Program Evaluation Experts' Competencies: A Delphi Study

Demet Sever[1]
[1]Anadolu University

To cite this article:

Program Evaluation Experts' Competencies: A Delphi Study

Demet Sever[1*]

[1]Anadolu University

Abstract

Due to the diversity of steps to be followed in program evaluation and knowledge, skills, qualification and competency requirements, the evaluation should be a field of expertise similar to other disciplines. Thus, it is important to consider program evaluation as a field of expertise and to determine the required competencies. This study aimed to determine the competencies of program evaluation experts based on the views of curriculum and instruction experts. A Delphi study was carried out with 23 participants in the first round and each was selected via the purposive sampling method. In the first round of the Delphi study, the open-ended questions were sent to the participants and qualitative data were analyzed with content analysis. At the end of the first round, a 5-point Likert-type questionnaire with 82 items was developed. In the second round, participants were asked to answer questions and also the views of the participants on main and sub-themes and their comments on these concepts were analyzed. In the third round, participants were asked to reevaluate their responses. The study was finalized with an 82 itemed-questionnaire under four main themes: "theoretical professional competencies of program evaluation experts", "practical professional competencies of program evaluation experts", "professional values that program evaluation experts should have", "professional skills that program evaluation experts should have".

Key words: Program evaluation, Program evaluation expert, Curriculum and instruction, Delphi technique

Introduction

In every part of life, the evaluation of the functionality, effectiveness and quality of the products and services based on different perspectives is of great importance to achieve better and current outcomes. Scriven (1996, p.395) stated that although evaluation has been an ancient practice, it is a new professional discipline. It could be suggested that evaluation is a human instinct based on individual taste. However, it is necessary to distinguish the evaluations made in daily life and those conducted with formal disciplines. In other words, professional evaluations should not entail subjectivity.

According to Ornstein and Hunkins (2009), evaluation is a decision-making process on a topic in the broadest sense. For this process to function properly, it is necessary to determine certain standards, to collect and analyze data with adequate methods and techniques, and to implement certain standards in the decision-making process. Program evaluation has been defined in various forms in numerous studies. According to Oliva (1988), program evaluation is a data collection process that requires scientific process skills. Barnes (1982, p.177), who considered program evaluation as a process, argued that it was necessary to seek answers to the questions regarding what will be evaluated, when and where it will be evaluated, why and by whom it will be evaluated in this process. Melrose (1998) also discussed program evaluation as a process where the value or usefulness of a program or its suitability for the individual, group, institution or community is determined. According to Melrose (1998), this analysis process varies based on the collected evidence, the questions asked to relevant individuals, and the value judgments of the practitioners. Özçelik (1992, p. 231) described program evaluation as the process of assigning a meaning to available information, the interpretation of the data based on its suitability for certain objectives, certain conditions, and certain meanings, etc.

Due to the diversity of steps to be followed in program evaluation and knowledge, skills, qualification and competency requirements, the evaluation should be a field of expertise similar to other disciplines. In addition, competition in a rapidly expanding global market requires service and product providers to provide evidence for

---
[*] Corresponding Author: *Demet Sever, dpala@anadolu.edu.tr*

individual or institutional accreditation (Engle, Altschuld, & Kim, 2006). Thus, evaluation experts have a great role to play in the process.

Undergraduate program evaluation departments are available in the USA, Canada, Australia, Iceland, Belgium and the West Indies (LaVelle & Donaldson, 2015). Although program evaluation education is not provided in a particular department in Turkey, program evaluation experts are trained in master's and doctorate programs in Curriculum and Instruction (CI) departments in education faculties in Turkish universities. The discipline of CI was initially introduced in the Educational Sciences Professorial Chair established within Ankara University, Faculty of Education in 1965. The undergraduate CI program, which provided undergraduate and graduate education services until 1997, was decommissioned during the restructuring of education faculties. The CI Program was responsible for the instruction of undergraduate vocational courses, aimed to improve the knowledge, skills and understanding of the students in the field of educational sciences at graduate level, and provide students with expertise in the field of CI through courses that emphasized the design, development, evaluation of programs, instruction design and scientific research skills. Lavelle and Donaldson (2015) reported that evaluators could be trained with various methods, including university programs, professional development workshops, webinars and field studies. Furthermore, Cousins, Goh, Elliott, and Bourgeois (2014) stated that program evaluation experts could be trained within the scope of formal education and special applications such as workshops, seminars, short courses and online training programs organized by private education institutions. Thus, since the evaluation experts are trained at graduate level in Turkey, this is not open to criticism. However, the study conducted by Atik Kara, Kürüm Yapıcıoğlu and Sever (2020) that evaluated CI graduate programs led to the question whether the education allowed the program evaluation experts to acquire the desired theoretical and applied knowledge since program evaluation topics were not adequately included in both master's and doctorate programs.

The consideration of program evaluation, which has a practical history of 50 years, as a field of expertise dates back to the 1980s (King, 2003; Stevahn, King, Ghere, & Minnema, 2005a). Evaluation, which was seen as a feasible job by a person who has an education in any field, was not only considered as an area of expertise, but was also considered important enough to suggest that the competencies that experts should have in this field should also be determined. Stevahn, King, Ghere, and Minnema (2005a) discussed the necessity of determining the competencies that program evaluators should possess and stated that since there was no undergraduate program for evaluation experts, anyone could claim to be a program evaluation expert. On the other hand, those who need a program evaluator may experience difficulties in determining the actual experts. However, individuals who want to be trained as program evaluators may also experience problems in selecting the adequate training program. Universities that aim to train program evaluation experts, on the other hand, could be limited by programs based on personal needs and priorities, when there is no specific qualification framework. Furthermore, the field may lack theoretical and descriptive research to guide effective applications. McGuire and Zorzi (2005) reported that it was also important to determine the field competencies to ensure the professional development of program evaluation experts who have diverse educational backgrounds and experiences.

Evaluation competencies include the skills, knowledge, abilities and qualifications required to evaluate (McGuire & Zorzi, 2005, p.74). It could be suggested that it is very difficult to define a fully competent framework for evaluators, since those who work as program evaluators have different educational backgrounds and experiences. On the other hand, certain studies also argued that the knowledge and skills required for evaluation may vary based on the purpose of the evaluation, employed evaluation model, the context, and the desires of the client or other stakeholders (Cooksy & Mark, 2012; McGuire & Zorzi, 2005; Volkov, 2011). Nevertheless, field experts have agreed that the competencies required for program evaluation, considered as a specialty, should be determined.

Studies on the competencies in program evaluation are available in the literature (Covert, 1992; Mertens, 1994; Scriven, 1996a). Scriven (1996a) summarized the required knowledge and skills for an evaluator in 10 items:

1. Basic qualitative and quantitative methodologies
2. Validity theory, generalizability theory, and meta-analysis
3. Legal constraints on data control and access, funds use, and treatment of the personnel (including the rights of human subjects)
4. Personnel evaluation
5. Ethical analysis
6. Needs assessment
7. Cost analysis

    8. Internal synthesis models and skills
    9. Conceptual geography
    10. Evaluation-specific report design, construction, and presentation (p.154)

It could be suggested that the evaluator competencies determined by the Joint Committee on Standards for Educational Evaluation (JCSEE) in 1994 and the American Evaluation Association (AEA) in 2004 are considered as fundamental criteria by the field experts. However, arguing that the predetermined principles and standards could not clearly indicate the features that evaluators would require, Stevahn, King, Ghere and Minnema developed a classification for basic evaluator competencies in 2001 in their study that aimed to determine the program evaluator competencies (King, Stevahn, Ghere, & Minnema, 2001). They updated their classification in 2005 with a few additions in 6 main categories and called the classification "basic competences for program evaluators." The 6 main categories and their scopes were as follows (Stevahn, King, Ghere, & Minnema, 2005):

> 1. Professional practice - competencies that focus on foundational norms and values that underlie program evaluation, such as adhering to standards and ethical principles.
> 2. Systematic inquiry - competencies that focus on the more technical aspects of program evaluation, such as study design, data collection and analysis, interpretation and reporting the findings.
> 3. Situational analysis - competencies that focus on analyzing and attending to unique interests, issues, and contextual circumstances in any given program evaluation.
> 4. Project management - competencies that focus on the nuts and bolts of conducting program evaluations, such as budgeting, coordinating resources, and supervising procedures.
> 5. Reflective practice - competencies that focus on one's awareness of evaluation expertise and needs for growth, including knowing oneself as an evaluator, assessing one's needs, and engaging in professional development for better evaluation practices.
> 6. Interpersonal competence - competencies that focus on the people skills used in conducting program evaluations, such as communication, negotiation, conflict, collaboration, and cross-cultural skills (p.106).

The classification of the competencies based on the dimensions of knowledge, skills and trends specific to the field and evaluation could lead to overlapping categories (King & Stevahn, 2015), and it is also broadly similar to previously determined principles and standards (Stevahn, King, Ghere, & Minnema, 2005).

Program evaluation is considered a field of expertise in several countries and various and specific training programs are offered by both formal and non-formal educational institutions. Furthermore, both institutional and individual studies have been conducted to determine program evaluation competencies for the effective practices in the field of expertise, to develop a foundation for professional development programs, and formal and non-formal curricula. In Turkey the program has been considered a sub-branch under evaluation expertise CI field, field-specific education is not available, and the curricula rarely include CI program outputs, courses, course objectives and content (Atik Kara, Kürüm Yapıcıoğlu, & Sever, 2020). It should not be forgotten that program evaluation is also an important stage of curriculum development process independent of the field. The program effectiveness would be ensured by the outputs of program evaluations conducted by competent evaluation experts. Thus, it is important to consider program evaluation as a field of expertise and to determine required competencies. The present study aimed to determine the competencies of program evaluation experts based on the views of CI experts. It could be suggested that the study findings would contribute significantly to both theoretical and practical developments in the program evaluation field. It was anticipated that the determined competencies would shed light on the renewal of the objectives and content of program evaluation courses, which are currently instructed in CI graduate programs, the self-assessment of professional competencies by program evaluation experts, and the structure of professional development programs. Furthermore, the distant aim of the present study was the acceptance of program evaluation expertise as a field of expertise in Turkey.

Method

Delphi technique was employed in the current study that aimed to determine the competencies of program evaluation experts based on CI expert opinion. Thus, the question "What are the professional and personal traits that program evaluators should possess?" was posed to field experts. Content analysis was employed to analyze

the responses. The list of items developed in the content analysis was presented to Delphi panelists as a Likert-type scale. Descriptive statistics was utilized to analyze participant responses.

Delphi Technique

Delphi technique, first employed in the military by Dalkey and Helmer (1963, p.458) in the 1950s, is a procedure that aims to obtain the most reliable agreement based on the views of a group of experts by subjecting them to a series of in-depth questionnaires interspersed with a controlled opinion feedback (Dalkey & Halmer, 1963, p.7).

There is a controversy about the research paradigm in which the Delphi technique should be included. Technique could be utilized with qualitative, quantitative and mixed research methods (Skulmoski, Hartman, & Krahn 2007, p.9). Although Delphi technique is considered among quantitative research techniques due to its focus on statistical agreement, collection of the research data based on subjective judgments and views of the experts (Fletcher & Childon, 2014, p.2; Habibi, Sarafrazi, & Izadyars, 2014, p.10). On the other hand, since it requires both quantitative and qualitative research skills in data analysis, it has also been considered as mixed method research technique (Creswell, 2008; Skulmoski, Hartman & Krahn, 2007).

Determination of the participants

The qualifications of the experts who participate in Delphi studies maximize the response quality, reduce stereotypes, and increase the reliability of the study (Powell, 2003, p.378; Okoli & Pawlowski, 2004, p.20; Nworie, 2011, p.25). Thus, it would be accurate to argue that the success of Delphi studies largely depends on the selection of experts. Şahin (2009, p.129) emphasized that the most important variable that improves the reliability of Delphi studies was the selection of adequate experts.

Purposive sampling is often used in the assignment of expert samples in Delphi studies. The assignment of experts or participants could be conducted based on the criteria associated with the research problem (Hasson, et al., 2000, p.1010). However, expertise is the most basic requirement in the assignment of the participants. Another criterion is the capacity and willingness of the assigned experts to contribute to the investigation of a particular problem. Participant willingness to contribute to the study is of great importance in the achievement of the study aim (Powell, 2003, p.379).

Previous studies argued different views on the expert sample size Delphi studies. Rowe and Wright (2001, p.128) stated that Delphi groups could include 5-20 experts. Clayton (1997, p.378) argued that 15-30 individuals would be adequate to represent homogeneous populations and 5-10 individuals could represent heterogeneous populations. According to Şahin (2001, p.217), the group size could be 100 or higher, but the most ideal Delphi group would include 10-20 experts. In conclusion, certain authors claimed that there should be at least 10 participants in certain Delphi studies, there could be more than 100 participants in certain others (Skulmoski, et al., 2007, p. 17-20). The sample size in a Delphi application may vary based on the topic, aim, and the scope of the research, the number of individuals that the author could reach, and whether the study was national or international (Clayton, 1997, p.378).

In the present study, criterion sampling, a purposive sampling method, which allows the assignment of individuals, events, objects or cares with certain qualifications that were predetermined about the problem, was employed to determine the expert participants (Büyüköztürk et al., 2010, p.91). The inclusion criteria were a "PhD degree in Curriculum and Instruction" and "teaching experience". Based on the above-mentioned criteria, the academic resumes of candidate participants were obtained from institutional websites. The review of the resumes led to a list of 50 expert individuals who met the study criteria. Faculty members were contacted via e-mail and invited to participate in the study. A message that included information on the aim, stages, and the response deadline of study was sent to 23 participants, who accepted to take part in the application among 50 experts. The responses to the questionnaires sent in each round and the sample size for each round are presented in Table 1. Twenty-three individuals participated in the first round of the study, 15 individuals participated in the second round and 12 participated in the third round. Participants who did not respond in any round of the study were excluded in the next round.

Table 1. The number of participants and responses in Delphi rounds

	1st round	2nd round	3rd round
Number of invitations	23	15	12

Number of responses	15	12	12
Response rate	%83.3	%80	%100

12 final round participants' demographics from 11 different university reported in Table 2.

Table 2. Participant demographics

		n
Gender	Female	7
	Male	5
Academic Title	Professor	3
	Associate Professor	5
	Assistant Professor	4

Data Collection

The study was conducted with three rounds of Delphi technique. The study conducted in each round is presented in a different section.

First Round

In the first part of the study, a literature review was conducted to determine the theoretical framework of the research. Then, various criteria were determined for the participants that would take part in the Delphi application and a list of experts was made based on these criteria. Invitation e-mails were sent to the experts. Correspondence between the author and the experts was conducted via e-mail. Detailed information on the research process was provided to the experts who agreed to participate in the study. In first phase of the Delphi application, the open-ended questions "What professional competencies should a program evaluator have?" and "What are the personal traits that a program evaluator should have?" were sent to the participants.

Qualitative data collected in the first round were analyzed with content analysis (Creswell & Clark, 2007). In short, content analysis involves the organization of the data to develop a general framework and entails definition, interpretation and classification of all data through reading and taking notes (Bauer, 2000). The main and sub-themes about the program evaluation expert competencies determined based on the agreement as a result of the comparison of the analysis findings with those determined by an another qualitative research and education curricula expert are presented in Table 3.

Table 3. The main and sub-themes

1. Theoretical Professional Competencies of Program Evaluation Experts	2. Practical Professional Competencies of Program Evaluation Experts
1.1. Program Development Theoretical Professional Competencies of Program Evaluation Experts	2.1. Program Analysis Practical Professional Competencies of Program Evaluation Experts
1.2. Program Analysis Theoretical Professional Competencies of Program Evaluation Experts	2.2. Practical Professional Competencies of Program Evaluation Experts on Research
1.3. Theoretical Professional Competencies of Program Evaluation Experts on Research	3. Professional Values That Program Evaluation Experts Should Have
	4. Professional Skills That Program Evaluation Experts Should Have

At the end of the first round, a 5-point Likert-type questionnaire with 82 items was developed based on the participating expert responses to open-ended questions for the themes presented in Table 2.

Second Round

Questionnaires developed at the end of the first round were sent to the experts. Experts indicated the level of their agreement with each item. The data collected in this stage were analyzed with descriptive statistics. Furthermore, the views of the experts on main and sub-themes and their comments on these concepts were analyzed. The new questionnaire form, which was the input of the third phase, were developed separately for each expert, including the individual and holistic analysis results in this stage.

Third Round

In the third round, experts were asked to reevaluate their responses from the previous round by comparing them with the results of holistic analysis. Thus, the participants had the opportunity to change their views based on their own decisions and those of other participants.

Delphi studies have two important criteria: stability and agreement. Stability entails the lack of a statistically significant difference between the responses in two Delphi rounds. Stability is not an indicator of agreement, but the analysis of the agreement level could be initiated only in the presence of stability eved (Dajani, Sincoff, & Talley, 1979). The stability of the participant views could be calculated with various statistical methods. Scheibe, Skutsch, and Schofer (2002) reported that when the percentage of the number of participants who changed their views is less than 15%, there is stability. There is no set criteria for agreement, which could vary based on the number of participants and the aim of the study (Hasson, et al., 2000, p. 1011). Measures of central tendency (means, median and mode) and distribution (standard deviation and quartiles) are commonly employed to determine the agreement level (Hasson, et al., 2000, p.1012). In the current study, the percentage of the number of participants who changed their views was determined as the agreement criterion. Three out of 12 experts changed their views. Although this meant that 25% of the participants changed their views, the fact that experts only changed their views on 1-2 items among 82 items, and the preexisting high agreement criteria allowed the authors to terminate the Delphi process on the third round. In the study, the agreement criterion was determined based on the mean and median, and frequencies of "I agree" and "I strongly agree" expert responses, which are commonly employed in the literature and provide effective feedback to the participants. The agreement criteria employed in the second and third rounds are presented in Table 4.

Table 4. Agreement criteria

2nd round (n=12)	3rd round (n=12)
Mean $\geq 3,7$	Mean $\geq 3,9$
Median ≥ 4	Median ≥ 4
Frequency 4+ frequency $5 \geq 9$	Frequency 4+ frequency $5 \geq 9$

An average questionnaire item score of at least 3.7 in the second round of the Delphi application, and 3.9 in the third round was considered sufficient for agreement. The requirement for median in the second and third rounds was at least 4. In both rounds, the frequency of the participants who selected "I agree" (4) or "strongly agree" (5) options was required to be at least 9.

Furthermore, a high level qualification criterion was defined based on study data. As presented in Table 5, a high level agreement was concluded when the mean was at least 4.5, the median was at least 5 and the frequency of the experts who responded "I agree" or "I strongly agree" was at least 10 in the third round.

Table 5. High level agreement criteria

3rd round (n=12)
Mean ≥ 4.5
Median $= 5$
Frequency 4+ frequency $5 \geq 10$

Ethical Concerns

In Delphi studies, the author has an ethical obligation to ensure that the identity and views of the participants were not disclosed to another panel member (participant). The decisions and views should remain anonymous throughout the process. In the present study, due attention was paid to the above-mentioned ethical principle. Furthermore, all participating experts were sent a written consent form, providing detailed information about the process, and their consent to voluntary participation was obtained.

Validity and Reliability

Delphi studies are considered reliable when the process is explained in detail. The participation of the same experts in each Delphi round, that is, the evaluation of the participants who raised certain issues in the first round in the next two rounds would improve the validity of the study (Seuring & Müller, 2008, p.458). In

Delphi studies, the questionnaire developed in each round is organized and sent back to the experts for feedback and re-analysis. Thus, the construct validity is inherent (Okoli & Pawlowski, 2004, p.27).

The measures implemented to establish validity and reliability in the study are listed below:
1. The views of 3 faculty members who were Curriculum and Instruction experts were consulted during the development of the open-ended questions in the first round.
2. The participants were assigned based on certain criteria:
    a. PhD degree in Curriculum and Instruction,
    b. Employment as a faculty member.
3. The data collected in the first round were analyzed with content analysis, and the categories and themes were determined individually by the author and a program evaluation and qualitative research expert. Coding reliability was calculated as 0.93 (Miles & Huberman, 1994).
4. The stages of the Delphi process were explained to the participants in detail before each data collection process.

Results and Discussion

In the first stage of Delphi study, two open-ended questions were posed to the experts, and the content analysis conducted on the collected data revealed the main and sub-themes presented in Table 2 on program evaluation expert competencies.

In the second stage, the questionnaire that was developed based on the views of the experts was sent to the experts via e-mail, and they were asked to critically examine and respond to the 82 items in the questionnaire. Fifteen out of the 23 participating experts responded the 5-point Likert-type questionnaire within the predetermined period. In addition to responding to the questionnaire, the experts also provided suggestions to improve the validity of the questionnaire. For example, they altered certain concepts (process-oriented, eclectic, intellectual courage, political dimension, research audience, etc.) or asked for examples to improve the comprehension. Furthermore, certain experts stated that some items were too broad to be considered as a qualification criterion for program evaluation experts and stated that these should be more focused. A small number of participating experts proposed to restructure the themes. However, since the majority of the experts had positive views on the themes, the themes were not changed. Certain experts suggested the removal of certain questionnaire items in the third round, and these suggestions were implemented based on the statistical values of the items. Furthermore, the views of the experts who suggested that certain items should be added to the questionnaire were meticulously evaluated. After the above-mentioned stages, the questionnaire developed for the third Delphi round was finalized after 3 items were removed and 3 items were added. The revised questionnaires in the second and third rounds are presented in Table 6.

Table 6. Themes and questionnaire items

Theme	Number of items in the 2nd round	Number of items in the 3rd round
1. Theoretical Professional Competencies of Program Evaluation Experts	-	-
1.1. Program Development Theoretical Professional Competencies of Program Evaluation Experts	10	10
1.2. Program Analysis Theoretical Professional Competencies of Program Evaluation Experts	5	6
1.3. Theoretical Professional Competencies of Program Evaluation Experts on Research	4	5
2. Practical Professional Competencies of Program Evaluation Experts	-	-
2.1. Program Analysis Practical Professional Competencies of Program Evaluation Experts	20	19
2.2. Practical Professional Competencies of Program Evaluation Experts on Research	15	15
3. Professional Values That Program Evaluation Experts Should Have	15	15
4. Professional Skills That Program Evaluation Experts Should Have	13	12
Total	82	82

Due to word-count limitations, findings associated with items with high-level agreement in each theme are presented in the following sections.

Under the "theoretical professional competencies of the program evaluation experts" main theme, 8 out of 10 questions met the high-level agreement conditions in "program development theoretical professional competencies of program evaluation experts" sub-theme. Based on these items, the program evaluation expert should have knowledge on the program components and the relationships between these components, the theoretical foundations of the program, the program development process and the theories associated with this process, the dimensions that affect the program development, and the approaches employed to determine requirements. Furthermore, the expert should follow the national literature on program development. Under the same main theme, it was determined that there was high-level agreement in 5 out of 6 items in "the theoretical professional program evaluation competencies of the program evaluation experts" sub-theme. Based on expert opinion, program evaluation experts should have knowledge on the program evaluation process, models and legal regulations (laws, regulations, directives, etc.). They could also explain the relationship between program development and evaluation and follow both national and international literature on program evaluation. The last sub-theme in the first main theme, the "theoretical professional research process competencies of the program evaluation expert", there was a high-level agreement in 3 of the 5 items. Thus, almost all experts agreed that curriculum evaluation experts should have knowledge on quantitative, qualitative and mixed research methods.

Under the second main theme of the questionnaire, "the practical professional competencies of the program evaluation experts", there was a high-level agreement in 15 items in the "practical professional program evaluation competencies of program evaluation experts" sub-theme. According to the participants, program evaluation experts should have knowledge on the functions of the program, could develop a program evaluation research design, and decide on the type, model and approach for program evaluation. Furthermore, they could correlate program evaluation types, models, approaches and designs. They could plan a program evaluation process and decide about the stakeholders (sample/participant) based on the predetermined evaluation model and approach. They could develop the criteria and standards required by the program evaluation model and approach. They could share their knowledge on the process with program stakeholders and work in collaboration with the related experts when necessary. In the program evaluation process, they should employ approaches to determine the requirements based on the conditions. They should decide on the termination, correction and improvement of the program based on the program evaluation research data and develop recommendations to overcome the flaws and deficiencies in the program. It was determined that a high-level agreement was established in 13 of the 15 items in the "practical professional research process competencies of the program evaluation experts" sub-theme under the same main theme. In this context, the program evaluation experts are required to conduct field research, employ quantitative, qualitative and mixed research methods, conduct meta-analysis, determine data sources, develop data collection instruments, collect and analyze data, analyze the validity and reliability of the data. Furthermore, they were expected to interpret the findings and to produce an analysis report based on the employed model.

There was a high-level agreement in 12 items under the main theme of "professional values that a program evaluation experts should have". The experts reported that program evaluation experts should respect the stakeholders, be open to criticism, take responsibility, be open to obtain assistance and participate in teamwork. Furthermore, it was emphasized that the program evaluation experts should be patient, willing to follow the developments in the field, have the intellectual courage to uphold the evaluation process, open to learning and innovation, and have an ethical approach with a versatile perspective.

In the last main theme of the questionnaire, "professional skills that a program evaluation expert should have", 10 items met the high-level agreement conditions. Based on expert opinion, program evaluation experts should possess investigative identity, logical consistency, high-level thinking, decision-making, problem solving and self-regulation skills. The expert should be objective, work systematically, communicate effectively and have a holistic perspective.

Conclusion

The content analysis conducted on the responses of field experts to the two open-ended questions in the first round of the study that aimed to determine the competencies of program evaluation experts led to a 82-item Likert-type questionnaire which was utilized as the data collection instrument in the second Delphi round. In the second round, 3 items were removed from the questionnaire based on the expert responses and suggestions, and

3 items were added to different themes. In the third round, the authors decided to end the process since there were high-level agreements in 68 out of the 82 questionnaire items, and no items were removed from the questionnaire based on the criteria.

In the study, the program evaluation experts' competencies were categorized in 4 main themes. It could be suggested that the participating experts considered theoretical knowledge on a topic and related practical skills was different competencies. Thus, "theoretical professional competencies of program evaluation experts" and "practical professional competencies of program evaluation experts" were determined as the main themes. Furthermore, "professional values that a program evaluation experts should have" and "professional skills that a program evaluation experts should have" main themes were determined.

In international studies where program evaluation expert competencies were determined, although the classification of the competencies were different when compared to the main themes determined in the present study, it could be argued that certain items were similar in sub-dimensions. The classifications reported in studies conducted by Stevahn, King, Ghere, and Minnema (2005) and Scriven (1996a) to determine program evaluation expert competencies were significantly similar to the findings obtained in the current study. In a study by Stevahn, King, Ghere and Minnema (2005), where program evaluation expert competencies included 61 items under 6 main categories, especially the findings obtained under "systematic inquiry", "reflective practice", "situational analysis" and "interpersonal competence" categories were consistent with the findings of the present study. The "systematic inquiry" category included the theoretical dimension of program evaluation and research paradigms, and the evaluation process analysis and reporting. Similarly, in the present study, the "theoretical professional competencies of program evaluation experts" main theme included theoretical foundation competencies in the field in the dimension of "theoretical professional program evaluation competencies." Furthermore, the theoretical and practical professional research process competences and the research paradigm competency sub-themes were included. Also, in the "reflective practice" category, the authors mainly included program evaluation expert competencies in the dimensions of professional involvement and development. It was observed that similar competences were included in the "professional values that a program evaluation experts should have" main theme in the current study. "Situational analysis" category included competencies associated with items on organizational and political analysis. Although there was no high-level agreement on the above-mentioned items in the study, there were agreements on similar items. Furthermore, it was determined that the "interpersonal competence" category and the "professional skills that a program evaluation experts should have" main theme had similar sub-dimensions such as effective communication, problem solving and teamwork skills.

In the classification by Scriven (1996), it would not be inaccurate to argue that the research methodology, self-assessment, ethical analysis, determination of the needs, conceptual organization and the theoretical and methodological design, planning and maintenance of the evaluation and reporting the findings were significantly similar to the present study findings.

On the other hand, there were also different findings in certain present study dimensions when compared to the competences reported in the above-mentioned studies. Stevahn, King, Ghere, and Minnema (2005) emphasized that evaluation was to benefit the client, participant, and institution in particular, and for public benefit in general, and addressed the program evaluation experts' competencies within the "professional practice" category. Furthermore, items such as development of the required agreements between the evaluation experts and the client, and cost justification included in the "project management" category were not included in the findings of the present study and budgeting the evaluation item was removed in the second round due to the lack of an agreement. The study conducted by Scriven (1996) reported similar competence items on budgeting and common effect dimensions when compared to Stevahn, King, Ghere and Minnema (2005).

As mentioned above, although there are countries with undergraduate programs for program evaluation experts such as the United States, Canada, Australia, Iceland, Belgium and the West Indies (LaVelle and Donaldson, 2015), individuals of different educational backgrounds could be trained as program evaluation experts in professional development workshops, webinars and online courses (Lavelle & Donaldson, 2015; McGuire & Zorzi, 2005). In Turkey, program expertise is not an area of expertise, and the graduates of the graduate programs that include program evaluation courses are not employed as program evaluation experts in the private sector. Abroad, program evaluation experts are not only employed in education but also in human resources departments of corporations that produce goods or services in various industries (McGuire & Zorzi, 2005). Thus, unlike Turkey, program evaluation experts in foreign nations who could find employment in the private sector should possess "customer satisfaction" and "budget" competencies.

In a study on the reasons for the underdevelopment of program evaluation theories, King (2003) reported that the field of program evaluation is a pragmatic field, and customer demands are dominant in program evaluation processes, while the theory and requirements are neglected. This could explain the differences between the present study findings and the findings of international studies. Thus, it would be accurate to argue that the experts in the present study did not consider program evaluation experts as a private sector employee but as an individual with a solid theoretical background in the program evaluation field and employed in education industry based on practical competencies. Thus, concepts such as "customer" and "budget" were not included among the competencies determined in the present study conducted with Turkish participants.

Although the items in the main themes of "professional values that a program evaluation experts should have" and "professional skills that a program evaluation experts should have" determined in the present study did not exist as a separate competency category in international literature, they were included as sub-dimensions in reported categories. Except "assigning significance to evaluation" mentioned in professional values and skills, it could be argued that responsibility, learning, problem solving, and high-level thinking skills were competencies that should be adopted by all occupational groups.

Program evaluation, which was described as an old practice but a new discipline abroad in late 20th century (Scriven, 1996), was a developing field in early 2000s (King, 2003), and is a field of expertise that defines various roles for evaluation experts today (Volkov, 2011). It could be suggested that the development of the field and expansion of the meaning of the field in Turkey would benefit several fields. Thus, various educational opportunities, similar to those provided in other countries, should be provided for program evaluation experts, who are primarily trained in master's and doctorate programs in CI in Turkey. Furthermore, it would be of great benefit to increase the number of program evaluation courses, which are currently quite limited and mostly theoretical in the above-mentioned educational programs, and to restructure these courses to emphasize the acquisition of practical competencies (Atik Kara, Kürüm Yapıcıoğlu, & Sever, 2020).

Recommendations

Updating the job description by altering the perception of "program evaluation expert" as an evaluation expert would allow program evaluation experts to work towards the improvement of service quality and promotion of systemic renewal in various industries. Thus, it could be suggested that program development graduate programs will be more in demand and graduates will have job prospects in various fields. However, as mentioned in the international literature, the uncertainty about the program evaluation expert competencies due to their training in different types of institutions will be eliminated, and the development and diversification of educational opportunities in the field will allow the management and implementation of the programs in the private sector and education by more competent individuals based on the theory and more effective operation of the programs due to the feedback of these individuals.

It could be suggested that the 82 items in 4 dimensions determined in the present study could be employed as a checklist for program evaluation experts. The questionnaire, which could be considered by field experts as a list of program evaluation expert competencies, could also be employed by experts to determine self-competencies or incompetence in certain fields. Thus, experts who determine the fields that they are competent and the fields that they are incompetent could utilize this information to plan their career development. Furthermore, the content of the programs that currently train program evaluation experts could be updated based on the presented items. Also, workshops, webinars or short-term training courses could be organized to train program evaluation experts who could benefit from the content organization dimension.

It was expected that the program evaluation expert competencies determined in the present study would fill a significant gap in the national literature. To improve the validity and reliability of the items, further studies could be conducted with a quantitative approach and larger samples. On the other hand, future qualitative studies could be conducted with experts with significant studies on the field to obtain in-depth information on basic program evaluation expert competences and sub-dimensions. Furthermore, the data collection instrument could be analyzed in a further scale development study on the 82 items in 4 sub-dimensions.

Today, the increasing knowledge base in every field requires individuals to specialize in particular fields. Thus, the program evaluation field should be accepted as a field of expertise in the 21st century in Turkey. Further research findings on program evaluation could be an important stone in the road of progress in the field by reducing challenges such as the lack of human resources and field education.

Acknowledgements

References

Atik-Kara, D., Kürüm-Yapıcıoğlu, D., & Sever, D. (2020). Eğitim programları ve öğretim lisansüstü eğitim programlarının incelenmesi [Examination of curriculum and instruction graduate education programs]. Anadolu Üniversitesi Eğitim Fakültesi Dergisi, 4(2), 163-190.

Barnes, D. (1982). Practical curriculum study. London: Routhledge.

Bauer, M. W. (2000). Classical content analysis: A review. In M. W. Bauer & G. Gaskell (Eds.), In qualitative researching with text, ımage, and sound: A practical handbook (pp. 131-151). London: Sage Publications.

Büyüköztürk, Ş., Çakmak, K. E., Akgün, E. Ö., Karadeniz, Ş., & Demirel, F. (2010). Bilimsel araştırma yöntemleri [Scientific researh methods]. Ankara: Pegem Akademi.

Clayton, M. J. (1997). Delphi: A technique to harness expert opinion for critical decision-making tasks in education. Educational Psychology, 17(4), 373-386.

Cooksy, L. J., & Mark, M. M. (2012). Influences on evaluation quality. American Journal of Evaluation, 33(1), 79-87.

Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. In J. B. Cousins & I. Bourgeois (Eds.), Organizational capacity to do and use evaluation. New Directions for Evaluation, 141, 7-23.

Covert, R.W. (1992). Successful competencies in preparing professional evaluators. Paper presented at the annual meeting of the American Evaluation Association, Seattle, WA.

Creswell, J. W. (2008). Research design: Qualitative, quantitative and mixed methods approaches. Thousand Oaks, CA: Sage

Creswell, W. J., & Clark, L. V. (2007). Designing and conducting mixed method research. London Sage Publications.

Dajani, J.S., Sincoff, M.Z., & Talley, W.K. (1979). Stability and agreement criteria for the termination of Delphi studies. Technological Forecasting and Social Change, 13, 83-90.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. Management Science, 9(3), 458-467.

Engle, M., Altschuld, J. W., & Kim, Y. (2006). 2002 Survey of evaluation preparation programs in universities: An update of the 1992 American Evaluation Association-sponsored study. American Journal of Evaluation, 27, 353-359.

Fletcher, A., & Childon, G. P. (2014). Using the Delphi method for qualitative, participatory action research in health leadership. International Journal of Qualitative Methods, 13, 1-18.

Habibi, A., Sarafrazi, A., & Izadyar, S. (2014). Delphi technique theoretical framework in qualitative research. The International Journal of Engineering and Science, 3(4), 8-13.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. Journal of Advanced Nursing, 32(4), 1008-1015.

King, J. A. (2003). The challenge of studying evaluation theory. New Directions for Evaluation, 97, 57-66.

King, J. A., & Stevahn, L. (2015). Competencies for program evaluators in light of adaptive action: What? So what? Now what? In J. W. Altschuld & M. Engle (Eds.), Accreditation, certification, and credentialing: Relevant concerns for U.S. evaluators. New Directions for Evaluation, 145, 21–37.

King, J.A., Stevahn, L., Ghere, G., & Minnema, J. (2001). Toward a taxonomy of essential evaluator competencies. American Journal of Evaluation, 22(2), 229-247.

LaVelle, J. M. & Donaldson, S. I. (2015). The state of preparing evaluators. In J.W. Altschuld & M. Engle (Eds.), Accreditation, certification, and credentialing: Relevant concerns for U.S. evaluators. New Directions for Evaluation, 145, 39–52.

McGuire, M., & Zorzi, R. (2005). Evaluator competencies and performance development. The Canadian Journal of Program Evaluation. 20(2), 73-99.

Melrose, M. (1998). Exploring paradigms of curriculum evaluation and concepts of quality. Quality in Higher Education, 4:1, 37-43.

Mertens, D.M. (1994). Training evaluators: Unique skills and knowledge. In J.W. Altschuld & M. Engle (Eds.), The preparation of Professional evaluators: Issues, perspectives, and programs. New Directions for Program Evaluation, 62, 17–27.

Miles, M. B., & Huberman, A. M. (1994). An expanded source book: Qualitative data analysis. London: Sage Publications.

Nworie, J. (2011). Using the Delphi technique in educational technology research. Tech Trends, 55(5), 24-30.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. Information & Management, 42(1), 15-29.

Oliva, P. F. (1988). Developing the curriculum. Glenview, Ill. : Scott, Foresman/Little, Brown College Division.

Ornstein, A.C., & Hunkins, F.P. (2009). Curriculum: Foundations, principles, and issues. USA: Pearson Education, Inc.

Özçelik, D. A. (1992). Eğitim programları ve öğretim [Curriculum and Instruction]. Ankara: ÖSYM Yayınları.

Powell, C. (2003). The Delphi technique: Myths and realities. Journal of Advanced Nursing, 41(4), 376-382.

Rowe, G., & Wright, G. (2001). Expert opinions in forecasting: The role of the Delphi Technique. In J. Armstrong (Ed.). Principles of Forecasting (pp. 125-144). Boston: Kluwer Academic.

Scheibe M, Skutsch M, Schofer J (2002). IV. C. Experiments in Delphi methodology. In: Linstone H, Turoff M (Eds.) The Delphi method: techniques and applications (pp 257-281).[Electronic version]. Newark, NJ: New Jersey Institute of Technology.

Scriven, M. (1996). The theory behind practical evaluation. Evaluation, 2, 393-404.

Scriven, M. (1996a). Types of evaluation and types of evaluator. American Journal of Evaluation, 17, 151-161.

Seuring, S., & Müller, M. (2008). Core issues in sustainable supply chain management–a Delphi study. Business Strategy and the Environment, 17(8), 455-466.

Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate research. Journal of Information Technology Education, 6, 1-21.

Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005a). Establishing essential competenciesfor program evaluators. American Journal of Evaluation, 26(1), 43-59.

Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005). Evaluator competencies in university-based evaluation training programs. The Canadian Journal of ProgramEvaluation, 20(2), 101-123.

Şahin, A. E. (2009). Türkiye' de ilköğretim okulu müdürlüğünün bir meslek olarak mevcut durumu: Bir Delphi çalışması [The state of elementary principalship as a profession in Turkey: A Delphi Study]. Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, 26 (26), 125-136.

Şahin, A. E. (2001). Eğitim araştırmalarında Delphi tekniği ve kullanımı [Delphi technique and its uses in educational research]. Hacettepe Üniversitesi Eğitim Fakültesi Dergisi 20: 215 - 220

Volkov, B. B. (2011). Beyond being an evaluator: The multiplicity of roles of the internal evaluator. In B. B. Volkov & M. E. Baron (Eds.), Internal evaluation in the 21st century. New Directions for Evaluation, 132, 25- 42.