



A Utilization Focused Evaluation of the Preparatory School of an ELT Program

Research Article

Gulumser EFEOGLU¹ Ferda ILERTEN² Ahmet BASAL³

¹Yildiz Technical University, Faculty of Education, Foreign Language Education, Istanbul, Turkey, ORCID: 0000-0003-2771-4401

²Yildiz Technical University, Faculty of Education, Istanbul, Turkey, ORCID: 0000-0002-7596-9030

³Yildiz Technical University, Faculty of Education, Istanbul, Turkey, ORCID: 0000-0003-4295-4577

To cite this article: Efeoglu, G., Ilerten, F., Basal, A. (2018). A Utilization Focused Evaluation of the Preparatory School of an ELT Program, *International Online Journal of Educational Sciences*, 10(4), 149-163.

ARTICLE INFO

Article History:

Received 24.04.2018

Available online

31.08.2018

ABSTRACT

Despite numerous attempts to evaluate language learning programs, longitudinal studies re-evaluating the effectiveness of the previous evaluation is highly scarce. The purpose of this study is to investigate the preparatory school's program evaluation (focusing particularly on reading, grammar and listening courses) based on the principles of Patton's Utilization Focused Evaluation in an ELT Department at a state university in Turkey. A descriptive mixed method research design is employed for this study. Data were collected via open-ended questionnaire, individual interviews, nominal group technique in two subsequent years. The number of the participants were 38 in total (first year, n=19; second year, n=19). The results of the first evaluation were disseminated to all parties involved (i.e. the students, instructors, policy-makers, and the head of the department). Thus, the second evaluation served for assessment of the first one as well as providing detailed analysis of the new program. Results indicated that all changes were all well accepted by almost all of the participants, particularly by the new-comers, highlighting the effectiveness of re-evaluation of the previous program.

© 2018 IOJES. All rights reserved

Keywords:

English Language Teaching, curriculum evaluation, preparatory school, utilization focused evaluation

Introduction

Evaluation is "an applied inquiry process for collecting and synthesizing evidence that culminates in conclusions about the state of affairs, value, merit, worth, significance, or quality of a program, product,

¹ Corresponding author's address: Yildiz Technical University, Faculty of Education, Foreign Language Education, Istanbul, Turkey
Telephone:
e-mail: gulumserefeoglu@gmail.com
DOI: <https://doi.org/10.15345/iojes.2018.04.008>

person, policy, proposal, or plan" (Fournier, 2005; pp.139-140). In setting up any program, there is a particular set of objectives and at some point, it is necessary to check whether the program meets all these predetermined goals since there is no one *best* educational program without any drawback. In addition, Blomquist (2003) argued that the program evaluation in general provides a wealth of valuable information for program developers regarding the effectiveness of the program itself that would be unavailable before the evaluation.

Program evaluation has only recently emerged as a significant part of academic inquiry. It has emerged from a pressing need to assess the outcomes of educational programs that must be congruent with the assumptions derived from learning theories. For instance, in line with Vygotsky's social constructivist view of learning (1962), Rust, O'Donovan and Price (2005) proposed "a social constructivist assessment process model" in which they focus the interaction of teachers and students sticking to four group-specific criteria-For teachers, these are tutor discussion of criteria, assessment guidance to staff, marking and moderation, assessment design and development of explicit criteria while for students, these are listed as follows: explicit criteria, active engagement with criteria, completion and submission of work, active engagement with feedback. (p. 233)

As for language program evaluations, there were few studies most of which were geared towards providing detailed descriptions of already existing language programs starting from 1960s (Kieyle & Rea-Dickens, 2005). These evaluations were for the purpose of judging the overall effectiveness of a given program and based on less systematic decision making. As the number of evaluations of programs have increased, a need for a more systematic and comprehensive evaluation of the educational programs emerged.

Review of Literature

Evaluation approaches in the 21st century can be classified based on their orientations and themes. Hogan (2007) specified the approaches as objectives, management, consumer, expertise, adversary and participant oriented. From another perspective, Brown (1995) claimed that all evaluation approaches center around three themes: time of evaluation (formative vs. summative), focus of evaluation (process vs. product) and type of data gathered for evaluation (qualitative vs. quantitative). Based on the priorities, it is possible to add new criteria to evaluation or some aspects could be at the forefront of program evaluation, yielding to a variety of evaluation approaches. In the following section, evaluation studies will be elaborated from various perspectives.

Previous Evaluations of Language Learning Programs Worldwide

In this section, we focused on the evaluation studies related to the types of language programs and projects in line with the purposes of the current study. While some of the large-scale studies have been funded by governments held by external evaluators such as Chan (2001) and Kiely and Rea-Dickens (2005), some have been conducted by the internal evaluators to develop the programs in use.

In Chan's (2001) large scale language program evaluation study, the needs and expectations of the university students were compared with those of language teachers and program coordinators at Hong Kong Polytechnic University. The results indicated that both instructors and students agree on improving listening and speaking skills for conferences and seminars, building discipline-specific vocabulary, raising motivation and confidence among students. In a very similar context, Peacock (2009) carried out the internal evaluation of TEFL program at the City University of Hong Kong and the results were similar to Chan's (2001) findings. This study has put forward some suggestions related to the course content and more technology integration. Both studies in question highlight the significance of involving all parties in the evaluation process. A relatively small-scale evaluation was held by an internal evaluator in English Language Institute (henceforth ELI) at the University of Hawai'i at Manoa (UHM) (Yang, 2009).

Additionally, Utilization Focused Evaluation (henceforth UFE) was settled as the evaluation framework. The institute aims to teach English to the international and immigrant students at the university. The graduate students in the Department of Second Language Studies were recruited as teachers in ELI. The evaluation was about the preservice induction practices and training of the ELI teachers. The directors of the department were chosen as intended users to decide on the process and method of the evaluation as one of the main requirements of UFE is the collaboration of the evaluator and the intended users. Interviews were made with ELI administrators, new ELI teachers, experienced ELI teachers and incoming ELI teachers in the following semester. ELI administrators defined the intended outcomes of the pre-semester induction practices, new teachers explained whether their needs and expectations were met in the induction practice, experienced ELI teachers listed what the new teachers should be taught during the practice and incoming ELI teachers stated what they expect from the practice. Based on the analysis of their replies, the emerging common belief was that observation and meetings with the lead teacher and the other ELI teachers are the most helpful practices. Some other essential findings were early-hiring of the teachers for giving them the needed time to become familiar with the class materials and videotaping of some classes as an alternative to classroom observation. The strength of the study is that the findings and suggestions were all presented to the administrators. In this type of evaluation, the internal evaluator, an ELI teacher, caused a change in the induction practice.

All the aforementioned studies exemplify distinct worldwide foreign language teaching contexts which benefit from program evaluation for a variety of purposes. Despite the differences in terms of the approach, theoretical framework and scope of the evaluations, it is evident that the results of evaluations have always been affirmative.

Previous Evaluations of Language Learning Programs in Turkey

Considering the evaluation studies conducted in Turkey, there has been an increasing amount of research differing to a great extent in terms of the framework and the context (e.g.; Gerede, 2003; Gullu, 2007; Karatas, 2007; Kesli-Dollar, Tekiner-Tolu, Doyran, 2014; Mede, 2012; Muslu, 2007; Ors, 2006; Ozkanal, 2009; Sarı, 2003; Tiryaki, 2009; Topcu, 2005; Yılmaz, 2009). This study only examines the evaluation studies conducted on the English preparatory programs of universities in Turkey. Gerede (2003) conducted a study comparing former and present curricula of Intensive English Program at Anadolu University. It was found that two curricula differ in terms of meeting students' needs.

In another study, Topcu (2005), Şahin (2006) and Muslu (2007) focused on the program evaluation of School of Foreign Languages at a university in Turkey. In their studies, interviews and surveys were conducted to determine possible weaknesses and strengths of the program. The results revealed that although differences can be found in terms of student and instructor' suggestions, they all contribute to the improvement of the already existing program. The aforementioned studies targeted program evaluation of language schools involving students with different job-orientations at the prep schools.

The current study was conducted in a different context, a preparatory school, where prospective English teachers are trained to improve their English knowledge in terms of both productive and receptive skills. In a similar study, Mede (2012) developed the English Preparatory program for undergraduate students in English Language Teaching Department, at Yeditepe University and evaluated it. The students were asked to fill in a survey and take part in semi-structured interviews. The first phase of the study was to pin down the needs and expectations of participants while the second one was geared towards evaluation of this program. Results revealed sound implications for both design and the evaluation of the program. For instance, based on the findings, students stated that activities that would foster their note-taking and oral summary skills should be employed much more frequently.

Utilization Focused Program Evaluation

The current study is the first study at a Turkish context which adopted Patton's Utilization Focused Program Evaluation Model to undertake the evaluation of a preparatory school's program of an ELT department at a well-known state university in Turkey. Utilization focused evaluation (henceforth UFE) is "evaluation done for and with specific intended users for specific intended uses" (Patton, 2008, p. 37). The main concern of UFE is how real people "apply evaluation findings in their actual use" and how they "experience" the evaluation process (Patton, 2008, p. 37). UFE does not favor any model, method or theory; intended users are encouraged to select the most suitable content, model, theory for their own settings instead. UFE is therefore a broad concept with multiple uses; it can be applied to any evaluative purpose (formative, summative and developmental), with quantitative, qualitative or mixed type of data and with any kind of focus (processes, outcomes, impacts, costs etc.). This wide range of variety lets the intended users feel the ownership of the process, encouraging active involvement of the participants during and after the evaluation process. Unlike traditional methods, the role of the evaluator is multifaceted; facilitator, negotiator, coordinator and collaborator. The decision-making body is the primary intended users (Yang, 2009).

According to Patton (1997), UFE has seven steps. Initially, the interests and commitments of the stakeholders are analyzed. The second step is to choose the intended users. Then, the evaluator and the intended users decide on the focus of the evaluation, the theory, research questions and discuss on the potential findings of the evaluation. This step is the building block of UFE. The fourth step is to decide on the data collection process. The evaluator and the intended users determine the instruments, methods and the procedure to follow. The evaluators and the intended users also interpret the findings together. Dissemination is the sixth step; dissemination and reporting are done both for the intended users who volunteer to benefit from the evaluation (planned utilization) and for unintended users. Finally, the evaluations are elaborated in a detailed manner.

There are few applications of UFE in language programs. Norris (2004) suggested a UFE framework for the validity of assessment in US college foreign language education. Yang (2009) conducted an evaluation of teacher induction practices in the English language program of a US university. Many of language program evaluations in Turkey aimed to propose a brand-new design that would meet the needs of the program based on need analysis. However, the current study aims to achieve meta-evaluation of the already existing program and highlight the ownership of the process and actual use of the outcomes. More precisely, this study aims to evaluate the effectiveness of a preparatory program in an ELT department longitudinally within UFE framework. Hence, changes in the program based on first evaluation will be re-evaluated in the subsequent year with the inclusion of and contributions from all involving parties. The following research questions guided the study:

- 1) What are opinions of participants with regard to the three courses namely reading, grammar and listening in the preparatory program of ELT department at a state university in Istanbul?
- 2) What are their suggestions in order to improve the preparatory program based on listening, reading and grammar courses?
- 3) How do participants' reactions differ regarding the effectiveness of courses in two successive years?
- 4) Does the evaluation of the first year within UFE framework have any impact on the following years' program?

Method

A descriptive mixed method research design (Johnson & Onwuegbuzie, 2004) was constructed by combining both questionnaires and following qualitative interviews that emphasize the understanding and interpretation of students' perceptions and experiences (Yin, 2011). According to Creswell and Plano Clark

(2007), mixed method "...focuses on collecting, analyzing, and mixing both quantitative and qualitative data in a study or series of studies and its central premise is that the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone" (p. 5). In this vein, the research design was adopted to be able to answer a wide range of research questions and for the corroboration of the findings. Thus, the questionnaire including open-ended questions, individual interview and nominal group technique were employed to gather information, providing valuable information to the final interpretation of the findings.

Participants

Participants were recruited from a preparatory school of an English Language Teaching Department at a state university in Istanbul. At the beginning of the university education, ELT students who score lower than 60 (out of 100) in the upper-intermediate level proficiency exam are required to study at the preparatory class for one year. The aim of the preparatory program is to develop students' language skills in English and prepare them for the academic and professional use of language.

Convenience sampling was used for selecting the participants of this study; participants were selected because of their availability and accessibility to conduct this study. A total of 38 students participated in this study; in the first year of evaluation, the number of students was 25; however, 19 of them volunteered to take part in the study. In the following year, out of 32 students 19 of them participated. Participants were informed that the questionnaire aims to pin down strengths and weaknesses of the program. It was not obligatory to take part in the study and participants were not graded based on their participation. Nonetheless, they were informed that their feedback was invaluable since they were the first two groups of students in the preparatory program.

Instrument(s)

Data of the study were collected via an open-ended questionnaire, individual interviews and nominal group technique. The questionnaire consisted of four open-ended questions: The first of which was about the teaching technique that the instructor preferred most frequently in order to compare participants' views on the technique that they believed to be the most effective for that particular course and the technique that the instructor employed most. The next two questions were about the weaknesses and the strengths of the courses. The last question asked for further suggestions of the participants was related to the program in question.

Individual interviews were conducted to elicit personal opinions that might not have been expressed in the group technique and to compare the individual views with the other elicitation tasks in the study (Taylor & Bogdan, 1998). Thirty-five (16 in the 1st evaluation + 19 in the 2nd evaluation years) prep-year students were interviewed. Their responses were audio-recorded, transcribed and listed for each course. The questions in the semi-structured interview were about their satisfaction from the program, strengths/weaknesses of each course and suggestions about the program.

Nominal Group Technique was the third data collection tool. According to MacPhail and Kirk (2001), unlike focus group interviews, brainstorming or the Delphi technique, Nominal Group Technique (henceforth NGT) aims at getting the opinions of individuals within the group. In focus group interviews, discussions are held within the group on a specific topic, however the limitation of this technique is that some vocal people in the group generally dominate and direct the flow of the discussion. In NGT, participants, initially, write their ideas independently in the same setting and they do not express their opinions verbally. Some advantages of the first-stage of the NGT are that participation is balanced among the interviewees and the researchers could avoid distractions such as note-taking and they do not have much influence in this process (MacPhail & Kirk, 2001). Finally, these independent ideas are gathered and discussed on a discussion stage and a mathematical voting procedure is incorporated (Delbecq et al, 1975). Moreover; Kiely (2001) notes that nominal group

technique let participants feel more responsible for their own learning process since they have a say on the evaluation of the program.

Following the procedure of NGT, the researcher met the students in the classroom, and asked them to write the main problems about Listening, Reading and Grammar classes, 3 main issues for each class. Since the students had conflicts with the instructor of the writing and speaking lessons, these courses were excluded from the study to avoid bias. Thirty-seven (18 in the 1st year + 19 in the 2nd evaluation year) students participated in this phase of the study. The individual writing period took 25 minutes. Then, the researchers chose a group discussion leader among the students and left the classroom to minimize the researchers' influence. The role of the discussion leader was to let everybody state what they wrote on their papers and they decided on the 3 most significant problems they have in those classes by voting. The leader wrote down the top-rated issues in each class and submitted them to the researchers. Finally, with the instructor of the grammar class, the students discussed the problems they stated for the class.

Data Collection and Analysis

Prior to collection data, participants were asked to sign the consent form and were informed that their participation in the study was voluntary. The participants were also notified that they could withdraw from the study at any time and that their responses were confidential. Pseudonyms were used to protect participants' identities and ensure confidentiality. They were told that the aim of the current study was to get feedback from them to make the preparatory program more effective in line with the needs and expectations of students. Then, semi-structured interviews were conducted individually. Each participant was interviewed over the single session lasting between 30 and 45 minutes. Interviews were video-taped with participants' permission. Lastly, NGT was employed to enable participants discuss and determine the priority of the problems that they detected with respect to the preparatory program itself. The study was conducted in 2013-2014 and 2014-2015 academic years. The same procedure was repeated in the second academic year with the new participants.

For objectivity, open-ended questionnaire was administered by the instructor whereas interviews were conducted by the research assistant. The researchers employed individually a peer-review strategy and independently read and analyzed the interview transcripts and compiled using the analysis and interpretation of the data to establish credibility, reliability, and confirmability.

Analyzing quantitative questionnaire, data exploration involved using descriptive technique to determine general trends in the data. The descriptive measures, theoretical analyses, and qualitative techniques used in analyzing the study data for answering specific research questions in relation questionnaire items, than followed more detailed explanation of analysis qualitative interviews and a brief narrative of the NGT discussion. Thus, data from three different sources were analyzed qualitatively after transcription and categorization. The qualitative analysis involved highlighting the recurrent patterns found in data. For interviews, the checklist matrix proposed by Miles and Huberman (1994) was employed while for both the questionnaire and nominal group technique, frequency count of the answers was taken into account. In addition to researchers, two external colleagues read and analyzed the collected data, checking for evidence of the patterns, assertions, and verifying them with the original data that helped to assess and strengthen the accuracy of the data.

Findings

Student Questionnaire

The questionnaire was based on 1) the most common techniques that the instructors use in the class, 2) the strengths, 3) weaknesses of the specific course and 4) the suggestions of the students. All questions were

open ended, and the students were allowed to write multiple statements for each section. The percentages were calculated through the frequency of the answers.

Table 1. Questionnaire results for Listening

Issues	1 st Evaluation	2 nd Evaluation
The most commonly-used technique(s)	Computer Assisted Teaching (48%)	Computer Assisted Teaching (52%)
Strength	Improvement in listening and inferencing skills (44%)	Improvement in listening skill in general (27%)
Weakness	Technical inadequacy (64%)	Mismatch in the proficiency level (19%)
Suggestion	More technical support (30%)	More technical support (23%)

*Each and every item in this table indicates the most commonly stated/highlighted issue

The questionnaire for the Listening course revealed that, the most common techniques have not changed in these years. The most common method is computer assisted teaching (1st evaluation 48%; 2nd evaluation, 52% respectively) as indicated in Table 1, followed by question-answer (23%, 24% respectively), individualized education (13%, 8%), brainstorming (13% appeared in the first evaluation only), discussion & criticism (3% - appeared in the first evaluation only), watching videos (8% - appeared in the second evaluation only) and others (8% - appeared in the second evaluation only) . The main strength of the course is that it improves their listening skills. Using the computer during the class hours (19%, 20%), student centeredness (13% appeared in the first evaluation only) and interesting topics (13% appeared in the second evaluation only) were other strengths of the course stated in addition to other factors (25%, 40%). In both years, one of the challenges of the Listening course has been the lack of a computer lab. The labs of other departments were used from time to time and the students had the most productive classes in these labs. However, most classes were held in regular classrooms and causing an acoustic problem which led to hearing difficulties. For this reason, inaccessibility of a lab has been the main weakness of the course (64%, 14%). In the second evaluation year, students also complained about the difficulty of the listening tasks (19%), the limited number of listening tasks in classes (14%), being exposed to only American accent (5%) and others (48%). They believe they should have the chance to listen to various audios from various accents. Mostly, students suggest to have a computer lab (30%, 23%), more listening tasks (25%, 14%) and topic variety.

Table 2. Questionnaire results for Reading

Issues	1 st Evaluation	2 nd Evaluation
Commonly-used technique(s)	Question-Answer (41%)	Question-Answer (48%)
Strength	Improvement in comprehension skills (20%)	Improvement in vocabulary (36%)
Weakness	Uninteresting texts (36%)	Uninteresting texts (35%)
Suggestion	More entertaining texts	More entertaining texts

*Each and every item in this table indicates the most commonly stated/highlighted issue

The questionnaire for the Reading course showed that the most common application in the classroom is question- answer (41%, 48% respectively), brainstorming (16%, 24%), individualized education (9%, 7%) and others (34%, 21%). It was observed that the amount of group work has decreased in the following year (19% to 3%). Students believed that they improved in reading skills: comprehension, vocabulary, grammar, pace etc. (20%, 36%). Also, they are content with the interactive and a student-centered course (54%, 48%). There are also some other factors (26%, 16%) such as teacher's being very helpful (7% appeared in the first evaluation only) and reading course fostering other skills as well (4% appeared in the second evaluation only). However,

they found the reading texts boring in both evaluation years (36% and 35%, respectively). This may be due to the simplicity (9%) of the reading texts in the first year, and the difficulty (35%) of them in the second year. While the students were dissatisfied with the difficulty of the exam compared to the classroom implication in the first year (9%), in the second year their dissatisfaction was about inability to improve several skills such as creative thinking, evaluation and problem-solving skills (12%). Finally, the suggestions of the students were based on material selection. They would like to work on enjoyable, interesting, authentic, visual and technology-based texts (7%, 22%).

Table 3. Questionnaire results for Grammar

Issues	1 st Evaluation	2 nd Evaluation
Commonly-used technique(s)	Question-Answer (59%)	Question-Answer (50%)
Strength	Student-centered (21%)	Contribution to their current knowledge (36%)
Weakness	Repetition of the similar topics over the years (38%)	Monotonous (33%)
Suggestion	Variety in the activity type (25%)	Stimulating tasks (20%)

*Each and every item in this table indicates the most commonly stated/highlighted issue

The questionnaire for the Grammar course showed that the most common technique in the classroom is question-answer (59%, 50%) in both evaluation years. The percentage of brainstorming (19%, 5%) and group work (19%, 5%), has decreased while individual practices increased (4%, 32%). Another technique noted in the second evaluation was textbook exercises (9%). The students found the course student centered (21%), informative (14%+14%) and inclusionary (14%) in the first year of evaluation. In the second year, they found the course informative as well (36%, 21%, and 14%). Strikingly, whereas the textbook was one of the main weaknesses (19%) in the first year, in the second year they found the book useful and detailed (14%). It is obvious that the instructor realized the weakness book in the first year and used a different one for the following year. Nevertheless, Grammar course is regarded as a boring class (13%, 33%). The students clarified that the class is just a revision, because at high school they had to study for the university entrance exam which requires an upper-intermediate/advanced level of grammatical knowledge. They demand to have an enjoyable, interesting, activity based, student centered grammar class (62%, 27%). In the second evaluation there were some other factors (53%) related to textbook (14%) and students (14%).

Individual Interviews

Interviews with participants were audiotaped in two subsequent years' preparatory program evaluation. The same questions were asked to participants to pin down discrepancies of curricula employed in two subsequent years. Each audio-recording was transcribed and categorized according to the checklist matrix proposed by Miles and Huberman (1994). In the 1st evaluation, out of 19 students, 11 of them stated that they are not satisfied or partly satisfied with the prep-year program. Three of them expressed that they are satisfied, and two of them did not answer. On the other hand, in the 2nd evaluation, majority of the participants (10 out of 19) noted that they are content with the curriculum while five of them reported the reverse.

The next step in the interview is to evaluate three courses respectively. According to the 1st year's evaluation, the main weakness of the grammar course is that the textbook is below their proficiency levels; 44% of the answers were in that way. Practicing (19%) and wasting time (13%) on the subjects which are already known were the next two issues they stated in the interviews. Sticking to the textbook all the time was another weakness (6%) along with others (18%). Contrary to 1st evaluation, 47% of participants noted that they do not observe any weaknesses whereas 26% of them reported that the course itself is boring in addition to

other weaknesses (27%). It is significant to note that the most frequently stated problem of the 1st evaluation was eliminated in the 2nd evaluation, which shows that precautions taken based on the UFE of the 1st year had a positive impact on participants' perceptions.

In terms of strengths, in the 1st evaluation 32% of the answers about the strengths of the course were that the repetition in the class let them remember the topics they forgot. Their good communication with the teacher followed it with 27%. Besides, online supplementary materials (13%) were noted to be a strength in addition to others (26%). In the 2nd evaluation, the change in the course book choice had a positive effect evidently. An overwhelming majority stated that they are content with the book (42%). The second most frequent strength was the contribution to their overall grammar knowledge (33%). Others were repetitive course content (17%) and activities in class (7%). This has already been obvious in the 1st evaluation in which they claimed the course content helps them to recall what they learned before. Thus, this triangulation of data proves that one of the major strength of grammar course within preparatory program is to help learners to restructure their current English grammar knowledge.

The participants were also asked for their suggestions to enhance the effectiveness of this course. They stated that the expectations of the students should be investigated at the beginning of the year, an advanced level textbook should be used, individual needs should be considered, more time should be devoted to analysis and discussions about the exceptions in the structures, more web-based exercises should be done, and internet and technology should be integrated within the course and the course should be more fun. In the 2nd evaluation, they stated that variety in the activity type helped them to have more fun.

As for the Reading course, they did not find the subjects of the texts interesting (22%). In the 2nd evaluation it is evident that the same problem persists. 53% of the participants noted that the reading texts were boring in addition to other problems (47%). Hence, it is obvious that changes made after the 1st evaluation are not satisfactory. On the other hand, in terms of the strengths, 53% of the answers showed that their inferencing, speed-reading, summarizing, comprehension skills improved and unlike the textbook of the Grammar course they are pleased with the level of the Reading course textbook in the 1st evaluation. Similarly, in 2nd evaluation participants claimed that their reading skills improved along with their lexical knowledge (37%) in addition to other factors such as text choice (41%) and techniques (22%). Their suggestions regarding the enhancement of this course were about making the class more fun and having more enjoyable and engaging reading texts. They also suggested reading a novel in a month and devoting s time in the class to read and discuss those novels together.

Their last evaluation was about the Listening course. In the 1st evaluation, 57% of the answers about the weaknesses of the course were related to the lack of computers for each student. The participants had problems in hearing the listening material when played from the tape recorder. The same problem was evident even in the 2nd evaluation. However, there was a dramatic drop in the percentage since in the second-year learners were directed to make use of listening sources in their spare time. In addition, from time to time Listening course was conducted in a computer lab. Thus, even if the condition was more or less the same technically, the reaction to this situation lessened.

In terms of the strengths, 70% of the answers showed an obvious improvement in listening skill in the 1st evaluation. Others were watching videos on Youtube (10%), teacher's being very helpful (10%) and being exposed to different accents (10%). In the 2nd evaluation they highlighted the sources of their contentment. The reason for this contentment was related with the frequency and variety of the activities and the attractiveness of the course content. Their having more chances to practice (29%), materials being more interesting (29%), using computer labs (14%), and accessing the online materials from home (14%) and lastly developing strategies for listening different texts (14%) were noted other strengths of the course. For a more effective

course, they suggested watching films, TV series, music videos in the class and having more guests in class speaking different accents of English.

Table 4. Interview results for courses (Grammar & Reading & Listening)

Issues	Grammar	Reading	Listening
1st Evaluation Weakness	Textbooks (44%) Repetition (32%)	Uninteresting topics (22%) Improvement in reading skill (53%)	Technical inadequacies (57%) Improvement in listening skill (70%)
1st Evaluation Strength			
2nd Evaluation Weakness	No weaknesses (47%)	Uninteresting topics (53%) Improvement in lexical knowledge (37%)	Mismatch lis. texts and students' current level of proficiency (33%)
2nd Evaluation Strength	Textbook (42%)		Practicing more (29%)

**Each and every item in this table indicates the most commonly stated/highlighted issue

All in all, individual interviews revealed the weaknesses and strengths of grammar, reading and listening courses clearly. Additionally, the subsequent evaluations highlighted the developmental pattern of preparatory program curriculum since each evaluation functioned as a feedback provider for the upcoming one. However, as it is obviously supported by UFE, feedback should be used to modify the curriculum so that it serves its purpose. Hence, it is significant to disseminate these results to all stakeholders.

Nominal Group Technique (NGT)

In the NGT, firstly each participant wrote the main three problems in the Listening, Reading and Grammar courses individually. Later, a group leader held a discussion and they formed a single paper including the discussion results based on the problems. The discussion leader submitted the paper to the researcher. Finally, the researcher negotiated the issues of each course with the students. The following table shows the main discussion issues of the courses.

Table 5. NGT Results: Weaknesses & Strengths

	1 st Evaluation			2 nd Evaluation		
	Listening	Reading	Grammar	Listening	Reading	Grammar
W	Computer Lab/ Technical support	Uninteresting reading texts	Simple books	Computer Lab/Technical support	Imbalance in the difficulty of the exams and in class activities	Sticking to the text book
S	Balance in the difficulty of the exams and in class activities	Interactive activities	Supplementary materials should be used	Textbook should be changed	More vocabulary should be taught	It should be geared towards the proficiency exam

For the listening course, the same problem persisted in the second year. The quality of the listening course depends on the accessibility to technological materials such as computer labs, headphones for each student, internet access and online components. In the second evaluation, the students also demanded having the tests in the labs. Besides, in the first year's evaluation they wanted to hear various accents, since they constantly listened to a British or American accent. The participants knew their interactions will not be limited to native speakers of English, but they will have contacts with the speakers of English as a second or foreign language too. In the second year's evaluation, although this issue was not stated as the most important

problem in the NGT, the questionnaire showed the opposite. Regarding the assessment in the first year, there seems to be an imbalance in the difficulty of the exam and class implementations. Finally, in the second year's evaluation, the students complained about using the textbook, because they mostly believed that the textbook detracts them from real life and authentic materials/internet access could compensate for the textbook.

The drawback in two years is not similar for the Reading course. Whereas, the participants' criticisms were based on the activities and materials in the first year, in the second year the main problem was about the implementation of the course in general. For instance, the participants complained about the level and the content of the texts in the first year. On the other hand, they questioned the equality in assessment and the benefits of the course in the second year. The participants believed that this course did not prepare them for the proficiency exam that they are required to take to continue their university education in the freshman class. They also would like to learn more vocabulary within the course.

In the first evaluation year of the grammar course, the level of the textbook was below the proficiency levels of most students. Therefore, they wanted to use a book which could develop their skills in grammar. They also suggested using supplementary materials to support the course. In the second year, an advanced level textbook was selected by the teacher. Although most of the students were satisfied with the level of the new book, some found it too detailed and monotonous. The book was based on deductive explanations and drills. They wanted to have various interactive activities for the class instead of sticking to the textbook. Finally, just as the reading course, they believed that the course was not adequate to prepare them for the proficiency exam.

Discussion & Conclusion

Triangulation of data via different instruments revealed similar course specific problems for each course evaluated in two subsequent years. Furthermore, evaluation of evaluation was also achieved via the inclusion of 2nd evaluation in the successive year. Thus, the researchers had the opportunity to observe participants' reactions to the precautions taken against the problems emerged in the 1st evaluation. For instance, in the case of grammar, the change in the course book was well-approved by the participants of the second year since they noted that they are content with the course book. This result supports Yang (2009)'s claim that when an evaluation is conducted internally, the outcomes are not regarded as a threat and necessary changes can be made. Nevertheless, there are persistent problems that appear in both evaluations. To illustrate, technical inadequacy of the listening course has been evident from almost all data collection sources in both evaluations. Hence, an immediate solution to such persistent problems should be introduced.

As for the listening course, the results of all data collection tools and both evaluations signal that technical inadequacy is the main problem that participants highlighted as they are highly familiar with advancements in technology. Thus, they seek ways of benefiting from these advancements in their courses as well. In brief, their call for technological facilities in listening course is understandable and the curriculum and physical conditions should be renewed in line with participants' suggestions.

In terms of reading, it seems the majority agreed that the course contributes to their English reading proficiency in both years' evaluations. However, in the 1st evaluation, participants reported that courses are boring since they deal with too many uninteresting texts. Hence, variety in the text content and the activity type may contribute to their motivation to take part in reading courses. Moreover, in both evaluations participants noted that there was no balance between the texts covered in class time and those in the exam. Thus, this stands as one of the persistent problems of preparatory program curriculum. To overcome this problem, the instructor may prepare mock-exams so that participants get used to the type of questions and text that they may encounter in the exam. Or at least, the exam should consist of text types and subjects covered in the class.

There were five courses in the prep program, yet this study focuses on only three of them namely reading, grammar, and listening excluding speaking and writing. The reason for this was the fact that participants had reported severe communication problems with the instructor and their views on the courses were highly biased in the first year. In the successive year, the evaluation for the same set of courses was conducted to compare and contrast the influence of the first evaluation on the program of the following year.

Lastly, participants noted hardly any problems with respect to the grammar course in both years' evaluations. The main problem of the 1st evaluation was the proficiency level of the textbook which was clearly below their proficiency level. Thus, no text book was used in the subsequent year. Instead, the instructor prepared a course pack. Besides, it was clear that participants did not find the content challenging enough, thus their motivation was low in 1st evaluation. In line with this, the course content was modified by adding some challenging structures such as ambiguous sentences. All these solutions have had their positive impact on 2nd evaluation, which is evident with the individual interviews revealing no weaknesses by 46% of the answers. In brief, it is clear that participants of 2nd evaluation are more content with the grammar course of preparatory program curriculum than those of 1st one.

In conclusion, this study sets out to provide a detailed analysis of a preparatory school's program evaluation (focusing particularly on reading, grammar and listening courses) based on the principles of Patton's Utilization Focused Evaluation in an ELT Department at a state university in Turkey. In this context, the results of this study have revealed sound implications for courses under evaluation in line with UFE process. Overall, some course specific problems are evident in all data collection tools in two subsequent years' evaluation, which were remedied in the subsequent year mostly. Similar to Peacock's (2009) findings, each course content required distinct sorts of remedies in these successive years.

Moreover, as it was the case in Mede's (2012) study, taking students' needs and expectations into consideration in the design of English language preparatory program in the second year provided invaluable contributions. The present findings also seem to be consistent with Yang's study (2009) which found that some changes implemented immediately after the first evaluation have been effective in remedying the problems highlighted by the first group of participants. Thus, both studies (i.e. Yang's (2009) and the current study) indicate that UFE might be a satisfactory tool in language program evaluation as well.

As this study was designed in line with the requirements of UFE, the to-do-list for the renewal (modification) of the program has emerged based on participants' suggestions. All parties involved in evaluation got feedback and benefited from the evaluation conforming to their roles in the program. As Patton (2008) noted, a number of evaluations fail to fulfill the dissemination and the follow-up stages to trace the progress in a specific program. It is obvious that the follow-up evaluation conducted in the successive year compensated such deficiencies.

Limitations and Further Study

Although the results were conclusive, this study is not without its limitations. Initially, it could have been better if data were gathered by the instructors as well as their perspective would enrich the analysis of the program in general. Moreover, the scope of the current study was limited to three courses out of five courses offered in the preparatory program due to problems related to the instructor of the course. Additionally, even though the researchers tried to eliminate it by asking the same questions in the interview with a different researcher, an arguable weakness of the study may be the Hawthorne Effect since the participants of the study were aware of being observed by one of the researchers. Notwithstanding these limitations, the study suggests that re-evaluation of language learning programs based on the prior evaluation may provide new windows of opportunities for improving the language learning programs in question.

Further studies may base their investigation on the needs analysis first as it is the case in Mede's (2012) study and then apply UFE framework to language program. It could be even better to involve some of the stakeholders in the evaluation process such as the head of the department and/or the dean of the faculty or a representative from the policy makers from the Council of Higher Education or Ministry of Education.

Acknowledgement

We would like to express our gratitude to Prof. Dr. Yasemin Bayyurt who introduced UFE framework in a detailed manner to us.

REFERENCES

- Blomquist, J. (2003). *Impact Evaluation of Social Programs: A Policy Perspective*. Retrieved from <http://siteresources.worldbank.org> July 25, 2017.
- Brown, J. D. (1995). *The elements of language curriculum*. Boston: Heinle and Heinle.
- Chan, V. (2001). Determining students' language needs in a tertiary setting. *English Teaching Forum July*, 16-27.
- Creswell, J. W., & Clark, V. L. P. (2007). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA Sage Publications.
- Delbecq, A. L., Van de Ven, A. H., & Gustafson, D. H. (1975). *Group techniques for program planning*. Glenview, IL: Scott, Foresman, and Co.
- Fournier, D. M. (2005) 'Evaluation', in Mathison, S. (Ed.) *Encyclopedia of Evaluation*. Thousand Oaks, Sage.
- Gerede, D. (2005). *A curriculum evaluation through needs analysis: Perceptions of intensive English program graduates at Anadolu University*. (Unpublished master's thesis). Anadolu University, Eskişehir.
- Güllü, S. A. (2007). *An evaluation of English program at Kozan Vocational School of Çukurova University: Students' point of view*. (Unpublished master's thesis). Çukurova University, Adana.
- Hogan, R. L. (2007). The historical development of program evaluation: Exploring past and present. *Online Journal for Workforce Education and Development*, 2(4), 1-14.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7), 14-26.
- Karataş, H. (2007). *Yıldız Teknik Üniversitesi modern diller bölümü İngilizce II dersi öğretim programının öğretmen ve öğrenci görüşlerine göre bağlam, girdi, süreç ve ürün (CIPP) modeli ile değerlendirilmesi*. (Unpublished Master's Thesis). Yıldız Technical University.
- Keşli-Dollar, Y., Tekiner-Tolu, A., Doyran, F. (2014). Evaluating a graduate program of English Language Teacher Education. *Turkish Online Journal of Qualitative Inquiry*, 5 (2), 1-10.
- Kiely, R. (2001) Classroom evaluation - values, interests and teacher development. *Language Teaching Research*, 5 (3), 241-261.
- Kieyle, R., & Rea-Dickens, P. (2005). *Program Evaluation in Language Education*. Basingstoke: Palgrave Macmillan.
- MacPhail, A., & Kirk, D. (2001). Coaching for Teachers: an evaluation of the programme in Leicestershire, *British Journal of Teaching Physical Education*, 32 (2), 45-47.
- Mede, E. (2012). *Design and Evaluation of a Language Preparatory Program at an English Medium University in an EFL Setting: A case study*. (Unpublished doctoral dissertation). Yeditepe University, Istanbul.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage. Thousand Oaks, CA.
- Muşlu, M. (2007). *Formative evaluation of a process-genre writing curriculum at Anadolu university school of foreign language*. (Unpublished Master's Thesis). Anadolu University, Eskişehir.
- Norris, J. M. (2004). *Validity evaluation in foreign language assessment*. (Unpublished doctoral dissertation). Honolulu, HI.
- Örs, M. (2006). *An analysis of the preparatory students' attitudes towards the appropriateness of the preparatory school program at the University of Gaziantep*. (Unpublished Master's Thesis). University of Gaziantep.
- Özkanal, Ü. (2009). *The Evaluation of English preparatory program of Eskisehir Osmangazi University Foreign Languages Department and a model proposal*. (Unpublished doctoral thesis). Anadolu University, Eskişehir.
- Patton, M. Q. (1997). Toward distinguishing empowerment evaluation and placing it in a larger context. *Evaluation Practice*, 18(2), 147-163.
- Patton, M. Q. (2008). *Utilization-focused Evaluation: Theory and Practice*. Cambridge: Cambridge University.
- Peacock, M. (2009). The evaluation of foreign-language-teacher education programmes. *Language Teaching Research*, 13(3), 259-278.

- Rust, C., O'Donovan, B., & Price, M. (2005). A social constructivist assessment process model: how the research literature shows us this could be the best practice. *Assessment & Evaluation in Higher Education*, 30(3), 231-240.
- Sarı, R. (2003). *A suggested English language teaching program for Gülhane Military Medical Academy*. (Unpublished master's thesis). Middle East Technical University, Ankara.
- Şahin, V. (2006). *Evaluation of the in-service teacher training program "the certificate for teachers of English" at the Middle East Technical University school of foreign languages*. (Unpublished Master Thesis). Middle East Technical University, Ankara.
- Tiryaki, S. A. (2009). *Ankara Üniversitesi Yabancı Diller Yüksekokulu İngilizce Hazırlık Programı Hakkında Program Koordinatörü, Okutman ve Öğrenci Görüşleri* (Unpublished MA Thesis). Hacettepe University, Ankara.
- Topçu, T. (2005). *An investigation of the effectiveness of the theme-based curriculum in the 2003-2004 academic year at the department of basic English at Middle East Technical University*. (Unpublished Master Thesis). Middle East Technical University, Ankara.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge MA: MIT Press.
- Yang, W. (2009). Evaluation of teacher induction practices in a US university English language program: towards useful evaluation. *Language Teaching Research*, 13(1), 77-98.
- Yılmaz, F. (2009). English language needs analysis of university students at a voluntary program. *Sosyal Bilimler Araştırma Dergisi*, 1, 148-166.
- Yin, R. K. (2011). *Qualitative Research from Start to Finish*. New York, NY: Guilford Press.