Developing a Program Evaluation Scale to Investigate Perceptions about Primary School English Language Teaching Program

Dinçay Köksal

Çanakkale Onsekiz Mart University, Turkey

Pınar Çankaya[1]

Tekirdağ Namık Kemal University, Turkey

Abstract

Evaluation serves as a major means to identify the benefits and effectiveness of a teaching program. To capture as many potential angles as possible and investigate the opinions of stakeholders with regards to the elements of teaching programs, researchers and program developers generally tend to rely on quantitative survey data. To this end, program evaluation questionnaires play a remarkable role since they ensure to gather quick and sufficient numerical data from a considerable number of respondents. However, the suitability, reliability, and appropriateness of questionnaires for investigating the opinions of stakeholders regarding English language teaching programs (ELTP, henceforth) are considered as narrow-scoped. Based on the argument, the primary concern of the current research is to develop a program evaluation scale for exploring opinions of language teachers by providing insights from the relevant literature on all possible angles of ELTP with a particular focus on objectives, content, teaching learning process and assessment of the program. To serve this purpose, a seven-step survey development framework was used for constructing the scale. The findings led to the development of a 33-item scale that involved five components: (a) general overview, (b) goals, (c) content (d) teaching and learning process, and (e) assessment. Moreover, the overall results indicated Cronbach's alpha reliability of .94 for the 33-itemed scale.

Key words: Program evaluation, scale development, ELTP

[1]Corresponding Author. Email: pcankaya@nku.edu.tr

Introduction

Foreign language teaching and learning have maintained its place and importance in education throughout the world for ages. Given the importance and a huge amount of effort, money and time; it is not still at a satisfactory level of success. Therefore, researchers have been carrying out a great number of studies on possible reasons from diverse angles ranging from language learners and teachers to language programs. Considering Turkish context, teaching and learning English is regarded as a significant problem demonstrated by many figures within the field (Demircan, 1988; Demirel, 2005; Demirel, 2006; Dinçer, Takkaç & Akalın, 2010; İnceçay, 2012; Kırkgöz, 2009; Kızıldağ, 2009; Soner, 2007; Tok &Arıbaş, 2008). Therefore, the possible reasons might include teachers' competences, learner differences and the teaching program itself. Of particular interest, ELTP is the central focus of this current study. In this sense, the literature review is going to focus on program evaluation in general terms, and ELT Pin particular.

Program evaluation

Program evaluation is a pivotal issue in education. In order to reach a clear understanding, three questions need to be posed and answered. Firstly, it is crucial to define what evaluation is. Brown defined evaluation as "the systematic collection and analysis of all relevant information necessary to promote the improvement of a curriculum, and assess its effectiveness and efficiency, as well as the participants' attitudes" (1995, p.223) while Lynch (1996) emphasized reaching judgments or decisions systematically with regards to evaluation.

Having defined the term, it is quite necessary to explain the main driving forces and reasons of evaluation. Without evaluating both the process and the outcomes of a teaching program, it is clear that the benefits and effectiveness of the program cannot be identified. Several additional significant reasons of evaluation can be listed as follows:

- To provide feedback to staff and participants,
- To shape policy development,
- To examine why a program succeeds or fails,
- To consider unexpected positive or negative effects,
- To examine whether the goals are appropriate for the learners

- To develop good practice,

- To make the best use of scarce resources

- To identify deficiencies in a program

- To improve the effectiveness of the program (Alderson & Beretta, 1992; Rolstad, Mahoney,& Gilas, 2005; Worthen, Sanders,& Fitzpatrick, 2004).

To put it short, evaluation is done for accountability, improvement, and program and teacher development. Having emphasized the importance and reasons of evaluation, the evaluation process of how to do it also deserves a serious consideration. There are many ways to conduct an evaluation neither of which is better than the other. Generally speaking, two main forms of evaluation come forefront. To illustrate, formative evaluation (known as process evaluation) occurring during the implementation process attempts to find out the problematic parts to make improvements while summativeevaluation (known as outcome evaluation) taking place after the intervention aims at determining the impact of the program taking into attainment of the objectives account (Rolstad,etal., 2005).

Having discussed the definition, purposes and types of program evaluation, the following section offers specific details about ELTP.

ELTP

Broadly speaking, a teaching program can be defined as "a series of courses linked with some common goal or end product (Lynch, 1996, p.2). It can be detailed as "any set of replicable procedures, materials, professional development, or service configurations that educators could choose to implement to improve student outcomes" by Slavin (2008, p.12). As the definitions highlight, a teaching program requires a purpose to achieve through a set of activities in which an assessment takes place for the end-product. Thus, the basic components of a teaching program might be listed as aims/objectives, content, teaching/learning process and assessment/evaluation. With regards to ELTP, the objectives can be written as the intended behaviors which are planned to be achieved by language learners. The objectives are of decisive importance as the other teaching program elements are organized in accordance with them. More importantly, they play crucial role in ordering the content and knowledge to be covered in addition to the organizing the teaching and learning process (Erden, 1995). Accordingly, content tries to cover what to teach within the program while teaching learning process attempts to arrange content and objectives based on contextual and situational

concerns. All issues concerned with the learning environment, timing, strategies, classroom activities, materials, interactions, techniques and methods are included in this category (Demirel, 2006). Last but not least, assessment answers the question of how much the objectives are achieved by learners thus providing necessary information about the deficiencies, strengths and weaknesses of the program.

Methodology

Since both quantitative and qualitative data were used, the current study embraced a mixed-type of methodology. In doing so, a comprehensive literature review and a clear in-depth understanding of what participants think about the program were achieved. As the main purpose of this paper is to develop a reliable instrument for evaluating a teaching program, the seven step framework was used as methodology of the research. The steps can be listed as follows based on the information by Gehlbach, Artino, and Durning (2010, p.925).

- Conducting a literature review,
- Conducting interviews,
- Synthesizing the literature review and interviews,
- Develop items,
- Conducting an expert validation,
- Conducting cognitive pre-testing,
- Pilot testing.

Instrument development

In an attempt to develop a reliable and comprehensive instrument to evaluate ELTP, the researchers developed the questionnaire based on the seven-step framework. As the first step, a literature review was conducted in order to achieve two goals, one of which was to build background knowledge about the construct under investigation within the study while the second purpose was to review already existing scales, thus attaining a combination of items (Adıgüzel & Özüdoğru, 2014; Alkan & Arslan, 2014; Bayrak & Erden, 2007; Erkan, 2009; Gömleksiz & Bulut, 2007; Güneş, 2009; İnam, 2009; Küçük, 2008; Mersinligil, 2002; Merter, Kartal & Çağlar, 2012; Orakçı, 2012; Seçkin, 2010; Yaman, 2010). Secondly, interviews through think-aloud protocol were carried out with practicing teachers as the active users of the program could help gain insights about the program which is worthy of

serious consideration. Moreover, synthesizing the literature review and interviews was done in order to overcome any ambiguous conceptualizations and expectations with regards to the program of interest. Incorporating the data obtained from the previous steps, item pooling was done as the fourth step with the central aim being to cover all relevant elements in detail. The items related to objectives, content, learning/teaching process and assessment in Ministry of National Education (MoNE, 2015) are listed as items in a questionnaire. The next issue arising in the discussion of developing questionnaires was expert consultation which took place as the fifth step in the current study in order to determine the relevance and suitability of the items. Through the feedback taken from the experts, necessary alterations including the number of the items, the language of the questionnaire with respect to the constructs in the study were adjusted. Having conducted cognitive interviews with four English language teachers, the researchers ensured the clarity of the items as intended. Then, a near-final version of the questionnaire was designed.

Having reached a near-final version of the questionnaire, the researchers used back translation method for translating the questionnaire before administering pilot testing. The main purposes of translating the questionnaire are to produce "a close translation of the original text so that we claim that the two versions are equivalent and to produce natural-sounding texts in the target language" (Mackey & Gass, 2011, p.79). As it is expected, badly translated questionnaires can cause a research study fails because it hinders collecting comparable data. In this sense, team-based brainstorming was also done with colleagues with an aim to back translate the target language version into the source language. The two source language versions were compared to find out if there were any problems in the target language text.

As the last step, pilot testing was carried out with the help of practicing teachers. The piloting process will be detailed in the following section.

Procedures for data collection

Pilot study

Piloting the questionnaire was implemented by applying the questionnaire with a group of participants who were similar to the target population via Google document link. The results of the pilot study are of crucial as they help to develop a final version of the questionnaire by excluding ambiguous, too difficult/ease, or irrelevant items and rehearsing

the administration procedures, finally improving the clarity (Mackey & Gass, 2011). With this purpose, a pilot testing was done to make necessary changes based on the participants' views.

Setting and participants

Having applied ethical terms, the researchers ensured volunteerism and confidentiality stating that the answers would be used only for academic purposes and kept confidential as anonymous. After that, the pilot testing was carried out with 115 practicing English teachers teaching in various cities in Turkey. Table 1 presents the distribution of the participants in terms of their gender, education degree and teaching experience.

Table 1

The Distribution of Gender, Experience, Degree of the Participants

	Number of the Participants	
Gender	Female	101
	Male	14
Degree	Bachelor	94
	Master	21
	Doctorate	_
Experience		
	1-5 years	66
	6-10 years	38
	11-15 years	9
	16- over	2

As demonstrated in Table 1, the total number of practicing teachers took place in the pilot testing was 115 of whom101 was female participants while 14 participants were male. The initial version of the "Program Evaluation Scale (PES)" was completed by those whose teaching experiences ranked from 1-5 years (*n*= 66) to 16 over (*n*=2). The ages of the participants in the piloting stage ranked from 23 to 50 at an average of 27. Additionally, six of the participants took place voluntarily in semi-structured interviews including four open-ended questions which were recorded and transcribed.

Procedures for data analysis

The obtained data were analyzed using the Statistical Package of Social Science (SPSS, version 20). Descriptive statistics, including frequencies, means, and standard deviations were used to explore the demographic data. Additionally, in order to explore any differences among participants in terms of their gender independent samples t-tests were computed. Besides, ANOVA statistics were done in order to determine the differences in terms of teaching experiences. In addition, the .05 level of statistical significance was set at all statistical tests.

Considering the qualitative data, as a first step the raw data were read to reach a general sense of the information which was based on the responses to open-ended interview questions. Later on, the data were organized into logical and meaningful categories. Moreover, similar and common themes emerged from the data were identified. In brief, identification and naming the categories were done by putting the similar words, notions and basic ideas into the same category. As a result, a set of categories were developed which reflect the underlying ideas of teacher participants (Saldana, 2009). The obtained data were analyzed by two independent raters who have experience on qualitative studies, thus increasing inter-rater reliability.

Results

Findings of factor analyses

In order to reduce and group the items, further to identify the reliability of the scale; exploratory factor analysis was computed. Taking the relevant literature and individual interview results into account, the researchers developed 55-itemed PES (initial version). Feedback taken from the respondents was analyzed through exploratory factor analysis. In the pursuit of implementing factor analysis, the suitability of the items was controlled, thus the Kaiser-Meyer-Olkin (KMO) coefficient value was measured as presented in the following table.

Table 2

KMO and Bartlett's Test Results

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.917
Bartlett's Test of Sphericity	Approx. Chi-Square	6185.802
	df	1485
	Sig.	.000

As demonstrated above, the factorability of the 55 items in the PES was fairly acceptable. The results revealed that KMO coefficient value was .917 which was regarded as considerable to calculate factor analysis (Büyüköztürk, 2007).Based on the results of KMO coefficient value, the first factor analysis was computed and the results were illustrated in Table 3.

According to the initial Eigen values, the 52.4 % of the variance was illustrated in the first factor, while the second factor explained 6.97 % followed by the third factor 5.14% and the fourth factor 4.18 %. Having explored factor loads in each component, the researchers attempted to exclude uncertain items with reference to two basic criteria based on the Büyüköztürk's study (2007). The first criteria necessitated to eliminate the items the factor loads of which were lower than .450. Moreover, the second criteria debated the difference between two highest factor loads. If the difference was lower than .100, the items were removed.

Based on this introductory discussion of certain criteria, the items 11, 23, 38, 33, 34, 37, 40 were excluded due to their factor loads being lower than .450. Additionally, according to the second criteria mentioned above, the items 2, 21, 44, 45, 25, 39, 42, 53 were omitted as the difference between two factor loads was smaller than .100. The factor loads of each item were shown in the following table.

Table 3

Factor Loadings Based on a PCA with Varimax Kaiser Normalization Rotation for 55 Items of the PES (N = 115)

Items									
Evaluation fosters learners to self-evaluate themselves (i47)	.740	.110	.280	.216	.121	.141	.185	.100	.002
Evaluation is explained in detail in the program (i48)	.729	.203	.218	.182	.227	.110	.022	.175	.108
Out of class assessment types are used (i46)	.689	.165	.045	.065	.337	.015	.268	.118	.035
Assessment types are in accordance with the goals (i51)	.676	.154	.266	.148	.162	.455	.128	.025	.040
Assessment is in accordance with the content (i50)	.601	.331	.397	.104	.051	.428	.026	.088	.063
Evaluation shows whether the goals are achieved by learners (i49)	.594	.261	.272	.220	.162	.360	.131	.156	.043
Speaking skill is adequately covered in the content (i27)	.498	-.005	.230	.409	.348	-.004	.149	.277	.062
Activities are organized based on learners' needs/interests (i38)	**.420**	.262	.408	.349	.239	.154	.170	.324	.097
The content is ranked from concrete to abstract items (i23)	**.420**	.271	.294	.373	.193	.039	.382	.002	.256
CEFR is strictly followed by the program (i11)	**.398**	.205	.081	.394	.050	.222	.357	.360	.053
Content includes enjoyable visual, audio, auido-visual materials (i33)	**.365**	.329	.018	.316	.333	.257	.175	.033	.191
Goals are suitable for learners' age (i14)	.143	.782	.218	.155	.179	.039	.184	.246	.078
Goals are suitable for learners' emotional development. (i16)	.182	.750	.326	.146	.173	.180	.151	.204	.112
Goals are suitable for learners' cognitive development (i15)	.075	.746	.334	.170	.100	.127	.184	.175	.096
Goals are in accordance with the content. (i18)	.332	.660	.195	.227	.098	.176	.121	.142	.147
Goals are attainable by learners (i17)	.259	.624	.257	.162	.214	.203	.292	.281	.088
Content provides learners with enjoyable, stress-free environment(i19)	.257	.552	.194	.379	.276	.240	.271	.014	.141
Content attracts the students' attention and curiosity (i20)	.280	.540	.268	.356	.132	.324	.021	.114	.086
The topics in units support each other (i26)	.239	.229	.632	.283	.076	.187	.204	.224	.105
C. activities are suitable to learners' physical development (i43)	.297	.331	.614	.083	.144	.205	.212	.227	.020
Content allows to use different methods and techniques.(i32)	.116	.318	.586	.217	.120	.226	.321	.040	.064
Content is suitable for learners' readiness level (i31)	.108	.277	.578	.139	.199	.051	.255	.343	.101
Teaching techniques are suitable for the level of classroom (i36)	.213	.315	.541	.232	.334	.024	.208	.207	.274
C. activities are suitable to learners' emotional development (i44)	**.494**	.325	**.540**	.165	.162	.211	.170	.077	-.023
T/L process is suitable for using a mix of instructional techniques (i35)	.240	.200	.529	.258	.358	.218	.128	.106	.279
C. activities are suitable to learners' mental development. (i45)	**.476**	.338	**.513**	.165	.245	.145	.150	.154	.114
Content is chosen from learners' daily life (i24)	.256	.366	.511	.388	.292	.202	.009	.020	.135
Content is in accordance with goals (i25)	.352	**.407**	**.482**	.350	.258	.161	.080	.205	.164
Teaching / learning process is in accordance with content (i37)	.267	.281	**.440**	.438	.288	.181	.319	.123	.043
Classroom activities are in accordance with the goals (i40)	.280	.312	**.415**	.302	.321	.273	.298	.145	.020
Program forms a basis for the students' future needs in English (i12)	.219	.232	.372	.659	.010	.161	.094	.183	.043
Goals support and complete each other (i13)	.209	.412	.199	.611	.098	.116	.254	.008	.053
Content allows learners to use Eng. as a means of communication (i21)	**.431**	**.371**	.212	.527	.233	-.055	.193	.057	.108
Classroom activities have relevance in students' daily lives (i39)	.285	.280	**.445**	**.504**	.344	.159	.011	.108	.089
A wide range of learning styles are addressed in the program (i42)	**.441**	.186	.288	**.469**	.114	.189	.407	.167	.026
The program provides a bridge between the known to the unknown (i9)	.287	.226	.269	.468	-.087	.207	.431	.204	.010
It is possible to evaluate listening skills in the program. (i54)	.358	.141	.319	.002	.739	.030	.087	.168	.118
It is possible to evaluate speaking skills in the program (i55)	.305	.111	.279	.063	.687	.175	.231	.146	.132
Listening skill is adequately covered in the content (i28)	.134	.315	.084	.267	.607	.125	.005	.223	.028
Goals of the program are able to be evaluated (i53)	.270	.291	.248	.074	**.450**	**.365**	.334	.046	.051
In-service training is essential to implement the program (i8)	.274	.107	.138	-.004	-.025	.670	.056	.077	.052
Portfolio evaluation is useful (i52)	.156	.313	.199	.131	.430	.566	.109	-.047	.086
The program allows learners to have an active role in class (i2)	.094	.167	.058	**.467**	.149	**.532**	.044	.419	.195
The program is student-centered (i1)	.052	.074	.121	.431	.241	.518	.119	.388	.073
It is possible to make learners develop positive attitudes to English (i5)	.159	.211	.220	.306	.212	.505	.401	.167	.239
It is possible to make learners enjoy English (i4)	.174	.194	.333	.261	.242	.458	.304	.321	.140
Activities are designed by taking learner differences into account(i41)	.245	.164	.334	.172	.239	.071	.686	.199	.092
The content is ranked from simple to complex items (i22)	.256	.351	.155	.246	.061	.130	.606	.150	.210
Time allocated to units is consistent with the degree of diffculty(i29)	.068	.341	.275	.055	.142	.140	.522	.300	.289
Time allocated to each unit is sufficient (i3)	.073	.170	.110	.006	.088	.031	.104	.740	.002
It is easy to understand and implement the program. (i7)	.374	.263	.232	.201	.144	.175	.150	.571	.016
The program guides teachers well (i6)	.357	.294	.010	.280	.248	.256	.201	.507	.130
The goals are clearly and explicitly stated (i10)	.371	.139	.225	.272	.058	.279	.266	.395	.241
The number of words to be taught in each unit is inadequate. (i30)	.008	.072	.081	.003	-.174	.064	.096	.031	.808
Both the target culture and international cultures are presented in a positive and non-threatening manner in the content (i34)	.292	.179	.292	.340	.271	.166	.070	.210	**.435**

After omitting the items mentioned above, a Principal Component Analysis (PCA) of the remaining 40 items was calculated one more time. Consequently, the items 12, 9, 24, 7 and 22 were dropped due to their factor loads. The factor analysis was repeated for the third time, which led to omit the items 29 and 3. For the remaining 33 items, another factor analysis was conducted and KMO coefficient value was measured. Since KMO coefficient was valued as .926 and factor loads were acceptable, it was ascertained that it was not necessary to exclude any items.Based on the PCA results in Table 4 below, five groups of factor loads were demonstrated.

Table 4

Factor Loadings Based on a PCA with Varimax Kaiser Normalization Rotation for 33 Items of the PES (N = 115)

Items	Goals	Teaching learning	Content	Evaluation	Overall
Goals of the program are attainable by learners (i11)	**.840**	.208	-.072	-.131	.105
Goals are suitable for learners' emotional development (i10)	**.805**	.255	-.156	-.308	.021
goals are in accordance with content (i12)	**.763**	.188	.085	-.301	.141
goals are suitable for learners' age (i8)	**.759**	.371	-.142	-.253	.157
goals are suitable for learners' cognitive development (i9)	**.750**	.408	-.054	-.201	.007
goals support and complete each other (i7)	**.684**	.145	.086	-.219	-.103
The goals are clearly and explicitly stated (i6)	**.671**	.109	-.298	.194	.396
C.activities are suitable to learners' physical development (i24)	.058	**.786**	-.056	.112	-.036
Teaching techniques are suitable for the level of classroom (i22)	.144	**.785**	-.076	.203	-.100
T/L process is suitable for using a mix of instructional techniques (i21)	-.041	**.726**	.112	.131	-.316
Activities are designed by taking learner differences into account(i23)	.011	**.702**	-.214	.321	-.058
Content provides learners with enjoyable, stress-free environment (i13)	.056	.123	**.810**	-.195	-.101
Content attracts the students' attention and curiosity (i14)	.155	.103	**.779**	-.227	.007
The topics in units support each other (i15)	.062	.121	**.756**	.223	-.023
Content is suitable for learners' readiness level (i19)	-.172	.191	**.721**	.309	-.024
Content allows to use different methods and techniques (i20)	.066	.173	**.716**	.151	-.288
Speaking skill is adequately covered in content (i16)	-.071	.302	**.628**	.335	.240
listening skill is adequately covered in content (i17)	-.317	-.062	**.598**	-.215	-.150
The number of words to be taught in each unit is inadequate (i18)	.160	.377	**.535**	.290	.035
Evaluation shows whether the goals are achieved by learners (i28)	-.031	-.284	.203	**.798**	.162
Assessment is in accordance with the content (i29)	-.091	-.230	.290	**.790**	.097
Assessment types are in accordance with the goals (i30)	-.007	-.431	.280	**.756**	.091
Evaluation is explained in detail in the program (i27)	-.053	-.390	.108	**.709**	.217
Evaluation fosters learners to self-evaluate themselves (i26)	.069	-.398	.166	**.703**	.085
It is possible to evaluate speaking skills in the program (i33)	.101	-.347	-.383	**.682**	-.180
Portfolio assessment is useful (i31)	-.170	-.157	-.021	**.658**	-.221
It is possible to evaluate listening skills in the program (i32)	.026	-.380	-.498	**.643**	-.125
Out of class assessment types are used (i25)	-.002	-.400	-.059	**.584**	.247
It is possible to make learners enjoy English (i2)	-.072	.009	.056	.213	**.778**
Possible to make learners develop positive attitudes to English (i3)	-.192	.056	.186	.152	**.749**
The program guides teachers well (i4)	-.009	.024	.307	.238	**.728**
The program is student-centered (i1)	.357	.167	.342	.198	**.576**
In-service training is essential to implement the program (i5)	.025	-.170	.361	-.040	**.456**

Note. Explained variance: Total = 69.7%; Factor1 = 36.4%; Factor2 = 13.7%; Factor3 = 5.8%; Factor4 = 5.1%; Factor5 = 4.2%.

While the first component included seven items, the second one constituted of four items. The third factor involved eight items, the fourth factor included nine and the last factor constituted five items. As a result, the five factored version of PES could explain 65% of the variance. In this sense, the final version of the scale included five components which were named as 'Overall characteristics', 'Objectives', 'Content', 'Teaching/learning processes and 'Assessment' respectively.

Having administered Factor Analysis, internal-consistency reliability was computed and Cronbach's alpha coefficients were revealed for the entire set of 33 items in the PES (See Appendix)as $\alpha = .966$ which specifies that the scale is highly reliable. Accordingly, the Cronbach's alpha values of each component were calculated as $\alpha =. 841$, $\alpha =. 921$, $\alpha =. 840$, $\alpha =. 843$ *and* $\alpha =.920$ respectively which suggests that the instrument is highly reliable (Büyüköztürk, 2007).

In order to interpret the qualitative data, the obtained responses were thematically coded so as to form a basis of program evaluation components. While analyzing the components, the recurrent themes were noted such as positive and negative aspects of an English language teaching program, the possible problems during implementing process. In doing so, the central framework of a program evaluation scale was drawn based on the participants' responses. In this sense, one of the teachers noted that, "Early start is positive for learning a language, but teachers themselves need to be ready for it (T1)". Thus, starting age, readiness level of both students and teachers in terms of psychological and emotional aspects were taken into account while developing the instrument. Another recurrent theme in the data set regarding positive aspects of the program was game based learning. One of the teachers commented: "It is important for young learners to learn English in an enjoyable way through songs, games, drama (T5)". Based on this point of view, the learning environment and the classroom atmosphere were discussed as items in the instrument.

In terms of teaching and learning process, some participants criticized the materials as in the following comment:

> The program lacks necessary materials such as teacher book, CD of the book, any kind of extra video or audio materials for young learners(T1,T2, T3, T5,T6).

Considering the assessment aspect of ELTP, the opinion of a teacher is prominently important to note here:

> I have difficulty in assessing my students as there are no clear borders in terms of assessment. So not to make them lose their motivation I give high marks (T1).

This comment led the researchers to discuss the assessment types, guidelines, procedures in ELTP in the instrument they developed.

In sum, having been tested, the "Program Evaluation Scale" developed by the researchers was found a well-established instrument with the reliability value of $\alpha = .966$. It consists of 33 items in which there are 5 choices ranging from "strongly disagree" to "strongly agree" for each item. Additionally, it includes 5 sub-dimensions, namely; overall characteristics, objectives, content, teaching /learning process, and assessment.

Discussion and Conclusion

The interpretation of the factor analysis results revealed a five-component PES. As displayed and discussed in the literature review, the program evaluation has various purposes, such as identifying deficiencies and problems, improving the program, making a final judgment about its success or weaknesses and exploring the opinions of stakeholders through a critical stance among others. Therefore, this study is significant as it achieves two of the purposes mentioned above, one of which is covering the opinions of stakeholders while the second one attempts to identify the problematic parts taking all possible angles into consideration such as overall characteristics, objectives, content, teaching/learning process and assessment through both qualitative and quantitative data.

In this respect, the first component -overall characteristics- is essential to realize the general understanding of the program including the main philosophy of it, students' attitudes towards it, and the guidance it provides for teachers. Furthermore, the second component – objectives- deal with the intended educational goals of the program by integrating the suitability and attainability of the objectives based on such variables as learners' age, their cognitive and emotional levels. As Erden (1995) noted, the objectives occupy a remarkable place in ordering and organizing the content and the teaching and learning process during program development process. Therefore, during the planning process, a large scale needs analysis which helps to set clear and attainable objectives should be done.The third component –content- constitutes of the topics and skills covered within the program with a specific focus on learners' readiness level and interests. What to teach is an important aspect of a program, thus the content should attract learners' attention and support the previous knowledge of learners.The fourth component -teaching / learning process- is made up of inclusive of classroom applications, teacher practices, course activities, materials, and learning environment which are considerably important to explore the practices within

classrooms. Therefore, timing, strategies, techniques and methods in order to reach the desired outcomes is highlighted in this process (Demirel, 2006). The fifth component – assessment- is based on the idea of the extent to which our objectives are achieved by learners. This aspect of the program is of crucial importance as it provides necessary information about the deficiencies, strengths and weaknesses of the program. As Cihan and Gürlen support, assessment should be in accordance with the content, objectives and teaching/learning activities (2013).

Implications and Suggestions

Developing a program evaluation scale based on practicing teachers' ideas and beliefs; the study concludes with several methodological and pedagogical implications in an attempt to draw some suggestions for further studies. To serve this purpose, a number of implications as indirect suggestions can be drawn from this present study to guide the further research studies.

With reference to the discussion above, it might be considered that teachers' beliefs are of decisive importance as they are the end-users of the program. In line with this, teachers' beliefs about the reasons and theoretical considerations underlying the program elements are prominently important as they apply the program based on their beliefs. For this reason, it would be fair to collect more detailed data via semi-structured interviews, observations, diaries and field notes from teachers through triangulation as it is necessary to explore or identify any concerns or potential problems within the program.

As researchers we expect that this program evaluation scale as a suitable and reliable data collection tool will be of utmost importance for program evaluators and researchers by shedding light on the details of the program. Further, we anticipate that the findings will also help teachers, program developers, and policy makers to gain insight into the strengths and weaknesses of the program of concern. More importantly, researchers from diverse interests might benefit from the instrument to evaluate various teaching programs by adjusting and modifying it in order to reach efficient data by numerous respondents.

In this study, teachers' opinions were explored while developing the scale; however, teacher educators and prospective teachers as well should have a say regarding the planning, designing and implementation process of the program as they directly use, practice and experience the program.

Acknowledgement

REFERENCES

Adıgüzel, O. C., & Özüdoğru, F. (2014). İlkokul 2. Sınıf İngilizce Öğretim Programına Yönelik Aydınlatıcı Değerlendirme Modeline Dayalı Program Değerlendirme Ölçeği Çalışması. *Trakya Üniversitesi Eğitim Fakültesi Dergisi, 4*(2), 124-136.

Alderson, J. C., & Beretta, A. (1992).*Evaluating Second Language Education*. Cambridge: Cambridge University Press.

Alkan, M. F., & Arslan, M. (2014). Evaluation of the 2nd grade English language curriculum. *Uluslararası Eğitim Programları ve Öğretim Çalışmaları Dergisi, 4*(7), 87-99.

Bayrak, B., & Erden, A. M. (2007). Fen bilgisi öğretim programının değerlendirilmesi. *Kastamonu Education Journal, 15*(1),137-154.

Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development.* Boston: Heinle & Heinle Publishers.

Büyüköztürk, Ş. (2007). *Sosyal bilimler için very analizi el kitabı* [Data analysis handbook for social sciences] (7th ed.). Ankara: Pegem Akademi.

Cihan, T., & Gürlen, E. (2013).Teachers' opinions on the English language curriculum of the 5th grade of primary education. *Anadolu University Journal of Social Sciences*,*13*(1),131-146.

Demircan,Ö. (1988). *Dünden bugüne Türkiye'de yabancı dil*. İstanbul: Remzi Kitabevi.

Demirel, Ö. (2005). Language project of the European council and its application in Turkey. *National   Education   Journal*,   *33*(167).   Retrieved   from http://dhgm.meb.gov.tr/yayimlar/dergiler/Milli_Egitim_Dergisi/167/index3-demirel.htm

Demirel, Ö. (2006). *Kuramdan Uygulamaya Eğitimde Program Geliştirme*. Ankara: Pegem Yayıncılık.

Dinçer, A., Takkaç, M., &Akalın, S. (2010). An Evaluation on English Language Education Process in Turkey from the Viewpoints of University Preparatory School Students. *2nd International Symposium on Sustainable Development*, Paper presented, Sarajevo.

Erden, M. (1995). *Eğitimde program değerlendirme*. Ankara: PegemYayıncılık.

Erkan, M.A. (2009). *İlköğretim Okulları 4. ve 5. Sınıflar Yeni İngilizce Öğretim Programının Uygulanmasında Karşılaşılan Sorunların İncelenmesi* (Yayınlanmamış Yüksek Lisans Tezi). Gaziantep Üniversitesi, Sosyal Bilimler Enstitüsü, Gaziantep.

Gehlbach H, Artino A R, Durning S. (2010). Survey Development Guidance for Medical Education Researchers, Academic Medicine, 85(5), p.925, DOI : 10.1097/ACM.0b013e3181dd3e88.

Gömleksiz, M. N., & Bulut, İ. (2007). Yeni Fen ve Teknoloji Dersi Öğretim Programının Uygulamadaki Etkililiğinin Değerlendirilmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi. 32*, 76-88.

Güneş, T. (2009). *İlköğretim 5. Sınıf İngilizce Dersi Öğretim Programına İlişkin Öğretmen Görüşleri* (Yayınlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

İnam, G. (2009). *İlköğretim okulları 4. Sınıf İngilizce dersi öğretim programının değerlendirilmesine ilişkin öğretmen görüşleri* (Yayınlanmamış yüksek lisans tezi). Adnan Menderes Üniversitesi, Aydın.

İnceçay, G. (2012). Turkey's foreign language policy at primary level: Challenges in practice. *ELT Research Journal*, *1*, 53-62.

Kırkgöz, Y. (2009). Globalization and English language policy in Turkey. *Educational Policy, 23*, 663-684.

Kızıldağ, A. (2009). Teaching English in Turkey: Dialogues with teachers about the challenges in public primary schools. *International Electronic Journal of Elementary Education, 1*(3), 188-201.

Küçük, Ö. (2008). *An Evaluation of English Language Teaching Program at Key Stage I and Opinions of Teachers Regarding the Program* (Published master thesis), Çanakkale Onsekiz Mart University, Çanakkale.

Lynch, B. K. (1996). *Language program evaluation: Theory and practice.* New York, NY: Cambridge University Press.

Mackey, A., & Gass, S. M. (2011). *Research Methods in Second Language Acquisition: A Practical Guide*. Oxford: Blackwell Publishing.

MoNE, (2015). *İlköğretim Kurumları İngilizce Dersi Öğretim Programı*. İstanbul: Milli Eğitim Basımevi.

Mersinligil, G. (2002). *İlköğretim dört ve beşinci sınıflarda uygulanan İngilizce dersi öğretim programının değerlendirilmesi: Adana ili örneği* (Yayınlanmış Doktora Tezi), Fırat Üniversitesi, Elazığ.

Merter, F., Kartal, Ş., & Çağlar, İ. (2012). The Evaluation of New High School English CurriculumIn Terms of Teachers' Opinions, *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 12*(23), 43-58.

Orakcı, Ş. (2012). *İlköğretim 7. Sınıflar için uygulanan 2006 İngilizce öğretim programının öğretmen görüşlerine gore değerlendirilmesi* (Yayınlanmamış yüksek lisans tezi), Gazi Üniversitesi, Ankara.

Rolstad, K., Mahoney,K., & Glas, G.V. (2005). The Big Picture: A Meta-Analysis of Program Effectiveness Research on English Language Learners, *Educational Policy, 19*(4), 572-594. Doi: 10.1177/0895904805278067.

Saldana, J. (2009). *The coding manual for qualitative researchers*. Los Angelas, CA: Sage Publications.

Seçkin, H. (2010). *İlköğretim 4. Sınıf İngilizce dersi öğretim programının değerlendirilmesi* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.

Slavin, R.E. (2008).Perspectives on evidence-based research in education-what works? Issues in synthesizing educational program evaluation. *Educational Researcher, 37*(1), 5-14, https://doi.org/10.3102/0013189X08314117

Soner, O. (2007). Türkiye'de Yabancı Dil Eğitimin Dünü Bugünü. *Öneri, 7*(28), 397-404.

Tok, H., & Arıbaş, S. (2008). Avrupa Birliğine uyum sürecinde yabancı dil öğretimi. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, *9*(15), 205-227.

Yaman, S. (2010). *İlköğretim birinci kademe İngilizce dersi öğretim programının öğretmen görüşlerine gore değerlendirilmesi* (Yayınlanmamış yüksek lisans tezi). Fırat Üniversitesi, Elazığ.

Worthen, B. R., Sanders, J. R.,& Fitzpatrick, J. L. (2004). *Program evaluation: Alternative Approaches and Practical   Guidelines.* Boston: Allyn and Bacon.

Developing a program evaluation scale to investigate perceptions about primary school English language teaching program

81

APPENDIX

The final version of Program Evaluation Scale

		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1.	The program is student-centred.	1	2	3	4	5
2.	It is possible to make learners enjoy English by implementing the program.	1	2	3	4	5
3.	It is possible to make learners develop positive attitudes to English by implementing the program.	1	2	3	4	5
4.	The program guides teachers well.	1	2	3	4	5
5.	In-service training is essential to understand and implement the program.	1	2	3	4	5
6.	The goals are clearly and explicitly stated.	1	2	3	4	5
7.	The goals support and complete each other.	1	2	3	4	5
8.	The goals are suitable for learners' age.	1	2	3	4	5
9.	The goals are suitable for learners' cognitive development.	1	2	3	4	5
10.	The goals are suitable for learners' emotional development.	1	2	3	4	5
11.	The goals of the program are attainable by learners.	1	2	3	4	5
12.	The goals of the program are in accordance with the content.	1	2	3	4	5
13.	The content provides learners with an enjoyable, stress-free learning environment.	1	2	3	4	5
14.	The content attracts the students' attention and curiosity.	1	2	3	4	5
15.	The topics in units support each other.	1	2	3	4	5
16.	Speaking skill is adequately covered in the content.	1	2	3	4	5
17.	Listening skill is adequately covered in the content.	1	2	3	4	5
18.	The number of words suggested to be taught in each unit is inadequate.	1	2	3	4	5
19.	The content is suitable for learners' readiness level.	1	2	3	4	5
20.	The content of the program allows to use different methods and techniques.	1	2	3	4	5
21.	The teaching and learning process is suitable for using an eclectic mix of instructional techniques simultaneously in classroom.	1	2	3	4	5

22.	The teaching techniques suggested in the program are suitable for the level of classroom.	1	2	3	4	5
23.	The classroom activities are designed by taking learner differences into account.	1	2	3	4	5
24.	The classroom activities are suitable to learners' physical development.	1	2	3	4	5
25.	Not only in-class but also out-of class assessment types are used.	1	2	3	4	5
26.	Evaluation fosters learners to self-evaluate themselves.	1	2	3	4	5
27.	Evaluation and assessment is explained in detail in the program.	1	2	3	4	5
28.	Evaluation is able to show whether the goals are achieved by learners.	1	2	3	4	5
29.	Assessment is in accordance with the content.	1	2	3	4	5
30.	Assessment types are in accordance with the goals of the program.	1	2	3	4	5
31.	Portfolio evaluation is useful.	1	2	3	4	5
32.	It is possible to evaluate listening skills in the program.	1	2	3	4	5
33.	It is possible to evaluate speaking skills in the program.	1	2	3	4	5