievement in meeting the standards. The matrix foci are overall in parallel with the original standards set by the Joint Committee (1981). When the utility standard scores are evaluated, it can be seen that most of the articles performed well, especially in items U2, U3.1, and U5.2. Item U2 focuses on the trustworthiness of the evaluators of the study. In the articles evaluated, researchers are respected academic staff employed in Turkish universities; therefore, the item was automatically checked for the evaluators. Regarding items U3.1 and U5.2, which focus on the needs and interests of the audience and the clear description of the purpose, it was not surprising to see all the articles performed well since they were all published articles with specific sections regarding their purpose and significance. Half of the articles got a full point on the feasibility standard, drawing a less successful picture compared to utility. These results seem to be parallel to the systematic reviews regarding the evaluation studies conducted in Türkiye (Akıncı & Köse, 2021; İpek, 2022). Only item F2.2 was checked for all articles because it focuses on the objectivity of the evaluator(s) and the interest groups. Now that the studies were conducted for academic and scientific purposes, and most stakeholders and audiences were not included in the study, the high achievement of the articles written by objective evaluators does not seem surprising. Propriety standard is the least successful standard for the articles evaluated, with only two articles getting a full score. Especially the items P1.1, P1.2, and P3.2, focusing on the approvals, limitations, and informing stakeholders, were not checked for most articles. Most authors did not mention the ethical approval procedures, limitations of the study, or the process of informing stakeholders. The reasons may include the lack of requirements of ethical approvals from the journals, the delay in the process of conducting research, underestimation of ethical considerations, and inadequate reporting. As also suggested by the results of the previous research studies in the field, the limitations and suggestions provided by the evaluation studies must be shared with the stakeholders (Akıncı & Köse, 2021; İpek, 2022; Stufflebeam, 2004). Most articles got almost full points from the accuracy standard, most probably because most items match the general standards of academic writing and publishing studies. However, for item A4.2, almost all articles got zero points. The item focuses on providing information from various sources, which most articles lack. The use of various data sources increases the quality of the research and helps evaluators to come up with thick descriptions (Fraenkel et al., 2012). Even though the overall success of the articles from the checklist seemed to be satisfying, the

checklist should not be seen as an achievement test where the average success is accepted as the standard. Instead, each item stands significant, referring to a crucial criterion. That is why, the results suggest that there are crucial standards to be worked on for curriculum evaluation studies in Türkiye.

In line with the findings, considering that programs for elective courses and especially pre-primary schooling level were considerably less addressed, it is crucial to approach the program as a whole and consider the need for diversifying program evaluation research studies. More importantly, this study showed that about half of the studies were not based on any program evaluation approach and/or a program evaluation model. However, having a theoretical approach and model in evaluation studies is seen as crucial to provide a systematic and evidence-based data production that addresses the main issues regarding the evaluated program, such as the needs, contexts, questions, instruments, stakeholders, and the position of the researcher (Fitzpatrick et al., 2004). There are several possible reasons why program evaluation studies in Türkiye may fail to include or suggest certain evaluation models to adopt. These reasons may include a lack of familiarity or training on certain models, limited scope of the evaluation, resource constraints, bias, and a lack of communication between researchers and stakeholders. It is important for program evaluation studies to consider a range of evaluation models and methods to ensure that the evaluation is comprehensive and effective. However, there may be constraints that prevent researchers from including or suggesting certain models, and researchers need to be transparent about their methods and explain why certain models were chosen and others were not. Overall, program evaluation studies should aim to provide a thorough and objective assessment of the program, using appropriate evaluation models and methods that align with its goals and objectives. This may be highlighted in program development and evaluation courses in university graduate programs. Moreover, the results from data sources used in studies revealed a need for diversifying the data sources and including varied stakeholders to see different perspectives. Hence, the evaluation studies need to be planned in a way to provide a more complete and deeper understanding of the evaluated program by considering its complexities.

The results from the evaluation of the articles using the standards revealed that there are certain points to be improved for the evaluation studies conducted in Türkiye. Firstly, all studies were conducted and published with academic concerns and by academic staff. It may be useful for the stakeholders and the participants to see a detailed picture of the program drawn by professionals, but this causes certain limitations, especially in terms of the use and the dissemination of the results of the evaluation studies. Because of the time and monetary constraints, academic staff have trouble including all or many of the stakeholders as information sources and also as partners/informants in the evaluation. This ends up in a one-sided look at the program and limits the use of the results by the stakeholders who were not included in the process. Thus, the researchers, funders, and decision-makers of the program should collaborate in large-scale evaluation studies (Stufflebeam, 2000). Besides, academic concerns may shadow the main purpose of curriculum evaluation, creating pseudo and shallow problems for the programs to be evaluated. If there was a collaborative environment, adequate funding, and communication among the stakeholders and the evaluators, the actual problems of the programs could be dealt with in more depth, and the results of the studies might have reflections on the authentic use of the program. Finally, the ethical approvals and limitations of an evaluation study should be transparently discussed to inform the researchers and the

practitioners who may be willing to conduct a further study or to use the results and implications from the evaluation study. This need is also emphasized in the standards of the Joint Committee (Stufflebeam, 1981). All in all, the evaluation of the articles provides significant implications in terms of collaboration, transparency, and incisiveness.

Finally, based on the findings of the current study, future curriculum evaluation studies need to scrutinize the under-explored dimensions, including pre-service teacher training, course hours, classroom arrangement and size, physical conditions, testing and evaluation, and teacher education policies to lead to more informed decisions as regards to curriculum development, implementation, and evaluation.

Limitations and related recommendations for further research

Finally, the curriculum evaluation articles were evaluated in the aspects of utility, feasibility, propriety, and accuracy through The Meta-Evaluation Checklist, which provides quantitative data. However, this checklist is limited to the subjective interpretation of evaluators. Therefore, for future studies, it is recommended to use the Checklist for curriculum evaluation to create a supportive tool for qualitative analysis instead of a grading rubric for evaluating the quality. Moreover, this study is limited to K12; further research would consider replicating this research design with a focus on higher education programs.

Author Contributions

The first author has made substantial contributions in framing the theoretical background, conducting the literature review, designing the research, developing the checklist, interpreting the data, presenting the findings related to the first question, and contributing to the discussions.

The second author has made substantial contributions in framing the theoretical background, conducting the literature review, designing the research, developing the checklist, interpreting the data, presenting the findings related to the third and fourth questions, and contributing to the discussions.

The third author has made substantial contributions in framing the theoretical background, conducting the literature review, designing the research, developing the checklist, interpreting the data, presenting the findings related to the second question, and contributing to the discussions.

The fourth author has made substantial contributions in framing the theoretical background, conducting the literature review, designing the research, developing the checklist, interpreting the data, presenting the findings related to the third and fourth questions, and contributing to the discussions.

The fifth author provided critical reviews of the manuscript drafts, checklist, and overall content and structure, bringing his expertise and offering valuable insight.

References

Akıncı, M., & Köse, E. (2021). Research trends of program evaluation studies conducted between 2010-2019 in Turkey. *Cukurova University Faculty of Education Journal, 50*(1), 77-120. https://dergipark.org.tr/tr/pub/cuefd/issue/59484/688142

Akşan, E., & Baki, A. (2017). Content analysis of curriculum-related studies in turkey between 2000 and 2014. *Educational Sciences: Theory and Practice, 17*(3), 877-904. https://doi:10.12738/estp.2017.3.0002

*Aksoy, E. (2020). Evaluation of the 2017 updated secondary school English curriculum of Turkey by means of theory-practice link. *Turkish Journal of Education, 9*(1), 1-21. https://doi.org/10.19128/turje.575392

*Altındağ, A., & Korkmaz, H. (2019). Ortaokul 5. sınıf matematik dersi öğretim programının Stake'in uygunluk-olasılık modeline göre değerlendirilmesi. *Türk Eğitim Bilimleri Dergisi, 17*(2), 463-501. https://dergipark.org.tr/tr/pub/tebd/issue/50950/648965

*Aslan, D. (2016). An evaluation of the private high school curriculum in Turkey. *Journal of Education and Practice, 7*(9), 205-214. https://www.iiste.org/Journals/index.php/JEP/article/view/29700/30493

*Aslan, M., & Altunova, N. (2019). Birinci sınıf Türkçe (ilk okuma yazma) öğretim programının değerlendirilmesi. *Adıyaman Üniversitesi Sosyal Bilimler Enstitüsü Dergisi,* (32), 31-75. https://doi.org/10.14520/adyusbd.490598

*Aslan, M., & Erden, R. Z. (2018). Beşinci sınıf fen bilimleri öğretim programının değerlendirilmesi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi, 12*(2), 508-537. https://doi.org/10.17522/balikesirnef.506464

Aslan, M., & Sağlam, M., (2017). Methodological investigation of the curriculum evaluation theses completed between the years 2006-2015 in Turkey. *Universal Journal of Educational Research, 5*(9), 1468 - 1478. https://doi.org/10.13189/ujer.2017.050904

*Aslan, M., & Uygun, N. (2019). Okul öncesi eğitim programının Stufflebeam'in bağlam, girdi, süreç ve ürün (BGSÜ) değerlendirme modeline göre değerlendirilmesi. *Eğitim ve Bilim, 44*(200), 229-251. https://doi.org/10.15390/EB.2019.7717

*Avci, N., Erikci, B., & Ok, A. (2021). The evaluation of the secondary education basic mathematics curriculum through Stake's Responsive Evaluation Model. *Journal of Qualitative Research in Education, 27*,1-25. https://doi.org/10.14689/enad.27.2

*Balıkçı, Ç., Tüysüz, C., Taşdere, A., & İnel-Ekici, D. (2021). 3. sınıf fen bilimleri dersi öğretim programının bağlam-girdi-süreç-ürün (CIPP) modeline dayalı öğretmen görüşlerine göre değerlendirilmesi. *Milli Eğitim Dergisi, 50*(229), 523-544. https://dergipark.org.tr/tr/pub/milliegitim/issue/60215/649677

*Basaran, M., Dursun, B., Gur Dortok, H. D., & Yilmaz, G. (2021). Evaluation of preschool education program according to CIPP model. *Pedagogical Research, 6*(2), 1-13. https://doi.org/10.29333/pr/9701

*Başaran, S. T., & Ulubey, Ö. (2018). 2013 Okul öncesi eğitim programının değerlendirilmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 51*(2), 1-38. https://doi.org/10.30964/auebfd.417643

*Batdı, V. (2017). Comparing the high school English curriculum in Turkey through multi-analysis. *Educational Sciences: Theory and Practice, 17*(4), 1255-1290. https://doi.org/10.12738/estp.2017.4.0490

*Berkant, H.G., & İncecik, A. (2018). Ortaokul matematik dersi beşinci sınıf öğretim programının öğretmenlerin görüşlerine göre değerlendirilmesi. *International Journal of Education Technology and Scientific Researches, 3*(6), 99-125. https://dergipark.org.tr/tr/pub/ijetsar/issue/40311/434240

Bulut, İ. (2006). *Yeni ilköğretim birinci kademe programlarının uygulamadaki etkililiğinin değerlendirilmesi (An evaluation of the effectiveness of the new primary school curricula in practice).* (Thesis No: 186204) [Doctoral Dissertation, Fırat University]. Turkish Council of Higher Education Theses Center. http://hdl.handle.net/11508/14563

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9(*2), 27-40. https://doi.org/10.3316/QRJ0902027

Connelly, F.M. (2013). Foreword. In: Deng, Z., Gopinathan, S., & Lee, C. K. E. (Eds.), Globalization and the Singapore curriculum: From policy to classroom (pp. vii-xii). Springer.

Cook, T. D., & Gruder, C. L. (1978). Metaevaluation research. *Evaluation Quarterly, 2*(1), 5-51. https://doi.org/10.1177/0193841X7800200101

*Çelik, K., & Büyükalan-Filiz, S. (2018). Ortaöğretim İngilizce dersi öğretim programının (2014) Eisner modeline göre değerlendirilmesi. *Eğitim ve Toplum Araştırmaları Dergisi, 5*(1), 50-67. https://dergipark.org.tr/tr/pub/etad/issue/37928/371740

*Çetin, E., & Gündoğdu, K. (2020). The evaluation of the 7th grade English curriculum according to the Eisner's educational connoisseurship and criticism model. *İlköğretim Online, 19*(3), 1242-1266. https://doi.org/10.17051/ilkonline.2020.728032

*Çiftçi-Cinkavuk, E., & Cesur, K. (2018). An evaluation of second grade English language teaching program of primary school: Tokat case. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi, 20*(3), 749-766. https://doi.org/10.17556/erziefd.437273

*Demir, E., Gacanoğlu, Ş., & Nakiboğlu, C. (2017). 2013 Kimya dersi öğretim programına yönelik öğretmen görüşleri doğrultusunda 2017 kimya dersi öğretim programının değerlendirilmesi. *Türkiye Kimya Derneği Dergisi Kısım C: Kimya Eğitimi, 2*(2), 135-184. https://dergipark.org.tr/tr/pub/jotcsc/issue/31858/345482

*Demiralp, N. (2017). Coğrafya öğretiminde programların tasarım ve program öğeleri açısından incelenmesi ve 2017 öğretim programı. *21. Yüzyılda Eğitim ve Toplum Eğitim Bilimleri ve Sosyal Araştırmalar Dergisi, 6*(17). https://dergipark.org.tr/tr/pub/egitimvetoplum/issue/35909/402872

Demirbaş, M., & Yağbasan, R. (2005). Türkiye'deki ortaöğretim kurumlarında uygulanan fen öğretim programlarının analizi: Modern fen öğretim program uygulamaları. *Gazi Üniversitesi Kırşehir Eğitim Fakültesi Dergisi, 6*(2), 33- 51. https://dergipark.org.tr/tr/pub/kefad/issue/59538/856361

*Demirel, M. (2010). İlköğretim ve ortaöğretim kurumları sınıf rehberlik programının değerlendirilmesi. *Eğitim ve Bilim, 35*(156), 45-60. http://egitimvebilim.ted.org.tr/index.php/EB/article/view/72

*Erdoğan, P., & Gürol, M. (2016). The evaluation of health education program (HEP) of 9th graders. *Journal of Education and Practice, 7*(33), 124-133. https://www.iiste.org/Journals/index.php/JEP/article/view/34138

Eğitim Reformu Girişimi (2005). *Öğretim programları inceleme ve değerlendirme* -1 [Education programs review and evaluation - 1]. http://apkm.short.gy/ERG-report

*Ersoy, M., & Merter, F. (2012). 9. Sınıf biyoloji dersi öğretim programının öğretmen görüşlerine göre değerlendirilmesi (Sivas İli Örneği). *E-International Journal of Educational Research, 3*(3), 1-17. http://www.e-ijer.com/en/pub/issue/8019/105317

Ertekin, B. E., & Bozkurt, Ü. (2020). Exploring findings of research on Turkish language curricula: a meta-synthesis study. *Journal of Language Education and Research, 6*(2), 613-634. https://doi.org/10.31464/jlere.767806

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternatives, approaches and practical guidelines* (3rd ed.). Pearson Education Inc.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw Hill.

*Geçitli, E., & Bumen, N. (2020). İlkokul ve ortaokulda bilişim teknolojileri alanında yer alan derslerin öğretim programları üzerine bir analiz: 1998-2018. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 20*(4), 1912-1934. https://doi.org/10.17240/aibuefd.2020.20.58249-627376

*Gelen, İ., & Alış, E. (2018). Ortaokul matematik ve fen bilimleri öğretim programının boyutlarının değerlendirilmesinde paydaşların görüşleri. *Disiplinlerarası Eğitim Araştırmaları Dergisi, 2*(4), 28-42. https://dergipark.org.tr/tr/pub/jier/issue/42063/482147

*Gökalp, M., & Köksaldi, G. (2019). Ortaokul 5. sınıf matematik programının öğretmenlerin görüşlerine göre değerlendirilmesi. *Elektronik Eğitim Bilimleri Dergisi, 8*(16), 218- 241. https://dergipark.org.tr/tr/pub/ejedus/issue/51353/645805

Gökmenoğlu, T. (2014). Geniş açı: Modeller ve yaklaşımlar açısından Türkiye'de program değerlendirme çalışmaları. *Uluslararası Eğitim Programları ve Öğretim Çalışmaları Dergisi, 4*(7), 55-70. https://www.ijocis.com/index.php/ijocis/article/view/14

*Gömleksiz, M. N., & Akyıldız, S. (2012). Vatandaşlık ve demokrasi eğitimi dersi öğretim programının uygulamadaki etkililiğinin değerlendirilmesi. *Milli Eğitim, 41*(196), 69-92. https://dergipark.org.tr/tr/pub/milliegitim/issue/36171/406676

Gözütok, F. D., Akgün, Ö. E., & Karacaoğlu, Ö. C. (2005, November 14- 16). İlköğretim programlarının öğretmen yeterlikleri açısından değerlendirilmesi (Paper presentation). Eğitimde Yansımalar VIII: Yeni İlköğretim Programlarının Değerlendirilmesi Sempozyumu. Kayseri, Türkiye.

*Gündüz, G. F., & Kuzu Demir, E. B. (2020). 2017 Bilişim teknolojileri ve yazılım dersi öğretim programının ortaokul 5. Sınıf öğrencilerinin görüşlerine göre değerlendirilmesi: Eskişehir ili örneği. *Trakya Eğitim Dergisi, 10*(3), 1024-1041. https://doi.org/10.24315/tred.735723

*Gürel, E., & Demirhan-İşcan, C. (2020). Reviewing the 9th grade English curriculum with Stake's responsive evaluation model according to teachers' opinions. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi, 49*(1), 501-554. https://dergipark.org.tr/en/pub/cuefd/issue/53758/623396

*Gürültü, E., & Gürol, M. (2020). Mesleki gelişim modüler öğretim programının katılımcı odaklı

yaklaşım ile değerlendirilmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 20*(1), 81-99. https://doi.org/10.17240/aibuefd.2020.20.52925-514361

Greene, J. C., Dumont, J., & Doughty, J. (1992). A formative audit of the ECAETC year 1 evaluation: Audit procedures, findings, and issues. *Evaluation & Program Planning, 15*, 81-90. https://doi.org/10.1016/0149-7189(92)90065-3

Harris, J. D., Quatman, C. E., Manring, M. M., Siston, R. A., & Flanigan, D. C. (2014). How to write a systematic review. *The American Journal of Sports Medicine, 42*(11), 2761-2768. https://doi.org/10.1177/0363546513497567

İpek, Ö. F. (2022). A systematic review of program evaluation studies in EFL: The Turkish case. *Dil ve Edebiyat Araştırmaları,* (25), 199-217. https://doi.org/10.30767/diledeara.1057707

*İzci, E., & Eroğlu, M. (2018). Yenilenen 9. sınıf kimya dersi öğretim programının öğretmen görüşlerine göre değerlendirilmesi. *E-Uluslararası Eğitim Araştırmaları Dergisi, 9*(1), 13-35. https://doi.org/10.19160/ijer.322892

Kablan, Z. (2011). İlköğretim matematik öğretim programının değerlendirilmesine yönelik araştırmaların analizi. *İlköğretim Online, 10*(3), 1160-1177. https://dergipark.org.tr/tr/pub/ilkonline/issue/8591/106802

*Kandemir, A., & Tok, Ş. (2017). İlkokul 2. sınıf İngilizce öğretim programinin katilimci odakli program değerlendirme yaklaşimiyla değerlendirilmesi. *Milli Eğitim Dergisi, 46*(215), 27-67. https://dergipark.org.tr/en/pub/milliegitim/issue/36134/405892

*Karakuş, H., & Akman, B. (2022). Okul öncesi matematik programının öğretmen ve ebeveyn görüşlerine göre değerlendirilmesi: Nitel bir çalışma. *Pamukkale Eğitim Fakültesi Dergisi*, (54), 297-327. https://doi.org/10.9779/pauefd.821103

*Karakuş Özdemirci, Ö., Aksoy, A. N., & Ok, A. (2018). Evaluation of human rights, civics and democracy curriculum through Eisner's evaluation framework. *Hacettepe University Journal of Education, 35*(1), 136-150. https://doi.org/10.16986/HUJE.2019051813

*Kayhan, E., & Gürol, M. (2019). Türkçe öğretim programı (2017)'nın Stufflebeam'in (CIPP) modeline göre değerlendirilmesi. *Türkiye Eğitim Dergisi, 4*(1), 48-67. https://dergipark.org.tr/tr/pub/turkegitimdergisi/issue/46280/548711

Kazu, H., & Aslan, S. (2012). 2004 İlköğretim programının "öğrenme-öğretme süreci" boyutu ile ilgili yapılan araştırmaların değerlendirilmesi. *E-Uluslararası Eğitim Araştırmaları Dergisi, 3* (2), 78-94. http://www.e-ijer.com/tr/pub/issue/8018/105333

*Kılınç, M. B., & Anılan, H. (2019). Birinci sınıf öğretmenlerinin birinci sınıf matematik dersi öğretim programına ilişkin görüşlerinin incelenmesi. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi,* (20), 1033-1061. https://doi.org/10.17494/ogusbd.555117

Koç, E. S. (2016). Türkiye de ilköğretim programlarının değerlendirilmesine yönelik yapılan lisansüstü tezlerin incelenmesi (2005–2014*). Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 16*(1). https://doi.org/10.17240/aibuefd.2016.16.1-5000182918

*Koçakoğlu, M. (2016). Ortaöğretim biyoloji dersi öğretim programının değerlendirilmesi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi, 10*(2), 65-91. https://doi.org/10.17522/balikesirnef.276943

Korkmaz, İ. (2020). Türkiye'de program geliştirme ve değerlendirme çalışmaları. In Oral, B. & Yazar, T. (Eds.), *Eğitimde program geliştirme ve değerlendirme* (pp.584-607). Pegem Akademi.

Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice, 24*(1), 120-124. https://doi.org/10.1080/13814788.2017.1375092

Kurt, A., & Erdoğan, M. (2015). Content analysis and trends of curriculum evaluation research: 2004-2013. *Education and Science, 40*(178), 199-224. https://doi.org/10.15390/EB.2015.4167

Kürüm-Yapıcıoğlu, D., Atik-Kara, D., & Sever, D. (2016). Trends and problems in curriculum evaluation studies in Turkey: The perspective of domain experts. *Uluslararası Eğitim Programları ve Öğretim Çalışmaları Dergisi, 6*(12), 91-114. https://www.ijocis.com/index.php/ijocis/article/view/159

Liberati, A, Altman D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Loannidis, J.P., Clarke, M., Devereaux P.J, Kleijnen J., Moher D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care intervention: Explanation and slaboration. *Annals of Internal Medicine, 151*(4), 65-94. https://doi.org/10.7326/0003-4819-151-4-200908180-00136

Ministry of National Education. (1995). MLO Müfredat Laboratuvar Okulları Modeli. Ankara: Milli Eğitim Basımevi.

Ministry of National Education. (2006). Öğretim programlarının değerlendirme raporu (İlköğretim 6. sınıflar Türkçe, Matematik, Fen ve Teknoloji, Sosyal Bilgiler). https://apkm.short.gy/earged

Ministry of National Education. (2012). 12 yıl zorunlu eğitim sorular-cevaplar. https://www.meb.gov.tr/duyurular/duyurular2012/12yil_soru_cevaplar.pdf

Ministry of National Education. (2017). *Müfredatta yenileme ve değişiklik çalışmalarımız üzerine*. https://uskudar.meb.gov.tr/meb_iys_dosyalar/2017_08/24010456_mYfredat_tanYtYm_kitapYY.pdf

Ministry of National Education. (2020). *Öğretim programlarını değerlendirme raporu*. https://ttkb.meb.gov.tr/meb_iys_dosyalar/2020_08/24113242_ogretimprogramlari_dr.pdf

*Nazlı, S., & Er, K. O. (2010). İlköğretim sınıf rehberliği programının değerlendirilmesi. *Balikesir University Journal of Social Sciences Institute, 12*(22), 129-147.

O'Leary, Z. (2014). *The essential guide to doing your research project* (2nd ed.). SAGE Publications, Inc.

Özdemir, S. M. (2009). Eğitimde program değerlendirme ve Türkiye'de eğitim programlarını değerlendirme çalışmalarının incelenmesi. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi, 6*(2), 126-149. https://dergipark.org.tr/tr/pub/yyuefd/issue/13712/166017

Sağlam, M., ve Yüksel, İ. (2007). Program değerlendirmede meta-analiz ve meta-değerlendirme yöntemleri. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, 18*, 175-187. https://dergipark.org.tr/tr/pub/dpusbe/issue/4760/65397

*Sağlık, M. A., & Aldan-Karademir, Ç. (2019). Teknoloji ve tasarım dersi 2018 öğretim programının öğretmen görüşlerine göre değerlendirilmesi. *Eğitimde Nitel Araştırmalar Dergisi, 7*(1), 302-319. https://doi.org/10.14689/issn.2148-2624.1.7c1s.14m

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

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Sage.

*Singer, E. N., & Taş, İ. D. (2020). 2017 3. Sınıf matematik öğretim programının CIPP modeline göre girdi boyutunda değerlendirilmesi. *Türkiye Sosyal Araştırmalar Dergisi, 24*(2), 469-484. https://dergipark.org.tr/tr/pub/tsadergisi/issue/56373/504177

Stufflebeam, D. L. (1974). *Meta-evaluation*. Western Michigan University Evaluation Center. Occasional Paper Series #3.

Stufflebeam, D. L. (1978). Meta evaluation: An overview. *Evaluation & the Health Professions, 1*(1), 17-43. https://doi.org/10.1177/016327877800100102

Stufflebeam, D. L. (1981). *Standards for evaluations of educational programs, projects, and materials*. McGraw-Hill Book Co.

Stufflebeam, D. L. (1999). *Program evaluations metaevaluation checklist*. https://usaidlearninglab.org/sites/default/files/resource/files/mod3_meta_evaluation_checklist_1.pdf

Stufflebeam, D. L. (2000). The CIPP model for evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models* (pp. 279-317). Springer. https://doi.org/10.1007/0-306-47559-6_16

Stufflebeam, D. L. (2001). The meta-evaluation imperative. *American Journal of Evaluations, 22*(2), 183–209. https://doi.org/10.1177/109821400102200204

Stufflebeam, D. L. (2004). A note on the purposes, development, and applicability of the Joint Committee Evaluation Standards. *American Journal of Evaluation, 25*(1), 99-102. https://doi.org/10.1177/109821400402500107

*Sunay, C., & Taşgın, A. (2020). Ortaöğretim TC İnkılap tarihi ve Atatürkçülük dersi öğretim programının eğitsel eleştiri modeli bağlamında öğretmen görüşlerine göre değerlendirilmesi. *Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi,* (41), 315-334. https://doi.org/10.33418/ataunikkefd.814962

Süer, S. (2022). Türkiye'de İngilizce dersi öğretim programlarını değerlendirmeye yönelik yapılan lisansüstü tezlerin sistematik analizi (2005-2021 yılları arası). *e-Kafkas Eğitim Araştırmaları Dergisi*, *9*, 528-544. https://doi.org/10.30.900/kafkasegt.963984

Syed, M., & Nelson, S. C. (2015). Guidelines for establishing reliability when coding narrative data. *Emerging Adulthood, 3*(6) ,375-387. https://doi.org/10.1177/2167696815587648

*Şahin, İ. (2009). Curriculum assessment: Constructivist primary mathematics curriculum in Turkey. *International Journal of Science and Mathematics Education,* 8, 51-72. https://doi.org/10.1007/s10763-009-9162-2

Tan-Şişman, G., Ödün-Başkıran, S., & Aktan, T. (2019). Fen ve matematik dersi öğretim

programları değerlendirme çalışmaları: 2005-2017 yıllarındaki lisansüstü tezler. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi, 39*(3), 1235-1262. https://doi.org/10.17152/gefad.568854

*Toraman, Ç., & Gözütok, D. F. (2014). Evaluation of 8th grade citizenship and democracy education curriculum. *Elementary Education Online, 13*(3), 954-979.

*Turan, R. (2020). 2017 Sosyal Bilgiler dersi öğretim programına ilişkin öğretmen görüşleri (Aksaray ili örneği). *Mavi Atlas, 8*(2), 256-277. https://doi.org/10.18795/gumusmaviatlas.744979

Turgut, M. F. (1983). Program Değerlendirme. *Cumhuriyet Döneminde Eğitim*. İstanbul: Milli Eğitim Basımevi.

Ünver, G. (2021). Program çalışmaları için öğretmen eğitimi. *Öğretmen Eğitimi ve Öğretim, 2*(2), 30-55. https://doi.org/10.55661/jnate.952088

Ünal, S., Coştu, B., & Karataş, F. Ö. (2004). Türkiye'de fen bilimleri alanındaki program geliştirme çalışmalarına genel bakış. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi, 24*(2), 183-202. https://dergipark.org.tr/tr/pub/gefad/issue/6759/90922

Yaşar, Ş. (1998). *Evaluation of educational programmes in Turkey*. AECT Annual Meeting, San Diego. (ERIC Document No. ED 419 846).

*Yazıcı, T., & Taşgın, A. (2021). Evaluation of a mathematics curriculum in accordance with the Eisner's educational connoisseurship and criticism model. *International Journal of Curriculum and Instruction, 13*(2), 1226-1240. https://ijci.globets.org/index.php/IJCI/article/view/473

Yüksel, D. S. (2000). Milli eğitim bakanlığındaki program geliştirme çalışmalarının değerlendirilmesi. *Kuram ve Uygulamada Eğitim Yönetimi, 24* (24), 501-608. https://dergipark.org.tr/tr/pub/kuey/issue/10372/126944

Yüksel, İ. (2010). *Türkiye için program değerlendirme standartları oluşturma çalışması (Development of Turkish program evaluation standards)*. (Thesis No: 262343) [Doctoral Dissertation, Anadolu University]. Turkish Council of Higher Education Theses Center.

Uluslararası Eğitim Programları ve Öğretim Çalışmaları Dergisi
13(*2*), 2023, 325-356

www.ijocis.com

TÜRKÇE GENİŞ ÖZET

Türkiye'de K-12 Eğitim Düzeyinde 2004-2022 Yılları Arasında Yürütülen Program Değerlendirme Araştırmalarının Sistematik Meta-Değerlendirmesi

Giriş

Program değerlendirme çalışmalarına ilişkin kararların geçerliliği ve güvenilirliği, değerlendirme sürecine bağlıdır (Sağlam & Yüksel, 2007). Sağlam ve Yüksel (2007) tarafından belirtildiği gibi, yanlış veya önyargılı değerlendirme sonuçları sunmak geri dönülmez sonuçlara yol açabileceğinden, değerlendirme kalitesi önemlidir. Stufflebeam (2000), değerlendirme raporlarının meta-değerlendirmeye tabi tutulmadığı durumlarda, karar vericilerin programın gelişimine katkı sağlamayacak kararlar almasına neden olabileceğini vurgulamaktadır. Meta-değerlendirme, Scriven (1969) tarafından ortaya atılan bir kavramdır ve değerlendirmenin değerini, karar verme sürecine rehberlik sağlama yeteneğini ve kaynak kullanımı açısından etik ve pratik olup olmadığını incelemeyi tanımlar (Stufflebeam, 1978).

Değerlendirme sürecine iliskin, teorik bir çerçeve oluşturulması ve değerlendirme standartlarının belirlenmesi konularında artan bir vurgu olduğu görülmektedir. Program değerlendirme, 1960'lardan bu yana gelişen ve olgunlaşan bir alan olarak ortaya çıkmıştır (Fitzpatrick vd., 2004). Türkiye'de de program değerlendirme çalışmalarına olan ilgi artmaktadır. Bu doğrultuda, program geliştirme ve değerlendirme çalışmaları, dünya genelinde program geliştirme alanının tanınmasıyla birlikte, 1950'lerden itibaren sistemli ve bilimsel bir şekilde yürütülmeye başlanmıştır (Özdemir, 2009).

Program değerlendirmeye verilen önem ve yapılan çalışmaların sayısı dikkate alındığında, Türkiye'deki çalışmalar hakkında yapılan sistematik bir meta-değerlendirme, bu konudaki genel eğilimler, prosedürler, güçlü ve zayıf yönler hakkında bütüncül bir anlayış sağlayabilir. Bunun, önceki değerlendirme çalışmalarının iyileştirilmesine rehberlik etme ve gelecekteki program geliştirme ve değerlendirme girişimlerine ışık tutma konusunda önemli olacağı düşünülmektedir. Bu çalışmanın amacı, Türkiye'de 2004-2022 yılları arasında K-12 eğitim düzeylerinde yapılan program değerlendirme araştırmalarını Eğitim Değerlendirme Ortak Komitesi tarafından önerilen dört standart (fayda, uygulanabilirlik, uygunluk ve doğruluk) temelinde meta-değerlendirme, önceki program değerlendirme araştırmalarının genel bir resmini sunmak ve değerlendiricilerin kaliteli ve sağlam değerlendirme konusundaki anlayışını geliştirmektir (Fitzpatrick et al., 2004; Stufflebeam, 2001; Stufflebeam, 2004).

Yöntem

Bu çalışmada, araştırmaya dâhil edilecek değerlendirme çalışmalarının belirlenmesi için sistematik analiz gerçekleştirilmiştir. Gerçekleştirilen analizde Dergipark, Web of Science, Scopus ve EBSCOhost veritabanlarında belirlenen anahtar kelimeler ile aramalar yapılmıştır. Bu aramaların ve içerik kontrol listesinin uygulanmasının ardından 42 program değerlendirme çalışması araştırmaya dâhil edilmiştir. Eğitim Değerlendirme Ortak Komitesi tarafından hazırlanan Meta-değerlendirme Kontrol Listesi araştırmacılar tarafından Türkiye bağlamına uyarlanmıştır. Bu süreçte bağlama uygun olmayan maddeler atılmış ya da değiştirilmiştir. Makalelerde yer alan önerilerin değerlendirilmesi amacıyla içerik değerlendirme yapılmıştır. Araştırmanın geçerlik ve güvenirliğini sağlamak amacıyla uzman görüşü alınmış, dört araştırmacı tarafından bağımsız değerlendirmeler ve uzman ve araştırmacıların katılmasıyla kalibrasyon toplantıları yürütülmüştür.

Bulgular

Yapılan araştırma sonucunda en erken program değerlendirme çalışmasının 2009 yılında, en güncel çalışmanın da 2022 yılında yapıldığı saptanmıştır. En çok çalışma 2018-2021 yılları arasında yürütülmüştür. Ders kapsamında en sık değerlendirilen dersler matematik ve İngilizce dersleriyken, okul seviyesi bazında en çok araştırma ilköğretim seviyesinde yürütülmüştür. Okul öncesi eğitim programlarına ilişkin araştırma sayısı azdır. Veri kaynakları değerlendirildiğinde, genellikle öğretmen ve öğrencilerden veri toplanmasıyla sınırlı kalındığı görülmüştür. Uygulanan meta-değerlendirme kontrol listesi değerlendirme çalışmalarını fayda, uygulanabilirlik, uygunluk ve doğruluk alanlarında değerlendirmiş ve sonucunda fayda ve uygulanabilirlik alanlarında çalışmalarının çoğu birçok kriteri karşılarken, uygunluk ve doğruluk alanlarında az sayıda çalışmanın bütün kriterleri karşıladığı saptanmıştır. En az karşılanan standartlar arasında etik izinler, şeffaf onay süreçleri ve kullanılan modellerin doğru ifade edilmesi yer almaktadır. Araştırmaların önerileri incelendiğinde paydaşlara ve değerlendirme çalışmalarının kapsamının genişletilmesine yönelik öneriler yapılmıştır.

Tartışma

Yürütülen araştırma sonucunda Türkiye'deki değerlendirme çalışmalarının genellikle ilköğretime odaklandığı saptanmış ve bunun olası nedenlerinden biri olarak Millî Eğitim Bakanlığı tarafından yapılan program değerlendirme çalışmalarının da aynı eğitim seviyesine odaklanması sunulmuştur. Bunun yanı sıra, veri kaynaklarının yalnızca öğrenci ve öğretmenler ile sınırlı olması, geniş bir kapsamda incelenmesi gereken çok katmanlı eğitim programları konseptinin anlaşılmasını zorlaştırmaktadır. Yürütülen değerlendirme çalışmalarının sonuçları, standartların karşılanmasında çeşitli parametreler olduğunu göstermektedir. Değerlendirme Kontrol Listesi, Ortak Komite (1981) tarafından belirlenen orijinal standartlarla genel olarak paralellik göstermektedir. Faydalılık standart puanları değerlendirildiğinde, özellikle U2, U3.1 ve U5.2 maddelerinde makalelerin çoğunun iyi performans gösterdiği görülmektedir. U2 maddesi, çalışmanın değerlendiricilerinin güvenilirliğiyle ilgilenmektedir. Değerlendirilen makalelerde, araştırmacılar Türkiye üniversitelerinde görev yapan akademik personeldir. Bu nedenle değerlendiriciler için madde otomatik olarak karşılanmıştır. Makalelerin yarısı uygunluk standardında tam puan alarak daha az başarılı bir görüntü çizmektedir. Bu sonuçlar, Türkiye'de

yapılan değerlendirme çalışmalarıyla ilgili sistematik incelemelerle paralellik göstermektedir. Değerlendirme çalışmalarının sınırlamaları ve önerileri paylaşması gerektiği daha önceki araştırma çalışmaları tarafından da önerilmiştir. Makalelerin çoğu doğruluk standardından neredeyse tam puan almıştır. Bunun nedeni, akademik yazma ve yayın standartlarıyla uyumlu olmalarıdır. Ancak, A4.2 maddesinden neredeyse tüm makaleler sıfır puan almıştır. Bu madde, makalelerin eksik olduğu çeşitli kaynaklardan bilgi sağlamaya odaklanmaktadır. Araştırma kalitesini artırmak için çeşitli veri kaynaklarının kullanılması önemlidir.

Genel olarak, Türkiye'de yapılan değerlendirme çalışmaları konusunda iyileştirilmesi gereken noktalar olduğu görülmektedir. Özellikle seçmeli dersler ve okul öncesi düzeyindeki programlar ele alınmamıştır. Bu nedenle, programı bütünsel bir şekilde ele almak önemlidir. Ayrıca, bu çalışma, çalışmaların yarısının herhangi bir program değerlendirme yaklaşımına veya modeline dayanmadığını göstermektedir. Ancak, değerlendirme çalışmalarında teorik bir yaklaşım ve modelin olması, değerlendirilen programla ilgili ana sorunları ele alan sistematik ve kanıta dayalı veri üretimi sağlamak açısından önemli görülmektedir. Türkiye'de program değerlendirme çalışmalarının belirli değerlendirme modellerini dâhil etmemesi veya önermemesine neden olabilecek birkaç olası sebep bulunmaktadır. Bunlar arasında belirli modellere ilişkin yetersiz bilgi, değerlendirmenin sınırlı kapsamı, kaynak kısıtlamaları, önyargılar ve araştırmacılar ile paydaşlar arasında iletişim eksikliği yer alabilir. Bununla birlikte, araştırmacıların belirli modelleri dâhil etmeme nedenlerini açıklıkla belirtmeleri ve yöntemlerini açıklamaları önemlidir. Program değerlendirme çalışmaları, hedefleri ve amaçlarıyla uyumlu değerlendirme modelleri ve yöntemlerini kullanarak programın kapsamlı ve tarafsız bir değerlendirmesini sağlamayı amaçlamalıdır. Bu, lisansüstü programlarında, program geliştirme ve değerlendirme derslerinde vurgulanabilir. Ayrıca, çalışmalarda kullanılan veri kaynaklarının çeşitlendirilmesi ve farklı paydaşların dahil edilmesi ihtiyacı olduğu ortaya çıkmıştır. Bu nedenle, değerlendirme çalışmaları, çok katmanlılığı göz önünde bulundurarak değerlendirilen programın daha eksiksiz ve derin bir değerlendirmesini sağlamak amacıyla planlanmalıdır. Sonuç olarak, standartlara göre yapılan makale değerlendirmesi, iş birliği, şeffaflık ve doğruluk açısından önemli sonuçlar ortaya koymaktadır. Ayrıca, tüm çalışmaların akademik kaygılarla yürütüldüğü ve akademik personel tarafından yayınlandığı belirtilmektedir. Bu nedenle, araştırmacılar, fon sağlayıcılar ve program karar vericileri büyük ölçekli değerlendirme çalışmalarında iş birliği yapmalıdır. Ayrıca, akademik kaygılar, program değerlendirmenin asıl amacını gölgeleyebilir ve programların gerçek sorunlarına yüzeysel bir bakış açısı getirebilir. Son olarak, bir değerlendirme çalışmasının etik onayları ve sınırlamaları tartışılmalıdır. Böylece bir çalışma yapmak veya değerlendirme çalışmasının sonuçlarını kullanmak isteyen araştırmacılar ve uygulayıcılar bilgilendirilmiş olur.

Appendix-I

The Metaevaluation Checklist

THE META-EVALUATION CHECKLIST FOR CURRICULUM EVALUATION RESEARCH ARTICLES

This checklist is constructed in order to assess the evaluation research articles based on the utility, feasibility, propriety, and accuracy standards set for curriculum evaluation of educational programs by the Joint Committee on the Evaluation Standards for Educational Evaluation. The checklist is composed of four sections, and each section includes items regarding the aforementioned standard areas. The evaluation research studies will be reviewed in the light of these items and will be evaluated based on each standard and overall quality.

PROGRAM DEĞERLENDİRME MAKALELERİ İÇİN META DEĞERLENDİRME KONTROL LİSTESİ

Bu kontrol listesi, Ortak Komite (Joint Committee on the Evaluation Standards for Educational Evaluation) tarafından eğitim programlarının değerlendirmesi için belirlenen yararlık, yürütülebilirlik, uygunluk ve doğruluk standartlarına dayalı program değerlendirme çalışmalarını değerlendirmek için oluşturulmuştur. Kontrol listesi, dört bölümden oluşmakta ve her bölümde yukarıda belirtilen standartlara ilişkin maddeler yer almaktadır. Program değerlendirme çalışmaları bu maddeler ışığında gözden geçirilecek ve çalışmaların niteliği, her bir standart özelinde ve bütün olarak değerlendirilecektir.

UTILITY/YARARLIK

1. U2 The evaluation study is conducted by competent and trustworthy evaluators.

(Değerlendirme, yetkin ve güvenilir değerlendiriciler tarafından yürütülmüştür.)

2. U3.1 The scope of the evaluation study is compatible with the needs and interests of the audiences (researchers, practitioners, experts, policymakers).

[Değerlendirmenin kapsamı kitlenin (araştırmacılar, uygulayıcılar, uzmanlar, karar alıcılar) ilgi ve ihtiyaçları ile uyumludur.]

3. U3.2 The information collected in the evaluation study is in line with the evaluation questions.

(Değerlendirme çalışmasında toplanan veri, değerlendirme soruları ile uyumludur.)

4. U5.1 The context of the evaluation is clearly described in the evaluation study.

(Değerlendirme çalışmasında, değerlendirmenin bağlamı detaylı bir şekilde tanımlanmaktadır.)

5. U5.2 The purposes of the evaluation are clearly described in the evaluation study.

(Değerlendirme çalışmasında değerlendirmenin amaçları detaylı bir biçimde tanımlanmaktadır.

6. U5.3 The procedures of the evaluation are clearly described in the evaluation study.

(Değerlendirme çalışmasında, değerlendirme yöntemi/işlemleri açık bir biçimde yazılmıştır.)

7. U5.4 The findings of the evaluation are clearly described in the evaluation study.

(Değerlendirme çalışmasında, değerlendirme bulgularına açık bir biçimde yer verilmektedir.)

8. U4 Relevant sources of values (perspectives/ procedures/ rationale) adopted for interpreting the findings are clearly described in the evaluation study.

(Değerlendirme çalışmasında, sonuçları yorumlamak amacıyla kullanılan perspektif, yöntem veya rasyonel gibi değerlerin ilgili kaynakları detaylı bir biçimde açıklanmaktadır.)

9. U8 The recommendations are provided in the evaluation study in ways that lead to further action by the stakeholders/researchers.

(Değerlendirme çalışması, paydaşlar ve araştırmacıları harekete geçmeye yönlendirecek öneriler ortaya koymuştur.)

FEASIBILITY/YÜRÜTÜLEBİLİRLİK

1. F1 The procedures utilized in the evaluation study are reported to be practical to minimize interruption in gathering information. (e.g., researchers, experts, financial resources, etc.)

Değerlendirme çalışmasında, yapılan işlemlerin veri toplama sürecini en az kesintiye uğratacak biçimde işe koşulduğu belirtilmiştir. Ör. Araştırmacılar, uzmanlar, ekonomik kaynaklar)

2. F3.1 The (human) resources used in the evaluation are described in detail in the evaluation study for accountability purposes.

[Değerlendirme çalışmasında, hesap verebilirliğin sağlanması amacıyla, değerlendirme sürecinde kullanılan kaynakları (insan) detaylı bir şekilde tanımlanmıştır]

3. F3.2 The time spent on the evaluation is described in detail in the evaluation study for accountability purposes.

(Değerlendirme çalışmasında, hesap verebilirliğin sağlanması amacıyla, değerlendirmeye ayrılan süre detaylı bir şekilde tanımlanmıştır.)

4. F2.2 The evaluation study is reported objectively so that any interference and misapplication by different interest groups (researchers, practitioners, experts, policymakers) are prevented.

[Değerlendirme çalışması, farklı ilgi gruplarının (araştırmacılar, uygulamacılar, uzmanlar, karar vericiler) müdahale ve yanlış uygulamasından kaçınılması amacıyla tarafsız bir biçimde rapor edilmiştir]

PROPRIETY/UYGUNLUK

1. P1.1 (If needed*) Necessary approvals and permissions for conducting the evaluation study are reported to be received.

[Değerlendirme çalışmasının yapılması için (gerektiğinde) gerekli onay ve izinler alındığı belirtilmiştir]

2. P1.2 Stakeholders involved in the study are informed about the details (e.g., purpose, method, timelines, and costs) of the evaluation study.

[Paydaşlar değerlendirme çalışmasının detayları (ör. Amaç, yöntem, zaman çizelgeleri ve bütçe) ile ilgili bilgilendirilmiştir]

3. P4 The findings of the evaluation study are reported within the limits of confidentiality and privacy of the stakeholders.

(Değerlendirme çalışmasının bulguları paydaşların gizlilik ve mahremiyet hakları çerçevesinde rapor edilmiştir.)

4. P7.1 The strengths of the program are reported in the evaluation study.

(Programının güçlü yönleri, değerlendirme çalışmasında rapor edilmiştir.)

5. P7.2 The weaknesses of the program are reported in the evaluation study.

(Programının zayıf yönleri, değerlendirme çalışmasında rapor edilmiştir.)

6. P3.1 The findings of the evaluation study are reported in an open, direct, and complete way.

(Değerlendirme çalışmasının bulguları açık, doğrudan ve bütünlük içinde aktarılmıştır.)

7. P3.2 The limitations of the evaluation study are reported in an open, direct, and complete way.

(Değerlendirme çalışmasının sınırlılıkları açık, doğrudan ve bütünlük içinde aktarılmıştır.)

ACCURACY/DOĞRULUK

1. A1 The program is adequately evaluated to clearly identify the design of the program.

(Değerlendirme, program tasarımını tanımlayacak şekilde uygulanmıştır.)

2. A2 The context in which the program is evaluated is described in detail to identify its effects on the program.

(Programın değerlendirme çalışmasının yapıldığı bağlam, programın üzerindeki etkilerinin saptanabilmesi için detaylı bir şekilde tanımlanmıştır.)

3. A4.1 The sources of information (institutions, individuals, research, documents etc.) in the evaluation study are clearly described so that adequacy of information can be assessed.

[Değerlendirme çalışmasında, elde edilen bilgilerin yeterliliğinin değerlendirilebilmesi için bilgi/veri kaynaklarına (kurumlar, kişiler, araştırmalar, dokümanlar vb.) açık bir şekilde yer verilmiştir]

4. A4.2 Various sources of information (institutions, individuals, research, etc.) in the evaluation study are utilized so that adequacy of information can be assessed.

[Değerlendirme çalışmasında, elde edilen bilgilerin yeterliliğinin değerlendirilebilmesi için çeşitli bilgi kaynakları (kurumlar, kişiler, araştırmalar, vb.) kullanılmıştır]

5. A5.1 The instruments used in the evaluation study for data collection are selected and/or developed appropriately to ensure validity.

(Değerlendirme çalışmasında kullanılan veri toplama araçları, geçerliliği sağlamak için uygun bir şekilde seçilmiş ve/veya geliştirilmiştir.)

6. A6 The instruments used in the evaluation study for data collection are implemented appropriately to ensure reliability.

(Değerlendirme çalışmasında kullanılan veri toplama araçları, güvenirliği sağlamak için uygun biçimde uygulanmıştır.)

7. A8-9 The analysis of the qualitative and/or quantitative data is conducted for supportable interpretations.

(Nitel ve/veya nicel verilerin analizi, desteklenebilir yorumlamalar yapılabilmesini sağlayacak şekilde yapılmıştır.)

8. A10 The conclusions of the evaluation study are justifiably reported for objective assessment by the audiences involved in or affected by the evaluation.

(Değerlendirme çalışmasının sonuçları, değerlendirmeye dâhil olan ya da değerlendirmeden etkilenen kitlenin nesnel değerlendirme yapabilmesi için savunulabilir bir şekilde ortaya konmaktadır.)

9. A11 The findings of the evaluation study are presented objectively without biased positions of the stakeholders and researchers.

(Değerlendirme çalışmasının bulguları, paydaşların ve araştırmacıların ön yargılı görüşleri olmadan nesnel bir şekilde sunulmuştur.)