

A CEFR BASED EVALUATION OF B1+ LEVEL PREPARATORY PROGRAM AT A TURKISH STATE UNIVERSITY: THE APPLICATION OF THE FOREIGN LANGUAGE SKILLS SCALE

Abstract: Despite the centrality of the Common European Framework of Reference for Languages (CEFR) in language teaching and assessment, studies investigating its learning outcomes in language program evaluations are quite scarce. This paper aims to evaluate the effectiveness of a modular English preparatory school program through the Foreign Language Skills Scale. The research sample consists of 357 preparatory school students having attained B1+ level of proficiency in the program. The results revealed that while the language program designed to be in line with the CEFR guidelines, in general, serves for the needs of the students, there are still some discrepancies between the learning outcomes of B1+ and opinions of students about their competencies, especially in listening sub-skills.

Duruk, Eda PhD
Assistant Professor
English Language Teaching
Department
Pamukkale University
Turkey
E-mail: easlan@pau.edu.tr
ORCID: 0000-0001-8564-2456

Keywords: Common European Framework, Foreign Language Skills Scale, Language Program Evaluation

INTRODUCTION

Program evaluation is the process of gathering information for judging decisions alternatives (Stufflebeam, 1971, p.4) to make new decisions or further revisions about the program. Hopkins (1989, p.14) defines it as a systematic description of education objectives or assessment of their merit or worth.

The purpose of program evaluation can be twofold; it can either improve the program, which is defined as formative evaluation, or reflect if a program should be continued, which is known as summative evaluation. The former one, as Scriven (1991) indicates, is used to gather and share information for program improvement by identifying problematic aspects whereas the latter one is applied at the end of a program to get information about program's success (Worthen, 1990). Scriven (1991) and Worthen, (1990) pointed out that program evaluation should be conducted regularly by using one of these methods, identifying whether students are meeting the learning objectives and leading to the necessary changes. In this respect, program evaluation in education is both a central and challenging task, and foreign language instruction is no exception (Lynch, 1990).

Once intensive foreign language instruction into the Turkish education system was introduced, a number of studies have attempted to evaluate language teaching programs (Erdem, 1999; Gerede, 2005; Topçu, 2005; Güllü, 2007; Karataş, 2007; Üstünlüoğlu et al, 2012). The common aim of these studies concerning language program evaluations is to investigate whether there is a match between what is desired for the program and the actual state of the program along with learners' level of skills and knowledge (Üstünlüoğlu et al, 2012). In the Turkish context, the demand for English in higher education is increasing, which, in turn, necessities intensive English programs being either compulsory or voluntary because either the medium of instruction in some universities is in English or some courses are offered in English (Arslan, 2020). From this aspect, preparatory programs' role is crucial to help students attain a proficiency level in English so that they can follow their courses in English effectively, which also makes it important to evaluate if preparatory schools serve for this aim or not (Coşkun, 2013; Ekşi, 2017).

It is in this context that the present study aims to investigate the effectiveness of an English preparatory school program by applying Foreign Language Skills Scale (Arslan, 2020), which can be used to maintain a comprehensive overview of the process of teaching language skills in the light of the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001), in an intensive modular preparatory program at a Turkish state university.

THE CEFR

The CEFR consists of communicative language activities and strategies as well as a set of communicative language competences for which descriptors were developed in the Swiss research project (North, 2000; North & Schneider, 1998). The project revealed the existence of the six levels plus mid-parts of the scale which came to be known as plus levels. As North (2007, p. 3) suggests:

Between what can be described as the criterion level for A2 and the criterion level for B1 there was found to be a plus level. The same was the case between levels B1 and B2 (B1+) and between levels B2 and C1 (B2+). Such plus levels were characterized by a stronger performance in relation to the same features found at the criterion level, plus hints of features that became salient at the next level.

To date, the CEFR has been noted to have a major influence on assessment (Jones & Saville, 2009; Coste, 2007; Fulcher, 2008). The CEFR, which has been set out to be a framework for the elaboration of language syllabi or examinations, was noted by all countries to be most useful for the planning and development of curricula as well as tests and certification (North, 2007).

On the other hand, its impact within the classroom environment was noted to be less because it is found difficult to understand (Jones & Saville, 2009). Therefore, assessment of language teaching programs is crucial not only for the administrators but also for the English language practitioners to get a clear understanding and feedback on the process.

In Turkey, the Turkish Ministry of National Education and the Higher Education Council adopted the CEFR as a guide for their foreign language teaching policies. Thus, in order to meet the criteria established by the CEFR, the education system including primary, secondary and higher education was reconstructed. With respect to higher education, which is the focus of the present study, the medium of instruction at many universities in Turkey is English and the number is increasing day by day (Kirkgöz, 2005), which, in turn, make it necessary for the universities to offer intensive English programs for their students. Despite the centrality of these programs, when the implementation of the CEFR in higher education is examined, it can be seen that some preparatory programs do not match with the aims of the CEFR and not all of them are grounding their applications on the basis of the principles of the CEFR (Kınsız & Aydın, 2008; Gökdemir, 2010) and that there is a conflict between the CEFR goals and the realities of Turkish higher education (Peaci, 2015). The present study aims to examine whether there is such kind of a mismatch in the Preparatory Program at the School of Foreign Languages at Pamukkale University by taking CEFR learning outcomes into consideration.

RESEARCH STUDIES IN LANGUAGE PROGRAM EVALUATIONS IN TURKEY

Although the idea of language program evaluation is not new, the way it is conducted and theorized has changed over the years. In the Turkish context, program evaluation in the preparatory programs at universities has become increasingly important in the last two decades (Üstünlüoğlu et al, 2012). One of the earliest studies regarding the program evaluation was conducted by Erdem (1999). In the study, the curriculum of Middle East Technical University was investigated, and it was found that a more student-centered program and improvements in in-service training were necessary. Another study conducted at the same university by Topçu (2005) revealed that there was a significant difference between the opinions of teachers and students when the curriculum and objectives of the Basic English Department were examined. A similar study by Gerede (2005) at Anadolu University examined the previous and new curricula of the Preparatory Program and significant improvements were found in meeting the students' language needs. Karataş (2007) evaluated the English instruction program at Yıldız Teknik University by examining the syllabus and the results showed that the teachers had negative opinions towards the program when its effects on students' listening, speaking and grammar knowledge were concerned. By analyzing the effectiveness of the preparatory program at Osmangazi University, Özkanal and Hakan (2010) found that the students were satisfied with the program; however, they also indicated that students' academic English needs should be taken into account while developing the program. In another study, Üstünlüoğlu et al. (2012) investigated the effectiveness of the curriculum at Izmir University of Economics, School of Foreign Languages and the results indicated that students had difficulty in tasks requiring higher order thinking skills. Coşkun (2013) evaluated the existing language program of a state university and he found that except the ones prepared for the speaking course, materials were sufficient. On the other hand, teachers believed that materials were not effective enough to serve for their students' needs. In his study conducted at Çanakkale Onsekiz Mart University, Tekin (2015) examined the program of ELT/ELL preparatory classes and found that majority of the students were satisfied with the program except for its physical conditions and added that students had difficulty in communicate skills. Lastly, in their longitudinal study, Efeoğlu et al. (2018) evaluated the language program of the preparatory program of ELT department at Yıldız Teknik University and the results indicated that almost all participants found the re-evaluation of the previous program effective.

All the studies discussed above tried to examine the effectiveness of language programs in terms of course materials, curriculum, assessment, activities and teaching objectives in general; yet there is only one study conducted to develop a scale assessing the success level of language learners in both language skills and core English course as described in the CEFR. In his study, Arslan (2020) developed the Foreign Language Skills Scale (FLSS) by focusing on the descriptors of the CEFR including A1, A2, B1, and B1+ levels; however, there is no evaluation of any preparatory school program through the use of the scale since it has been currently developed. Therefore, being the first of its kind, the study aims to fill this gap in the literature by evaluating B1+ level program at an English Preparatory School in a Turkish State University.

AIM OF THE STUDY

The present study seeks to examine the success of learners from the level of B1+ in accordance with the learning outcomes of the CEFR; and thus, evaluating the language program of the preparatory school by using FLSS (Arslan, 2020). To achieve these aims, the following research questions are addressed;

When students having attended B1+ module are considered;

1. What are their overall opinions about their learning outcomes in four language skills in relation to their current level of English proficiency?
2. What are their opinions about their learning outcomes in the relevant sub-skills of each of the four language skills in relation to their current level of English proficiency?
3. What are their opinions about the contribution of Core English to their language development in the four skills?
4. What are their overall opinions about their learning outcomes in four language skills in relation to their genders?
5. What are their overall opinions about their learning outcomes in four language skills in relation to their ages?

METHOD

RESEARCH MODEL

The methodology of this study was survey research. Creswell (2015) mentioned that the survey research design enables researchers to describe population characteristics, values, attitudes or opinions both qualitatively and quantitatively through the studies applied on the sampled units that were selected from the population itself. Survey research may use a variety of data collection procedures with the most common being questionnaires and different type of interviews. In his study, Arslan (2020) used this research model as a scale development study to depict population tendencies, attitudes or opinions.

CONTEXT OF THE STUDY

The evaluated program is B1+ level English preparatory program in the School of Foreign Languages at Pamukkale University in Turkey in the 2020-2021 academic year. Students are from different departments such as International Trade and Finance, English Language Teaching, English Language and Literature, Textile Engineering and Business Administration, where the medium of instruction is in English in either all or in some selected courses. The preparatory program is based on a modular system where, at the beginning of the academic year, students are placed to the relevant level according to their level of English proficiency based on the placement exam. The programs of different levels in each module are designed to be in line with the Common European Framework of Reference for Languages (CEFR) guidelines, including A1, A2, B1, and B1+ levels. An entire academic year consists of 4

modules, each lasting 8 weeks and 192 hours in total. In a given module, the weekly program includes courses such as listening (2 hours), speaking (3 hours), writing (5 hours), reading (5 hours) and core language (9 hours). Students are supposed to attend 2 modules at least since the opening module of the year starts from B1 level and they are supposed to complete the program at B1+ level.

SAMPLE

The study was conducted during the Fall and Spring Term of 2020-2021 academic year. Convenience sampling method was used as a sampling method in this study. Convenience sampling is a specific type of non-probability sampling method that relies on data collection from population members who are conveniently available to participate in study. Seçer (2015) also used this method by claiming that the participants were already attending the foreign language preparatory program and they were easy to reach for research purposes. In line with his study, the present study applied the same sampling method by considering participants' availability and accessibility because they were also attending the program already.

DATA COLLECTION

PARTICIPANTS

The data was comprised of 357 students studying at the preparatory school at Pamukkale University., with 163 female and 194 male students. Table 1 demonstrates the descriptive statistics related to the demographic variables.

Table 1. Descriptive statistics related to the demographic variables of B1+ students

Variable	Category	N	%
Gender	Female	163	45.6
	Male	194	54.1
Age	18	130	36.7
	19	106	29.6
	20	80	22.4
	21	21	5.8
	Other	20	5.5
Faculty Type	Education	31	8.6
	Science and Letters	86	24.0
	Economics and Administrative Science	184	51.7
	Engineering	56	15.7

Table 1 shows that while the ages of the participants vary, the majority are between the ages of 18-20; and they are from four different faculties. All students involved in the study had experienced at least two modules in the program since during the initial module students from the highest level of proficiency started from B1 module depending on their level of entry to the program. Thus, every student having completed B1 module successfully and attended B1+ module was asked to complete the questionnaire just before taking their final exam in B1+ module in the 2020-2021 academic year.

FOREIGN LANGUAGE SKILLS SCALE

Foreign Language Skills Scale (FLSS) developed by Arslan (2020) was used to collect the quantitative data during the study. FLSS consists of 27 items and 5 factors named by examining the contents of the items gathered under five factors. There were eight items in the first factor named *writing skills*. There were five items in each of the factors named *speaking skills*, *listening skills* and *reading skills*. In addition, there were 4 items in the fifth factor named *core English*. The scale items were designed to be scored as 4-point Likert-type items; namely, (1) Completely Disagree, (2) Disagree, (3) Agree, and (4) Completely Agree. The lowest possible score from this scale was 27, while the highest was 108.

Arslan (2020) tested the validity and the reliability of the scale with sample of 326 preparatory school students for the Exploratory Factor Analysis and 350 preparatory school students for the

Confirmatory Factor Analysis. Confirmatory and Exploratory Factor Analysis results supported 27 items and 5 factors named by examining the contents of the items gathered under these five factors. There were eight items in the first factor named *writing skill*. There were five items in each of the factors named *speaking skill*, *listening skill*, and *reading skill*. In addition, there were 4 items in the factor named *core English* and gave the best validity result model (together ($\chi^2/sd = 1.893$ RMSEA=.049; GFI=.89; AGFI=.86; CFI=.95; NNFI=.90; RFI=.89).

In this study, validity and reliability results of the FSL and 5 factors were calculated by the researcher on 357 preparatory school students. For the Validity of the scale, the fit indices calculated after the confirmatory factor analysis process were $\chi^2 / Sd = 1.92$; RMSEA = .042; GFI = .92; AGFI = .88; NFI = .92; NNFI = .92; CFI = .94; IFI is .90 and RFI is .93.

Reliability of the test or scale is one of the important indices that is related to whether or not a measurement instrument provides the consistent and sensitive results in times of repeated application (Buyukozturk, 2002; Baykul, 2000). Arslan (2020) indicated that the Cronbach's Alpha reliability coefficients of the factors related to the FLSS were as 0.913 for Reading, 0.879 for Listening, 0.838 for Speaking, 0.834 for Writing and 0.853 for Core English. The Cronbach's Alpha value for the whole scale was 0.957.

In this study, Cronbach's Alpha reliability coefficients of the factors and whole scale were also calculated. The results showed that the Cronbach's Alpha reliability coefficients of the factors about FLSS were as 0.875 for Reading, 0.833, for Listening, 0.863 for Speaking, 0.895 for Writing and 0.911 for Core English. The FLSS Cronbach's Alpha value for the whole scale was 0.944.

DATA ANALYSIS

In order to examine the research questions, the data obtained from the samples were uploaded into the SPSS 22.00 software program and analyzed. First, Descriptive Statistics was used for the first three research questions. The lowest and highest means scores for each item and domains were calculated and interpreted based on the characteristics of the items and domains. Next, for the last three research questions the extreme values were controlled before the analysis and the assumptions about the analyses were tested. As a result of tests of normality, it was found that the data fits a normal distribution, and the values of skewness and kurtosis were between +2 and -2 (George & Mallery, 2010). Therefore, independent samples t-test and One Way ANOVA were used for comparison of the groups. For the Post Hoc procedure in One Way Anova, Tukey's HSD was used to identify the source of the mean differences among groups. The effect size for independent samples t-test and One Way Anova results were calculated with Cohen's d. Cohen (1988) suggested that $d = 0.2$ be considered a 'small', 0.5 represents a 'medium' and 0.8 a 'large' effect size.

FINDINGS

OVERALL EVALUATION

Data were collected using a questionnaire named the Foreign Language Skills Scale (FLSS). Mean and standard deviations for the five factors and overall results of the FLSS are given below in Table 2.

Table 2. Mean and standard deviation of the overall evaluation

Factors	N	Mean	Sd
Reading Skills	357	3.25	.49
Listening Skills	357	3.03	.56
Speaking Skills	357	3.10	.54
Writing Skills	357	3.21	.55
Core English	357	3.23	.63
Overall Evaluation	357	3.17	.45

In regard to the first research question, the results in Table 2 reveal that students are satisfied with the improvement in their language skills in general (overall evaluation mean: 3,17). However, they see themselves as more competent in “Reading Skills” and less competent in “Listening Skills”, with mean scores of 3,25 and 3,03, respectively.

FACTORS

READING SKILLS

“Reading Skills” is the first factor of the FLSS and it is represented by five questions. Among the items belonging to the “Reading Skills” in the questionnaire, it can be seen in Table 3 that item 2, “I can answer questions related to a reading text”, has the highest mean whereas item 1, “I can guess the meaning of words I do not know in a reading text” has the lowest mean in the questionnaire. Thus, with respect to the second research question, it is found that students evaluate themselves as more successful in skimming and scanning, and less successful in deducing the meaning from context which are the subskills of reading.

Table 3. Mean and standard deviation of the Reading Skills

Item No	Item	N	Mean	Sd
1	I can guess the meaning of words I do not know in a reading text.	357	3.18	.58
2	I can answer questions related to a reading text	357	3.34	.58
3	When answering a question about a reading text, I can easily find the section related to the question	357	3.28	.64
4	can understand the main idea of a text I read	357	3.19	.68
5	I can deduce from a text I read	357	3.24	.62
Overall	Reading Skills	357	3.25	.49

LISTENING SKILLS

“Listening Skills” is the second factor of the FLSS and it is comprised of five questions. Table 4 shows that with respect to the items of the “Listening Skills” in the questionnaire, item 10, “During the listening process, I can catch phrases such as ‘the door of the room’, and ‘students in the class’”, has the highest mean while item 6, “I can take notes when somebody speaks”, has the lowest mean among all the items in the questionnaire. Therefore, concerning the second research question, the results indicate that among the subskills of listening students have more positive opinions in dictation and more negative opinions in note-taking.

Table 4. Mean and standard deviation of the Listening Skills

Item No	Item	N	Mean	Sd
6	I can take notes when somebody speaks.	357	2.90	.72
7	During the listening process, when I am asked, I can catch the details such as who, where, and when,	357	3.08	.67
8	I can understand the main idea of any conversation I listen to.	357	3.07	.71
9	I can deduce the meaning of a word I do not know from the context when I listen to a conversation	357	2.86	.75
10	During the listening process, I can catch phrases such as ‘the door of the room’, and ‘students in the class’.	357	3.23	.68
Overall	Listening Skills	357	3.03	.57

SPEAKING SKILLS

“Speaking Skills” is the third factor of the FLSS and it contains five questions. As shown in Table 5, among the items in the “Speaking Skills” part, item 15, “I can express personal information about myself”, has the highest mean; on the other hand, item 13, “I can communicate with native speakers of English”, has the lowest mean in the questionnaire. Hence, regarding the second research question, it can be seen that students evaluate their competency higher in giving personal information but lower in being able to interact when speaking subskills are concerned.

Table 5. Mean and standard deviation of the Speaking Skills

Item No	Item	N	Mean	Sd
11	I can answer any question when somebody asks me.	357	3.14	.62
12	I can communicate with non- native speakers of English.	357	3.02	.73
13	I can communicate with native speakers of English.	357	2.93	.75
14	I can participate in a conversation.	357	3.01	.72
15	I can express personal information about myself.	357	3.38	.61
Overall	Speaking Skills	357	3.10	.55

WRITING SKILLS

“Writing Skills” is the fourth factor of the FLSS and it includes eight questions. When the items in “Writing Skills” of the questionnaire are taken into account, it is found that item 19, “*I can write coherent texts*”, has the highest mean as seen in Table 6.

Table 6. Mean and standard deviation of the Writing Skills

Item No	Item	N	Mean	Sd
16	I can write sentences with meaning relations such as cause-effect, contrast, and comparison.	357	3.23	.68
17	I can write a paragraph	357	3.13	.70
18	I can express my feelings and thoughts in writing	357	3.28	.65
19	I can write coherent texts.	357	3.36	.63
20	I can enrich the text I write by using conjunctions	357	3.08	.71
21	I can use examples, quotes, or statistics to support my ideas when I write a paragraph.	357	3.28	.67
22	I can write the sections of a paragraph such as topic sentence, supporting sentences, and concluding sentence.	357	3.15	.74
23	I can rewrite a given sentence with the same meaning.	357	3.19	.78
Overall	Writing Skills	357	3.21	.55

On the other hand, item 20, “*I can enrich the text I write by using conjunctions*”, has the lowest mean in the questionnaire. Consequently, in regard to the second research question, the results reveal that students feel more competent in writing coherently and less competent in enriching texts by conjunctions.

CORE ENGLISH

“Core English” is the fifth factor of the FLSS and it consists of four questions. As for the items in the “Core English” part of the questionnaire, Table 7 presents that item 26, “*My writing skill has improved*”, has the highest mean whereas item 24, “*My speaking skill has improved*” and item 27, “*My listening skill has improved*” have the lowest mean in the questionnaire. Thus, with respect to the third research question, the results show that students think Core English course contributed more to their development in writing skill while it has less contribution in their speaking and listening skills.

Table 7. Mean and standard deviation of Core English

Item No	Item	N	Mean	Sd
24	My speaking skill has improved.	357	3.14	.76
25	My reading skill has improved	357	3.30	.72
26	My writing skill has improved.	357	3.32	.72
27	My listening skill has improved	357	3.14	.75
Overall	Core English	357	3.23	.63

VARIABLES

GENDER

Table 8 shows descriptive statistics for average FLSS scores for the female and the male group. Independent-samples t-test results illustrated that there was a significant difference between the female and the male groups in terms of overall evaluation [$t_{(355)} = 4,559, p < .05, r = 0.16$],

Reading Skills [$t_{(355)} = 2,853, p < .05, r = 0.04$], Listening Skills [$t_{(355)} = 2,032, p < .05, r = 0.07$], Speaking Skills [$t_{(355)} = 2.650, p < .05, r = 0.09$], Writing Skills [$t_{(355)} = 5,458, p < .05, r = 0.19$], and Core English [$t_{(355)} = 4.548, p < .05, r = 0.16$]. The results shown in Table 3.7 reflect that female students have higher mean scores than male students. Hence, it can be argued that the FLSS scores of the participants are affected by gender variable. It can be understood that female students are more satisfied with the B1+ preparatory program when compared to male students.

Table 8. Independent Samples t-Test Results Regarding FLSS Scores of Gender

Variable	Group	n	M	SD	t	p
Reading	Female	163	3.30	.50	2.853	0.004
	Male	194	3.20	.48		
Listening	Female	163	3.07	.61	2.032	0.042
	Male	194	2.99	.51		
Speaking	Female	163	3.15	.56	2.650	0.008
	Male	194	3.05	.52		
Writing	Female	163	3.32	.52	5.458	0.000
	Male	194	3.11	.55		
Core	Female	163	3.33	.58	4.548	0.000
	Male	194	3.13	.65		
Overall	Female	163	3.24	.45	4.559	0.000
	Male	194	3.09	.43		

AGE

In regard to the FLSS overall score, Tukeys’ HSD demonstrated that mean score of 18-year-old students is significantly higher than the ages of 19, 20, 21 and other ages (Table 9). In addition, mean score of 19-year-old students is significantly higher than the ages of 20, 21 and other ages. These results demonstrated that this difference is to the favor of the 18- and 19-year-old students. In other words, 18- and 19-year-old students’ evaluations are meaningfully more positive than those of the students from the other age groups. No significant difference was found among the students from other age groups.

Table 9. One-Way ANOVA Results Regarding FLSS Scores of Age

Variable	Group	n	M	SD	F	p	Difference
Overall	18 year-olds	130	3.27	.40	12.523	0.000	1 – 2;
	19 year-olds	106	3.19	.44			1 – 3;
	20 year-olds	80	3.02	.47			1 – 4;
	21 year-olds	21	3.04	.42			1 – 5;
	Others	20	2.97	.48			2 – 3;
Reading	18 year-olds	130	3.35	.46	8.131	0.000	2 – 5;
	19 year-olds	106	3.24	.47			1 – 2;
	20 year-olds	80	3.12	.53			1 – 3;
	21 year-olds	21	3.11	.46			1 – 4;
	Others	20	3.16	.47			
Listening	18 year-olds	130	3.12	.51	11.103	0.000	1 – 3;
	19 year-olds	106	3.09	.57			1 – 5;
	20 year-olds	80	2.85	.57			2 – 3;
	21 year-olds	21	2.93	.56			2 – 5;
	Others	20	2.74	.52			
Speaking	18 year-olds	130	3.19	.46	10.187	0.000	1 – 3;
	19 year-olds	106	3.16	.55			1 – 5;
	20 year-olds	80	2.92	.54			2 – 3;
	21 year-olds	21	2.97	.72			2 – 5;
	Others	20	2.91	.59			
Writing	18 year-olds	130	3.29	.50	5.503	0.000	1 – 3;

	19 year-olds	106	3.23	.57			1 – 5;
	20 year-olds	80	3.10	.54			
	21 year-olds	21	3.15	.56			
	Others	20	3.01	.62			
Core	18 year-olds	130	3.39	.53	11.922	0.000	1 – 2;
	19 year-olds	106	3.21	.66			1 – 3;
	20 year-olds	80	3.08	.62			1 – 4;
	21 year-olds	21	2.94	.65			1 – 5;
	Others	20	3.01	.74			2 – 4;

1=18 year-olds; 2=19 year-olds; 3=20 year-olds; 4=21 year-olds; 5= Others

With respect to the reading skills, Tukeys’ HSD demonstrated that mean score of 18-year-old students is significantly higher than the ages of 19, 20 and 21. These results demonstrated that this difference is to the favor of the 18-year-old students. In other words, 18-year-old students’ evaluations are meaningfully more positive than those of the students from 19-, 20- and 21-year-old students. No significant difference was found among the students from the other age groups.

As for the listening skills, Tukeys’ HSD demonstrated that mean score of 18-year-old students is significantly higher than the ages of 20 and other age groups. In addition, 19-year-old students’ mean score is significantly different and higher than 20 and other age groups. In other words, 18- and 19-year-old students’ evaluations are meaningfully more positive than those of the students from 19 and other age group. No significant difference was found among the students from other age groups.

In regard to the speaking skills, Tukeys’ HSD demonstrated that mean score of 18-year-old students is significantly higher than the ages of 20 and other age groups. In addition, 19-year-old students’ mean score is significantly different and higher than 19 and other groups of ages. In other words, 18- and 19-year-old students’ evaluations are meaningfully more positive than those of the students from 20 and other age groups. No significant difference was found among the students from other age groups.

With respect to the writing skills, Tukeys’ HSD demonstrated that mean score of 18-year-old students is significantly higher than the ages of 20 and other age groups. In other words, 18-year-old students’ evaluations are meaningfully more positive than those of the students from 20 and other age groups. No significant difference was found among the students from other age groups.

For Core English, Tukeys’ HSD demonstrated that mean score of 18-year-old students is significantly higher than the ages of 19, 20, 21 and other age group. In other words, 18-year-old students’ evaluations are meaningfully more positive than those of the students from 19 and other age groups. In addition, 19-year-old students’ mean score is significantly higher than the mean score of 21-year-old students. No significant difference was found among the students from other age groups.

DISCUSSION AND CONCLUSION

To begin with, the results of the quantitative data indicate that, in general, the modular system designed to be in line with the Common European Framework of Reference for Languages (CEFR) guidelines serves for a high-quality language program for the students enrolled in the preparatory school at Pamukkale University since B1+ students performed a high level of agreement in the acquisition of all language skills. This can be because these students went through at least two modules, namely B1 and B1+. In these modules both skill courses and core English course include lots of communicative tasks and integrated skills activities and the syllabi include fostering more challenging subskills such as note-taking, deducing meaning

from context, and inferencing. Thus, in general, considering that they are exposed to such kind of activities 24 hours a week and for at least 2 modules, they might have a sense of achievement by having practiced these skills again and again.

The findings also show that while students generally find themselves competent in all four language skills at the end of the B1+ module, they feel themselves most confident in the reading skills. This finding is in parallel to Tekin's (2015) study stating that most of the students see themselves as competent in reading skills. One important finding of the study is that in reading skills, guessing the meaning of words from context and getting the main idea of a text seem to be the subskills students do not feel satisfied as much as the other ones. This can stem from the fact that while studying on their own, students can directly conduct their dictionaries and do not force themselves to get the meaning by making use of contextual clues. It is also possible that because of the heavy syllabus teachers are supposed to cover each week, they may not spend enough time on practicing general strategies, recognition and interpretation of specific context clues or feedback.

The second skill students see themselves as more competent is writing skills. Contrary to these findings, Tekin's (2015) study revealed that 68% of the participants perceived themselves as insufficient in writing skills. This difference might be since his participants were all from ELT and ELL departments and the expectations of those students can be higher than the preparatory students of other departments. It is also possible that since those students from language departments were separately placed in the preparatory program, their syllabus might include different genres with challenging tasks which can negatively affect students' perceptions about their competencies. On the other hand, participants of the present study are from various departments placed in the same class and the syllabus is common regardless of their departments which means that there is not a separate and more challenging writing syllabus for ELT and ELL students. The present study also revealed that, in writing skills, there remains some inadequacies among students in enriching a text by using conjunctions.

Another important finding concerns the speaking skills in that students seem to have a relatively low number of competencies in communicating with both native speakers and non-native speakers of English and participating in a conversation. This finding of the research matches up with Tekin's (2015) study having found that speaking skills seem to be one of the most problematic skills based on participants' perceptions. This finding is also in parallel to Özkanal & Hakan's (2010) study stating that while 83,7% of the students perceived speaking skills as very important, only 31,7% thought that their level of language skills is good. In his study, Coşkun (2013) also found that the skill least frequently focused on in the class is speaking skills. Possible reasons for these consistent findings can be three-fold; firstly, the language programs do not focus on and foster speaking skills by placing more communicative activities in their syllabi and implementing them in the class by decreasing teacher talk and increasing student talk; secondly, as Coşkun (2013) suggested, the materials are not sufficient enough to foster their speaking skills; and finally it is also probable that students do not force themselves to use English in the classroom environment because they are afraid of making mistakes and they do not feel secure about it.

The most prominent discrepancy is between the learning outcomes of B1+ module related to listening skills and the opinions of students about their competencies in the skill since the results signal students' relevant inadequacy in listening skills, especially in deducing the meaning of an unknown word in a conversation and note-taking skills. In his study, Tekin (2015) also found that half of the participants see themselves as insufficient in listening skills. Overall, these findings are in accordance with the findings reported by Özkanal & Hakan (2010) in which students suggested that for the program to be more effective listening skills should be more actively conducted by having more emphasis on note-taking skills.

Finally, students' opinions about the contribution of Core English to their language development in four language skills point out the effectiveness of the course in their language development in each skill. Students' positive opinions about the course can stem from the fact that Core English materials and the coursebook used in the course include integrated skills with various activities from the language skills and students have more chance to practice them as the Core English course takes up two days with nine hours a week.

IMPLICATIONS

The present study was carried out as an evaluation of B1+ level program at an English Preparatory School in a Turkish State University by applying Foreign Language Skills Scale (Arslan, 2020) to maintain a comprehensive overview of the process of teaching language skills in the light of the CEFR learning outcomes. The study revealed the necessity of placing more focus on fostering listening skills by increasing the use of different language activities in the class which are designed to serve for these purposes. Moreover, the English Preparatory Program should place different types of activities such as long public talks, a-2-minute talk, individual and group presentations, communicative activities, and design different implementations like speaking clubs to contribute to the students' sense of achievement in speaking skills.

REFERENCES

- Arslan, Recep Ş. "The Development of a Scale to Evaluate Foreign Language Skills at Preparatory Schools". *International Journal of Assessment Tools in Education*, 2020, 7 (2) (2020): 223-235.
- Cohen, Jacob. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.
- Coşkun, Abdullah. "An investigation of the effectiveness of the modular general English language teaching preparatory program at a Turkish university". *South African Journal of Education*, 33 (3) (2013): 1-18.
- Council of Europe (CoE). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg, France: Council of Europe. Cambridge University Press, 2001. Retrieved from <https://www.coe.int/en/web/common-european-frame-work-reference-languages>.
- Creswell, John W. *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (5th ed.). Boston, MA: Pearson, 2015.
- Ekşi, Gonca Y. "Designing curriculum for second and foreign language studies". In A. Sarıçoban (Ed.), *ELT methodology*. Ankara: Anı Yayıncılık, 2017.
- Erdem, Hikmet E. "Evaluating the English language Curriculum at a Private School in Ankara: A Case Study" Yayınlanmamış Doktora Tezi. Middle East Technical University, Ankara, 1999.
- Fulcher Glenn. "Criteria for evaluating language quality". In E. Shohamy (Ed.), *Language testing and assessment. Encyclopedia of language and education*, (pp. 157-176), 7. Amsterdam: Springer, 2008.
- George, Darren and Mallery, Paul. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update (10a ed.)* Boston: Pearson.
- Gerede, Derya. "A Curriculum Evaluation Through Needs Analysis: Perceptions of Intensive English Programme Graduates at Anadolu University" Yayınlanmamış Yüksek Lisans tezi. Anadolu University, Eskisehir, 2005.
- Gökdemir, Cem V. "Üniversitelerimizde Verilen Yabancı Dil Öğretimindeki Başarı Durumumuz (Our Success in Foreign Language Teaching in Our Universities)". *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 6 (2) (2010): 251–264. Retrieved from <http://dergipark.gov.tr/ataunisobil/issue/2816/37912>
- Güllü, Seda. A. "An evaluation of English programme at Kozan Vocational School of Çukurova University: Students' point of view" Yayınlanmamış Yüksek Lisans Tezi. Çukurova Üniversitesi, Adana, 2007.
- Hopkins, David. *Evaluation for school development*. Milton Keynes: Open University Press, 1989.
- Jones, Neil and Saville, Nick. (2009). "European language policy: Assessment, learning, and the CEFR". *Annual Review of Applied Linguistics*, 29 (2009): 51–63.
- Karataş, Hakan. "Yıldız Teknik Üniversitesi Modern Diller Bölümü İngilizce 2. Dersi Öğretim Programının Öğretmen ve Öğrenci Görüşlerine Göre Bağlam, Girdi, Süreç ve Ürün (CIPP) Modeli ile Değerlendirilmesi" Yayınlanmamış Yüksek Lisans Tezi. Yıldız Teknik Üniversitesi, İstanbul, 2007.
- Kınsız, Mustafa and Aydın, Hüseyin. "Yükseköğretimde yabancı dil öğrenim ve öğretim sorununun değerlendirilmesi ve yabancı dil politikası üzerine düşünceler (Evaluating the problem of foreign language learning and teaching in higher education and reflecting on foreign language policy)". *Yabancı dil Bölümleri*

- ve Yükseköğretim Kurumları Yabancı Dil Öğretimindeki Sorunları Kurultayı (Problems of Foreign Languages Departments and Schools in Foreign Language Teaching) (pp. 71-82). Muğla University, 7-9 May, 2008.
- Kırkgöz, Yasemin. "Motivation and student perception of studying in an English-medium university". *Journal of Language and Linguistic Studies*, 1 (1) (2005): 101-123.
- North, Brian. *The development of a common framework scale of language proficiency*. New York: Peter Lang, 2000.
- North, Brian. *The CEFR common reference levels: Validated reference points and local strategies*. In *Language Policy Forum Report* (pp. 19-29). Strasbourg: Cambridge University Press, 2007.
- North, Brian and Schneider, Günther. "Scaling descriptors for language proficiency scales". *Language Testing*, 15 (1998): 217-262.
- Özkanal, Ümit and Hakan, Ayhan G. "Effectiveness of University English Preparatory Programs: Eskişehir Osmangazi University Foreign Languages Department English Preparatory Program". *Journal of Language Teaching and Research*, 1 (3) (2010): 295-305.
- Peaci, Davut. "A Class of Cultures: European and Turkish Goals in Language Education versus University Academic Requirements". *Anthropologist*, 22 (2) (2015): 355-361.
- Scriven, Michael. *Evaluation Thesaurus* (4th ed.). Newbury Park, CA: Sage, 1991.
- Seçer, İ. (2015). *Psikolojik test geliştirme ve uyarlama süreci*. Ankara: Anı Yayıncılık.
- Stufflebeam, Daniel L. "The relevance of the CIPP evaluation model for educational accountability". *Journal of Research and Development in Education*, 5 (1971): 19-25.
- Tekin, Mustafa. "Evaluation of a Preparatory School Program at a Public University in Turkey". *The Journal of International Social Research*, 8 (36) (2015): 718-733.
- Topçu, Tansu. "An investigation of the Effectiveness of Theme-Based Curriculum in the 2003-2004 Academic Year at the Department of Basic English at Middle East Technical University" Yayınlanmamış Yüksek Lisans Tezi. Middle East technical University, Ankara, 2005.
- Üstünlüoğlu, Evrim, Zazaoğlu, Kısmet F., Keskin, Michelle N., Sarayköylü, Beril & Akdoğan, Gülfem. "Developing a CEF based curriculum: a case study". *International Journal of Instruction*, 5 (1) (2012): 115-128.
- Worthen, Blaine. "Program evaluation". In H. Walberg & G. Haertel (Eds.), *The international encyclopedia of educational evaluation* (pp. 42-47). Toronto, ON: Pergamon Press, 1990.