

The Development of a Scale to Evaluate Foreign Language Skills at Preparatory Schools

Recep S. Arslan ^{1,*}

¹Pamukkale University, Faculty of Education, Kınıklı Campus, 20070, Denizli, Turkey

ARTICLE HISTORY

Received: Dec 18, 2019

Revised: Apr 01, 2020

Accepted: May 05, 2020

KEYWORDS

Scale development,
English language teaching,
Teaching language skills

Abstract: The aim of the present study is to develop a valid and reliable scale evaluating the effectiveness of language preparatory programs in the acquisition of language skills. In the development of Foreign Language Skills Scale (FLSS) in this study, research sample consisted of 326 preparatory school students for the exploratory factor analysis (EFA) and 350 preparatory school students for the confirmatory factor analysis (CFA). Based on the data obtained from the first sample, an EFA was carried out on the FLSS. EFA has identified that 27 items of the scale have factor loads between 0.519 and 0.729, while they explain 65.376% of the total variance and are distributed under five factors. These factors are named as *writing skill*, *speaking skill*, *listening skill*, *core skills*, and *reading skill*. A CFA was applied on the data obtained from the second sample that consisted of 350 students. As a result of the CFA, it was confirmed that the FLSS consisted of 27 items and five factors. For all the items in the scale, item-subscale, item-test correlation coefficients and mean differences between the upper and the lower 27% of the participants were calculated, and it is determined that each item is consistent with not only the subscale it is under but also the whole test. In addition, the Cronbach's Alpha reliability coefficients of the total scale's and five sub-scales' internal consistency is quite high. The FLSS is expected to offer a comprehensive evaluation of the acquisition of four language skills in foreign language teaching programs.

1. INTRODUCTION

In the Turkish context there is an emerging need for individuals with a sound knowledge of at least one foreign language, which is usually English. With respect to higher education, the increasing demand for English, in turn, makes it necessary for the universities to offer intensive English programs being either compulsory or voluntary since either the medium of instruction at a number of state universities in Turkey is in English or some courses are offered in English. To this end, preparatory programs offer intensive English courses for tertiary level students before they are admitted to their own field of studies in faculties. Due to their crucial role in enabling tertiary level students to gain a proficient knowledge of English so that they can follow their courses in English effectively, it has, therefore, become essential to evaluate whether preparatory schools serve such ends or not (Coşkun, 2013; Ekşi, 2017).

The Higher Education Council (2016) responsible for the coordination of universities in the Turkish context states the aim of foreign language education as “to teach students basic

CONTACT: Recep S. Arslan ✉ rsarslan@pau.edu.tr 📍 Pamukkale University, Faculty of Education, Department of Foreign Languages Teaching, Kınıklı Campus, 20070, Denizli, Turkey

ISSN-e: 2148-7456 /© IJATE 2020

principles of the foreign language that they are taught, to enhance their foreign language vocabulary, and to ensure that they can understand what they read and listen in a foreign language and they can express themselves orally or in writing” as declared in the Official Gazette dated 23.03.2016, with number 29662. However, curriculum design as well as its implementation and evaluation is left to universities. Regardless of compulsory or elective foreign language instruction offered in preparatory programs at tertiary level in Turkey, the Common European Framework of Reference for Languages (CEFR) that “describes in a comprehensive way what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively” (Council of Europe, 2001, p.1) is taken into consideration by almost all state and private foundation universities in designing preparatory programs. CEFR places students in six varying levels, including A1 level as breakthrough/beginner/basic user, A2 level as waystage/elementary/basic user, B1 level as threshold/intermediate/independent user, B2 level as vantage/upper intermediate/independent user, C1 level as effective operational proficiency/advanced user, and C2 level as mastery/proficiency/proficient user under understanding (listening and reading), speaking (spoken interaction/spoken production) and writing with illustrative scales for each skill (Council of Europe, 2001). North (2007) also suggests the existence of six levels plus mid-parts of the scale which came to be known as *plus levels* such as B1+ between levels B1 and B2 and B2+ between levels B2 and C1.

To date, the CEFR, which has been set out to be a framework for the elaboration of language syllabi or examinations, was noted to be the most useful for the planning and development of curricula as well as designing tests and certification (North, 2007). Therefore, evaluation of foreign language teaching programs based on the CEFR guidelines is crucial not only for administrators but also for English language practitioners to get a clearer understanding of and give feedback on the process as it would help administrators and instructors see success, reveal strengths and weaknesses, and make necessary improvements (Black, Harrison, Lee, Marshall, & Wiliam, 2004). It is, therefore, of paramount importance to evaluate language programs systematically and effectively in order to improve the quality expected from such efforts (Coşkun, 2013; Ekşi, 2017; Kiely & Rea-Dickins, 2005; Peacock, 2009).

A number of studies have attempted to evaluate preparatory programs at tertiary level in the Turkish setting from different perspectives. A few researchers evaluated language programs using Context, Input, Process and Product (CIPP) model and mostly reported that the language curriculum components were viewed positively; however, some improvements as to physical conditions, content, materials and assessment in the curriculum needed to be made (Akpur, Alcı, & Karataş, 2016; Coşkun, 2013; Karataş & Fer, 2009; Tunç, 2010). Moreover, Karcı-Aktaş and Gündoğdu (2020) applied ‘Bellon and Handler model’ to evaluate the English preparatory curriculum of a state university and stated the problems as lack of philosophy or goals of the English preparatory curriculum, inefficacy of the skills courses, communication problems between the administration and other participants, need to improve the physical facilities, need for professional English language teaching, and the necessity to involve all stake-holders in decision making processes.

Some other studies also evaluated language programs using survey techniques and also came up with similar results. Language programs were viewed as effective in general; however, content, course materials, and teaching equipments (Güllü, 2007), physical contexts and the necessity to develop communicative skills (Tekin, 2015), and objectives, teaching materials, assessment, evaluation, and general structure (Uysal, 2019) were stated among problematic issues. In addition, curriculum needed revision in line with students’ needs (Sağlam & Akdemir, 2018), curriculum needed to include academic or English for specific purposes courses (Balcı, Ügüten, & Çolak, 2018; Özkanal & Hakan, 2010) or technical English (Özkanal, 2009), a

preference towards teaching academic skills rather than general English was needed (Keser & Köse, 2019), there were some motivational and attendance problems (İşcan, 2017), speaking skill needed to receive more attention and also content, materials and activities were to be modified (Öner & Mede, 2015), speaking and listening skills considered weak also needed to be included more in the program (Yılmaz, 2009), and four language skills were to be tested through contextualised and communicative test items for backwash effect (Paker, 2013).

All these studies have attempted to evaluate foreign language teaching preparatory programs in terms of objectives, content, course materials, teaching equipments, physical contexts, and language components in general (Akpur, Alcı, & Karataş, 2016; Balcı, Ügüten, & Çolak, 2018; Coşkun, 2013; Karataş & Fer, 2009; Özkanal & Hakan, 2010; Tunç, 2010); however, no study has yet attempted to investigate learners' success level in specific language skills; namely, speaking, reading, writing, or listening skills. Development of a scale to evaluate the effectiveness of preparatory programs in the acquisition of language skills has, therefore, been essential, and it is in this context that the present study aims to develop a scale which can be used to maintain a comprehensive overview of the process of acquisition of language skills within the field of foreign language teaching in an intensive modular preparatory program at a Turkish state university.

2. METHOD

This study used the basic survey model as a scale development study.

2.1. Context of the Study

This study was conducted in the School of Foreign Languages at Pamukkale University in Turkey in the 2018-2019 academic year. The preparatory school founded in 2004 has been offering intensive English language instruction since 2007-2008 academic year for about 1000 students each year. Students are enrolled in various departments such as Business Administration, International Trade and Finance, English Language Teaching, English Language and Literature, Textile Engineering, and Electric and Electronics Engineering, where medium of instruction is in English in either all or in some selected courses. With the idea that the modular system can be effective as students can be placed according to their level of English proficiency, and they can also receive appropriate education designed in line with the CEFR guidelines, the preparatory program has been based on a modular system since 2015-2016 academic year (Erarslan, 2019). Pamukkale University Preparatory Program is also based on the descriptors of the Common European Framework of Reference for Languages (CEFR) including A1, A2, B1, and B1+ levels. Students admitted to the program for at least two modules and at most four modules depending on their level of entry to the program are all supposed to complete the program at B1+ level. Volunteering students have the chance to attend B2 level as well. Each module lasts 8 weeks and the program runs 24 hours weekly with 192 hours of courses in total in a module. The weekly schedule includes such language skills courses as listening (2 hours), speaking (3 hours), writing (5 hours) and reading (5 hours) as well as a core language course for 9 hours. Students in the program go through formative and summative assessment through quizzes, performance assignments, one midterm examination and one final examination for each module.

2.2. Samples

The study was carried out during the Spring Term of 2018-2019 academic year. Convenience sampling method was used to reach the sample since all the participants were already attending the preparatory program and they were easy to reach for research purposes. In this study, different samples were chosen from different levels to conduct a scale development study.

During the scale development phase of the study 326 students studying at Pamukkale University preparatory school participated in exploratory factor analysis (EFA) conducted on the data obtained from the samples. Of the participants 111 (34%) were B1 level, and 204 (62.6) were B1+ level and 11 (3.4 %) were B2 level students. 142 (43.6%) were female and 183 (56.1%) were male students. 1 student (.3 %) did not mention the gender. The validity and reliability work of Foreign Language Skills Scale was obtained at the end of the pilot study conducted on the selected sample. For Comrey and Lee (1992), 300 is good for a sufficient sample size for factor analysis while Kline (1994) finds 200 individuals enough for a sample size with reliable factors.

A confirmatory factor analysis (CFA) was also carried out on the data obtained from the sample group of 350 students. The number of participants per item was more than 10 individuals as the scale consisted of 27 items. Of the participants 105 (29.2 %) were A1 level, 99 (27.5 %) were A2 level, 109 (30.3%) were B1 level, and 47 (13.1%) were B1+ level. 194 (53.9%) were male and 165 (45.8%) were female male students while 1 student (.3 %) did not mention the gender.

2.3. Data Collection

The validity and reliability analyses of the scale were conducted at the end of the pilot study with the selected sample.

2.3.1. Foreign language skills scale for preparatory schools

Foreign Language Skills Scale (FLSS) for Preparatory Schools was developed similar to the scaling approach based on grading totals developed by Likert (1932). During the scale development, first, literature on CEFR and evaluation of language programs was reviewed. Since review of the related literature did not show any measurement tools evaluating language skills in English Language Teaching Preparatory Programs based on CEFR, no specific sample was used while developing the scale items. Based on the review of literature, a number of 67 items were developed for the scale in line with the CEFR descriptors. An item pool of 67 items related to evaluation of language program was then submitted for the opinions of 35 experts in preparatory schools or English Language Teaching departments to consult their views on the development of items in order to validate the item pool of the scale.

During the pilot study stage, the items in the pools were examined by two English language teaching experts and one measurement and evaluation expert as well. According to expert views, researchers removed 34 items of the pilot scale as to the experts such items did not measure what was intended for or such items were found ambiguous. After the pilot study, there were 33 scale items based on 5-point Likert-type; namely, Strongly Agree (5), Agree (4), Neither Agree or Disagree (3), Disagree (2) and Strongly Disagree (1).

2.4. Data Analysis

In order to examine the validity and reliability analyses of the instrument, the data obtained from the first and second samples were uploaded onto the SPSS 22.00 and AMOS 16 software programs and analyzed. Firstly, for the purpose of determining the construct validity of the scale, KMO (Kaiser-Meyer-Olkin) and Bartlett's tests were carried out on the data obtained from the first sample to see the data's suitability for factor analysis. KMO value was obtained to determine if data structure suits factor analysis based on the sampling adequacy. Bartlett's Test of Sphericity was obtained to see the multivariate normal distribution of the data. In determining whether the data are appropriate for factor analysis, Kaiser-Meyer-Olkin value is to be greater than .70. For Bartlett's sphericity test, it was checked whether $p < .05$. .30 for the contribution value to common variance; .40 was used as a criterion for factor load value. While deciding the number of factors, the scree plot graph was used. Based on the obtained values, an EFA was carried out on the data.

Additionally, for each item in the scale, between item-subscale and the item-test correlation coefficient scores were calculated with a purpose to see whether each item was consistent with the subscale and whole scale. In addition, the statistical difference between item scores' means between groups of the upper and lower 27% were examined with 0.05 alpha level. Subsequently, a CFA was applied on the data obtained from the second sample. During the confirmatory factor analysis phase, data set of another 350 students was examined, and extreme and missing values were checked. In order to calculate the reliability coefficient of the scale, the Cronbach's Alpha reliability coefficient method was used.

3. FINDINGS

3.1. Findings on Validity

3.1.1. Exploratory factor analysis

Construct validity was applied to the measurement scale in order to determine the extent to which the FLSS as the measurement instrument can measure the variable it aims to measure without confusing it with other variables (Balci, 2009; Gorsuch, 1983). To determine the construct validity of the FLSS, firstly, Kaiser-Meyer-Olkin and Bartlett's test analyses were conducted on the data collected from the first sample, and the values were obtained as KMO= 0.940; Bartlett's test value $\chi^2 = 5390.619$; $sd=351$ ($p=0.000$). As KMO values of higher than 0.60 are seen to be sufficient for factor analysis in the social and educational sciences (Büyüköztürk, 2002), it was decided that factor analysis could be conducted on the 33 item in the scale.

In Exploratory Factor Analysis, Principal Component Analysis (PCA) is a technique that is used to reveal whether or not the items in a scale could be divided into a lower number of factors that eliminate each other (Büyüköztürk, 2002). In order to classify the factors that were formed by collecting the items, Varimax orthogonal rotation technique was preferred as a rotation method since it was not expected that there would be a high degree of correlation among the factors that emerged in the principal component analysis (Kline, 1994). Items that have factor load values under 0.30 and those that are distributed under more than one factor with less than a difference of 0.10 between their factors loads need to be removed from the scale (Balci, 2009; Büyüköztürk, 2002). As a result of the analyses in this study, the eigenvalues of the items had to be at least 1.00, while their factor loads at least 0.50. Items that were distributed under multiple factors were eliminated, 6 items were removed, and the analyses were carried out on the remaining 27 items.

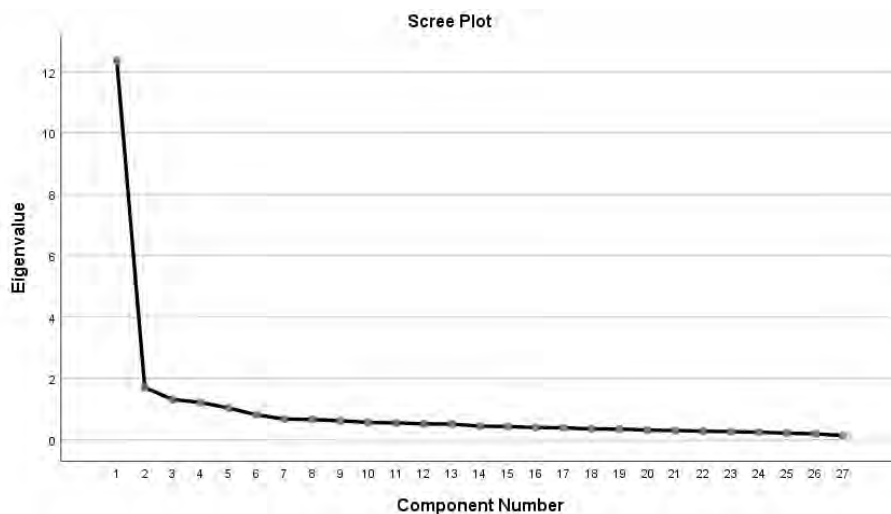


Figure 1. Eigenvalues based on the factors

As can be seen from the scree plot graph in Figure 1, 27 items can be collected under five factors. Without subjecting the remaining 27 items to rotation, it was found that the factor loads varied between 0.614 and 0.770. After subjecting the items to the Varimax orthogonal rotation technique, these factor loads were found to vary between 0.663 and 0.780. Additionally, it was identified that the items and factors in the scale explained 65.37% of the total variance. As it was stated that this ratio needs to be at least 40% (Kline, 1994; Scherer, Wiebe, Luther, & Adams, 1988), the obtained value was found sufficient. This finding obtained by EFA is shown in Figure 1 based on the eigenvalues. When Figure 1 is examined, it is seen that after five factors there is a routinized variation, and therefore, these factors have significant contribution to the variance.

Furthermore, the factors were 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 skills*. Table 1 presents findings on the item loads of the remaining 27 items based on the factors, factor eigenvalues and variance explanation ratios.

Table 1. FLSS common variances, item factor loads, variances explained by sub-scales and item analysis results

| Items | Common Variance | Factor 1 Writing Skill | Factor 2 Speaking Skill | Factor 3 Listening Skill | Factor 4 Core Skills | Factor 5 Reading Skill |
|--------------------|-----------------|---------------------------|----------------------------|-----------------------------|-------------------------|---------------------------|
| Q54 | .696 | .784 | | | | |
| Q51 | .709 | .726 | | | | |
| Q52 | .691 | .713 | | | | |
| Q56 | .624 | .708 | | | | |
| Q53 | .594 | .671 | | | | |
| Q55 | .614 | .658 | | | | |
| Q46 | .671 | .655 | | | | |
| Q47 | .577 | .609 | | | | |
| Q35 | .705 | | .731 | | | |
| Q36 | .665 | | .710 | | | |
| Q40 | .683 | | .703 | | | |
| Q37 | .646 | | .667 | | | |
| Q38 | .729 | | .647 | | | |
| Q31 | .678 | | | .731 | | |
| Q29 | .662 | | | .700 | | |
| Q30 | .628 | | | .623 | | |
| Q33 | .673 | | | .616 | | |
| Q28 | .559 | | | .608 | | |
| Q61 | .710 | | | | .761 | |
| Q63 | .697 | | | | .741 | |
| Q60 | .707 | | | | .737 | |
| Q65 | .606 | | | | .609 | |
| Q15 | .711 | | | | | .755 |
| Q16 | .734 | | | | | .746 |
| Q18 | .571 | | | | | .519 |
| Q21 | .519 | | | | | .479 |
| Q19 | .591 | | | | | .474 |
| Eigenvalue | | 5.04 | 3.67 | 3.12 | 2.93 | 2.89 |
| Explained variance | | 18.68 | 13.58 | 11.57 | 10.85 | 10.70 |
| Total Variance | | | 65.37 | | | |

As seen in Table 1, the factor loads of the items in the factor *writing skill* of the scale varied between 0.609 and 0.784. The eigenvalue of this factor in the general scale was 5.04, and its contribution to the general variance was 18.68%. The factor loads of the items in the factor *speaking skill* varied between 0.647 and 0.731. The eigenvalue of this factor was 3.67, and its contribution to the general variance was 13.58%. The factor loads of the items in the factor *listening skill* varied between 0.608 and 0.731. The eigenvalue of this factor was 3.12, and its contribution to the general variance was 11.57%. The factor loads of the items in the factor *core skills* varied between 0.609 and 0.761. The eigenvalue of this factor was 2.93, and its contribution to the general variance was 10.85%. And finally, the factor loads of the items in the factor *reading skill* varied between 0.474 and 0.755. The eigenvalue of this factor was 2.83, and its contribution to the general variance was 10.70%.

In addition, the relationship between the four factors in the FLSS was determined and for this reason, the correlations among the factors were checked. The findings are shown in Table 2.

Table 2. Correlation analysis results among the factors of the FLSS

| Factors | Writing Skill | Speaking Skill | Listening Skill | Core Skills | Reading Skill |
|-----------------|---------------|----------------|-----------------|-------------|---------------|
| Writing Skill | - | | | | |
| Speaking Skill | 0.641** | - | | | |
| Listening Skill | 0.667** | 0.684** | - | | |
| Core Skills | 0.625** | 0.642** | 0.692** | - | |
| Reading Skill | 0.606** | 0.602** | 0.592** | 0.574** | - |

** $p < 0.01$

As seen in Table 2, based on the correlation values among the factors of the FLSS, the five factors were found to be significantly related, while there was no problem of autocorrelation.

3.1.2. Item Discrimination

The correlation coefficients between the Item and Subscale correlation and Item and Test correlation were also calculated, and the discrimination rate of each item was determined in order to reveal the degree to which each item served the general purpose of the subscale it was in and the entire scale (Balçı, 2009; Baykul, 2000). Table 3 presents the items, item-factors, item-subscale correlations and item-test correlations.

As seen in Table 3, the item-subscale correlations were in the ranges of 0.665-0.748 for the first factor, 0.642-0.735 for the second factor, 0.593-0.683 for the third factor, 0.623-0.683 for the fourth factor and 0.608-0.692 for the fifth factor. Each item had a significant and positive relationship with the general scale ($p < 0.001$).

When the item-test correlation coefficients for the whole scale were examined, the lowest correlation value was found as 0.570, while the highest one was 0.739. Each item had a significant and positive relationship with the overall scale ($p < 0.001$). These coefficients that were calculated were the validity coefficients of all items, and they indicated the consistency of the items with the entire scale. In other words, these referred to the degree to which the scale served its general objective (Baykul, 2000).

The statistically significantly difference between item scores' means between groups of the upper and lower 27% were examined. It was found that all the items in FLSS were discriminated and the mean difference between the lower and upper groups was at a significant level of 0.05.

Table 3. Item discrimination analysis results

| Initial Item No | Updated Item No | Item | Factor | Item-Subscale Correlation | Item-Test Correlation | Upper/Lower 27% <i>t</i> |
|-----------------|-----------------|--|--------|---------------------------|-----------------------|-----------------------------|
| Q54 | 26 | I can enrich the text I write by using conjunctions | 1 | .748 | .635 | 11.481** |
| Q51 | 23 | I can write a paragraph. | 1 | .758 | .692 | 16.394** |
| Q52 | 24 | I can express my feelings and thoughts in writing | 1 | .743 | .701 | 14.743** |
| Q56 | 28 | I can write the sections of a paragraph such as topic sentence, supporting sentences, and concluding sentence. | 1 | .712 | .649 | 12.908** |
| Q53 | 25 | I can write coherent texts. | 1 | .665 | .602 | 10.508** |
| Q55 | 27 | I can use examples, quotes, or statistics to support my ideas when I write a paragraph. | 1 | .739 | .696 | 14.032** |
| Q46 | 21 | I can write sentences with meaning relations such as cause-effect, contrast, and comparison. | 1 | .689 | .662 | 12.673** |
| Q47 | 22 | I can rewrite a given sentence with the same meaning. | 1 | .692 | .662 | 11.839** |
| Q35 | 12 | I can answer any question when somebody asks me. | 2 | .738 | .664 | 11.698** |
| Q36 | 13 | I can communicate with non- native speakers of English. | 2 | .718 | .641 | 11.847** |
| Q40 | 17 | I can express personal information about myself. | 2 | .642 | .619 | 13.305** |
| Q37 | 14 | I can communicate with native speakers of English. | 2 | .705 | .637 | 12.858** |
| Q38 | 15 | I can participate in a conversation. | 2 | .755 | .739 | 12.935** |
| Q31 | 10 | I can deduce the meaning of a word I do not know from the context when I listen to a conversation | 3 | .658 | .594 | 11.506** |
| Q29 | 8 | During the listening process, when I am asked, I can catch the details such as who, where, and when, | 3 | .660 | .627 | 12.841** |
| Q30 | 9 | I can understand the main idea of any conversation I listen to. | 3 | .683 | .669 | 12.927** |
| Q33 | 11 | During the the listening process, I can catch phrases such as 'the door of the room', and 'students in the class'. | 3 | .623 | .634 | 11.268** |
| Q28 | 7 | I can take notes when somebody speaks. | 3 | .593 | .570 | 12.200** |
| Q61 | 30 | My reading skill has improved. | 4 | .697 | .597 | 11.783** |
| Q63 | 32 | My listening skill has improved. | 4 | .668 | .594 | 12.565** |
| Q60 | 29 | My speaking skill has improved. | 4 | .692 | .617 | 13.599** |
| Q65 | 31 | My writing skill has improved. | 4 | .608 | .616 | 12.186** |
| Q15 | 1 | I can guess the meaning of words I do not know in a reading text. | 5 | .689 | .610 | 12.498** |
| Q16 | 2 | I can answer questions related to a reading text. | 5 | .687 | .637 | 12.773** |
| Q18 | 3 | When answering a question about a reading text, I can easily find the section related to the question. | 5 | .666 | .670 | 12.326** |
| Q21 | 5 | I can deduce from a text I read. | 5 | .627 | .646 | 10.455** |
| Q19 | 4 | I can understand the main idea of a text I read. | 5 | .668 | .702 | 13.602** |

** $p < 0.01$

3.1.3. Confirmatory factor analysis

The dimensions of the FLSS were determined to consist of five factors as a result of the EFA. To confirm these factors, the scales that consisted of 27 items was applied on the second sample and a CFA was carried out on the data. CFA is based on the relationship among observable and unobservable variables and testing them as hypotheses (Pohlmann, 2004).

According to the results that were obtained, the χ^2/df ratio was calculated as 1.893. A χ^2/df ratio of 5 or lower is considered to be sufficient for model data fit (Schumacker & Lomox, 2004; Wang, Lin & Luarn, 2006). Moreover, a χ^2/df ratio of smaller than 3 shows a high model-data fit (Schumacker & Lomox, 2004). The χ^2/df value obtained as 1.893 in this study was a significant indicator that the measurement instrument had single dimension. Another important index, the RMR value was calculated as 0.021. It is known that the RMR index needs to be between 0 and 1 (Golob, 2003).

Other fit indices were also computed to evaluate the fit of the model. The calculated goodness of fit indices values were as: IFI=0.951; CFI=0.951; GFI=0.888; NFI=0.902; AGFI=0.864, and RFI=0.890. It is generally acceptable that the indices to be in the range of 0.80-0.90 and the values higher than 0.90 refer to a good fit (Yap & Khong, 2006; Wang et al., 2006). The RMSEA analysis result was determined as 0.049. RMSEA values of lower than 0.10 show an acceptable level of model-data fit, while those lower than 0.05 are an indicator of a good fit (Bayram, 2013). Based on the χ^2/df , RMSEA and RMR values obtained from the data in the study, it may be stated that the measurement instrument consisted of five factors. Figure 2 shows the standardized Structural Equation Modelling parameter values on the obtained findings.

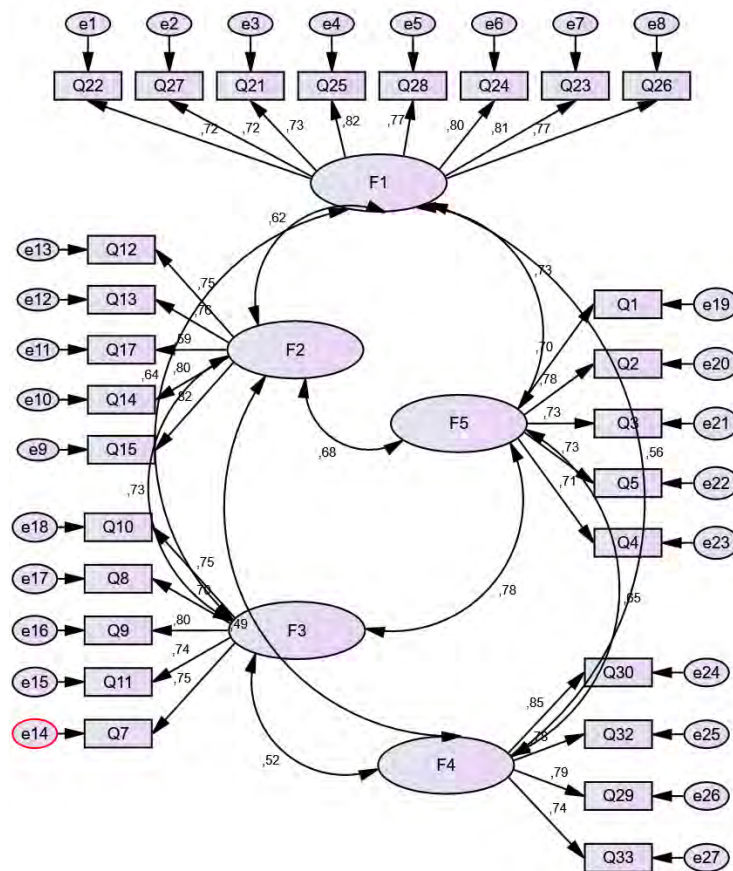


Figure 2. Confirmatory factor analysis results of the scale

As a result of the Confirmatory Factor Analysis, it was confirmed that the FLSS consisted of 27 items and five factors.

3.2. Findings on reliability

Reliability is a concept that is related to whether or not a measurement instrument provides the consistent and sensitive results in times of repeated application (Balçı, 2009; Baykul, 2000). As a result of the EFA, it was determined that the FLSS consisted of a total of 27 items and five factors. In order to identify the reliability indices of these five factors in relation to internal consistency, their Cronbach's Alpha reliability coefficients were calculated. The Cronbach's Alpha reliability coefficients of the factors were as 0.913 for *writing skill*, 0.879 for *speaking skill*, 0.838 for *listening skill*, 0.834 for *core skills* and 0.853 for *reading skill*. The Cronbach's Alpha value for the whole scale was 0.957.

The Cronbach's Alpha coefficient takes values in the range of 0.00 to 1.00. As the coefficient gets up to 1.00, the reliability of the measurement instrument increases, while as it gets closer to 0.00, the reliability decreases. In the educational and social sciences, in general, Cronbach's Alpha coefficients of 0.60 or higher are seen to be acceptable. On the other hand, the reliability indices used for preparing and applying psychometric tests is expected to be 0.70 or higher (Büyüköztürk, 2002). According to the findings obtained, the internal consistency coefficients for the factors and the entire scale were quite high in this study.

4. DISCUSSION and CONCLUSION

With a purpose to develop a scale in order to evaluate language skills in preparatory language teaching programs, 326 students studying at Pamukkale University preparatory school were asked to participate in the explanatory factor analysis phase of the scale development. Prior to the application of the scale, an item pool of 67 items was developed for the scale. An EFA was conducted on the data related to 67 items of the scale and 34 items that were found statistically insignificant were removed from the scale after calculations based on item-factor and item-test correlations. According to the results of EFA, it was decided that factor analysis could be conducted on the 33 items in the scale since Kaiser-Meyer-Olkin (KMO) and Bartlett's test values were obtained as KMO= 0.940; Bartlett's test value $\chi^2 = 5390.619$; $sd=351$ ($p=0.000$). As a result of the analyses, items that were distributed under multiple factors were eliminated, 6 items were removed, and the analyses were carried out on the remaining 27 items. A confirmatory factor analysis was carried out on the data obtained from the sample group of 350 students. The dimensions of the FLSS were determined to consist of five factors as a result of the EFA. To confirm these factors, the scale that consisted of 27 items was applied on the second sample and a CFA was carried out on the data. The χ^2/df value obtained as 1.893 in this study was a significant indicator that the measurement instrument had a single dimension. Another important index, the RMR value was calculated as 0.021. Other fit indices were also computed to evaluate the fit of the model. The calculated goodness of fit indices values were as: IFI=0.951; CFI=0.951; GFI=0.888; NFI=0.902; AGFI=0.864, and RFI=0.890. Based on the χ^2/df , RMSEA and RMR values obtained from the data in the study, the measurement instrument can be considered to consist of five factors.

In order to identify the reliability indices of these five factors in relation to internal consistency, their Cronbach's Alpha reliability coefficients were calculated. The Cronbach's Alpha reliability coefficients of the factors were as 0.913 for *writing skill*, 0.879 for *speaking skill*, 0.838 for *listening skill*, 0.834 for *core skills* and 0.853 for *reading skill*. The Cronbach's Alpha value for the whole scale was 0.957. These findings show the internal consistency coefficients for the factors and the entire scale quite high in this study.

Accordingly, in this particular study the Foreign Language Skills Scale that consisted of five factors and included 27 items was found to be a valid and reliable scale based on the statistical

data. This scale is expected to contribute to the field of foreign language teaching being a unique one that specifically addresses the evaluation of four main language skills in foreign language teaching programs. By using this scale, curriculum designers can evaluate the process of teaching language skills within the field of foreign language teaching and determine whether it is necessary to make changes, modifications or eliminations in the light of program goals and specific objectives. Since the main goal of foreign language teaching is to equip learners with an overall competency in understanding what they read and listen and also in expressing themselves orally or in writing in a foreign language, all items included in the scale would also help all those parties involved in such ventures to see how the actual practice fits the proposed goals of such programs in the acquisition of listening, speaking, reading and writing skills. Scores obtained as result of the application of this scale may either approve the programs as successful ones or may reveal the weaknesses and prompt immediate actions to tackle possible problems. Moreover, the application of the FLLS can also provide language instructors with valid data as to their own performance in teaching four language skills, and may, therefore, suggest whether they should revise their methods, materials, and activities.

However, the FLSS is not free from limitations. Since the scale consists of only 27 items, it assesses a limited number of subskills; thus, other scale attempts can be made to develop more comprehensive scales. Moreover, as the FLSS attempts to evaluate foreign language programs in terms of four language skills only, it excludes evaluation of other essential components of language programs such as the effect of course materials followed, course hours allocated, nature of programs (e.g. general or academic), teaching equipments used, physical contexts, roles of instructors and administrators, and involvement of stakeholders in decision-making processes of curriculum design. Therefore, more comprehensive scales that can investigate foreign language programs from such diverse points are timely.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author(s).

ORCID

Recep S. Arslan  <http://orcid.org/0000-0002-2475-5884>

5. REFERENCES

- Akpur, U., Alci, B., & Karataş, H. (2016). Evaluation of the curriculum of English preparatory classes at Yıldız Technical University using CIPP model. *Educational Research and Reviews*, 11(7), 466-473.
- Balcı, A. (2009). *Sosyal bilimlerde araştırma: yöntem, teknik ve ilkeler*. Ankara: PegemA Yayıncılık.
- Balcı, Ö., Durak Ügüten, S., & Çolak, F. (2018). Zorunlu İngilizce hazırlık programının değerlendirilmesi: Necmettin Erbakan Üniversitesi yabancı diller yüksekokulu örneği [The evaluation of compulsory preparatory program: The case of Necmettin Erbakan University school of foreign languages]. *Kuramsal Eğitim Bilim Dergisi*, 11(4), 860-893.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- Bayram, N. (2013). *Yapısal eşitlik modellemesine giriş* (2. baskı). Bursa: Ezgi Kitabevi.
- Black, P, Harrison, C, Lee, C., Marshall, B., & Wiliam, D. (2004) Working inside the Black Box: assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8-21.
- Büyüköztürk, Ş. (2002). *Sosyal bilimler için veri analizi el kitabı*. Ankara: PegemA Yayıncılık.

- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Coşkun, A. (2013). 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), 1-18.
- Council of Europe (CoE). (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg, France: Council of Europe. Cambridge University Press. Retrieved from <https://www.coe.int/en/web/common-european-framework-reference-languages>
- Ekşi, G. Y. (2017). Designing curriculum for second and foreign language studies. In A. Sarıçoban (Ed.), *ELT methodology*. Anı Yayıncılık: Ankara.
- Erarşlan, A. (2019). Progressive vs modular system in preparatory school English language teaching program: A case of system change at a state university in Turkey. *Dil ve Dilbilimi Çalışmaları Dergisi*, 15(1), 83-97.
- Golob, T. F. (2003). Structural equation modeling for travel behavior research. *Transportation Research*, 37(1), 1-25.
- Gorsuch, R. L. (1983). *Factor Analysis*. Hillsdale: Lawrence Erlbaum Associates.
- Güllü, A. S. (2007). *An evaluation of English program at Kozan Vocational School of Çukurova University: Students' point of view* (Unpublished master's thesis). Çukurova University, The Graduate School of Social Sciences, English Language Teaching Department, Adana.
- İşcan, S. (2017) The efficacy of modular EFL syllabus in prep classes. *International Journal of Management and Applied Science*, 3(8), 91-94.
- Karataş, H & Fer, S. (2009). Evaluation of English curriculum at Yıldız Technical University using CIPP model. *Education and Science*, 34, 47-60.
- Karcı-Aktaş, C. & Gündoğdu, K. (2020) An extensive evaluation study of the English preparatory curriculum of a foreign language school. *Pegem Eğitim ve Öğretim Dergisi*, 10(1), 169-214.
- Keser, A. D., & Köse, G. D. (2019). Determining exit criteria for English language proficiency in preparatory programs at Turkish universities. *The Online Journal of Quality in Higher Education*, 6(2), 45-49.
- Kiely, R. & Rea-Dickins, P. (2005). *Program evaluation in language education*. Pelgrave Macmillan.
- Kline, P. (1994). *An easy guide to factor analysis*. New York: Routledge.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 1-55.
- North, B. (2007). The CEFR illustrative descriptor scales. *The Modern Language Journal*, 91(4), 656- 659.
- Öner, G. & Mede, E. (2015). Evaluation of A1 level program at an English preparatory school in a Turkish university: a case study. *International Association of Research in Foreign Language Education and Applied Linguistics. ELT Research Journal*, 4(3), 204-226.
- 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.
- Özkanal, Ü., & Hakan, A. G. (2010). Effectiveness of university English preparatory programs: Eskisehir Osmangazi University foreign languages department English preparatory program. *Journal of Language Teaching and Research*, 1(3), 295-305.
- Paker, T. (2013). The backwash effect of the test items in the achievement exams in preparatory classes. *Procedia-Social and Behavioral Sciences*, 70, 1463-1471.

- Peacock, M. (2009). The evaluation of foreign-language-teacher education programmes. *Language Teaching Research*, 13(3), 259-78.
- Pohlmann, J. T. (2004). Use and interpretation of factor analysis in the journal of educational research: 1992-2002. *The Journal of Educational Research*, 98(1), 14-23.
- Sağlam, D., & Akdemir, E. (2018). İngilizce hazırlık öğretim programına ilişkin öğrenci görüşleri [Opinions of students on the curriculum of English preparatory program]. *Yükseköğretim ve Bilim dergisi/Journal of Higher Education and Science*, 8(2), 401-409.
- Scherer, R. F., Wiebe, F. A., Luther, D. C. & Adams, J. S. (1988). Dimensionality of coping: factor stability using the ways of coping questionnaire, *Psychological Reports*, 62(3), 763-770. PubMed PMID: 3406294.
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). NJ: Lawrence Erlbaum Associates, Mahwah.
- Tekin, M. (2015). Evaluation of a preparatory school program at a public university in Turkey. *Uluslararası Sosyal Araştırmalar Dergisi. The Journal of International Social Research*, 8(36), 718-733.
- The Higher Education Council (2016). Yükseköğretim Kurumlarında Yabancı Dil Öğretimi ve Yabancı Dille Öğretim Yapılmasında Uyulacak Esaslara İlişkin Yönetmelik [Regulation of the Higher Education Institutions on foreign language teaching and rules to be obeyed]. Retrieved from <https://www.mevzuat.gov.tr/Metin.Aspx?MevzuatKod=7.5.21475&MevzuatIliski=0&sourceXmlSearch=Y%C3%BCksek%C3%B6%C4%9Fretim%20Kurumlar%C4%B1nda%20Yabanc%C4%B1%20Dil%20%C3%96%C4%9Fretimi%20ve%20Yabanc%C4%B1%20Dille%20%C3%96%C4%9Fretim%20Yap%C4%B1lmas%C4%B1nda%20Uyulacak%20Esaslar>
- Tunç, F. (2010). *Evaluation of an English Language Teaching Program at a Public University Using CIPP model* (Unpublished masters' thesis). Ankara: Middle East Technical University. Retrieved from <http://etd.lib.metu.edu.tr/upload/12611570/index.pdf>
- Uysal, D. (2019) Problems and solutions concerning English language preparatory curriculum at higher education in view of ELT instructors. *International Journal of Contemporary Educational Research*, 6(2), 452-467.
- Wang, Y., Lin, H., & Luarn, P. (2006). Predicting consumer intention to use mobile service. *Information Systems Journal*, 16(2), 157-179.
- Yap, B.W., & Khong, K.W. (2006). Examining the effects of customer service management (CSM) on perceived business performance via structural equation modelling. *Applied Stochastic Models in Business and Industry*, 22, 587-605.
- Yılmaz, F. (2009). English language needs analysis of university students at a voluntary program. *The Journal of Social Sciences Research*, 4(1), 148-166.