60-Distance education system evaluation of English language teaching: A scale development study

Ömer Faruk İPEK[1]

Hakan Tahiri MUTLU[2]

APA: İpek, Ö. F. & Mutlu, H. T. (2022). Distance education system evaluation of English language teaching: A scale development study. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, (26), 978-994. DOI: 10.29000/rumelide.1074140.

Abstract

As educational curricula have been transformed from face-to-face to distance education, new and different tools in educational program evaluation have become a necessity. Therefore, in this study we aimed to develop a five-point Likert-type scale to measure the effects of the distance education system of English language teaching. For the pilot study, a 40-item questionnaire was applied to 121 students studying in the English preparatory program at a Turkish university. First, the factor analysis study of the scale was created, and then the item analysis studies were carried out and the scale was updated by removing 10 items with a low factor load or collected under more than one factor. The new scale, consisting of 30 items, was applied to 370 students, and as a result of the factor analysis study, the scale consisted of five different structures named as "Language Skills", "Communication", "Content Evaluation", "Instructors" and "Assessment". The Cronbach Alpha internal consistency coefficient for the entire scale, whose validity and reliability studies were completed, was determined as 0.952. Finally, confirmatory factor analysis studies were carried out and it was determined that the goodness of fit indices were at an acceptable level and the exploratory factor analysis results were confirmed.

Keywords: EFL, Program evaluation, Distance education, Scale development, Exploratory and confirmatory factor analyses

İngiliz dili öğretiminin uzaktan eğitim sistem değerlendirmesi: Ölçek geliştirme çalışması

Öz

Eğitim müfredatlarının yüz yüze eğitimden uzaktan eğitime dönüşmesiyle birlikte eğitim programlarının değerlendirilmesinde yeni ve farklı araçlar bir zorunluluk haline gelmiştir. Bu nedenle bu çalışmada, uzaktan eğitim sisteminin İngilizce öğretimine etkilerini ölçmek için beşli Likert tipi bir ölçek geliştirmek amaçlanmıştır. Pilot çalışma için bir Türk üniversitesinde İngilizce hazırlık programında öğrenim gören 121 öğrenciye 40 maddelik anket uygulanmıştır. Önce ölçeğin faktör analizi çalışması oluşturulmuş, ardından madde analizi çalışmaları yapılmış ve faktör yükü düşük olan veya birden fazla faktör altında toplanan 10 madde çıkarılarak ölçek güncellenmiştir. 30 maddeden oluşan yeni ölçek 370 öğrenciye uygulanmış ve faktör analizi çalışması sonucunda ölçek "Dil Becerileri", "İletişim", "İçerik Değerlendirme", "Eğitmenler" ve "Değerlendirme" olmak üzere beş

[1] Dr. Öğr. Üyesi, Bolu Abant İzzet Baysal Üniversıitesi, Yabancı Diller Yüksekokulu, Hazırlık Bülümü (Bolu, Türkiye), theipekk@gmail.com, ORCID ID: 0000-0003-1921-3332 [Araştırma makalesi, Makale kayıt tarihi: 09.09.2021-kabul tarihi: 20.02.2022; DOI: 10.29000/rumelide.1074140]

[2] Dr. Öğr. Üyesi, Bolu Abant İzzet Baysal Üiversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, Sayisal Yöntemler ABD (Bolu, Türkiye), tahirimutlu@ibu.edu.tr, ORCID ID: 0000-0002-8964-2696

farklı yapıdan oluşmuştur. Geçerlik ve güvenirlik çalışmaları tamamlanan ölçeğin tamamı için Cronbach Alpha iç tutarlılık katsayısı 0,952 olarak belirlenmiştir. Son olarak doğrulayıcı faktör analizi çalışmaları yapılmış ve uyum indekslerinin kabul edilebilir düzeyde olduğu belirlenmiş ve açıklayıcı faktör analizi sonuçları doğrulanmıştır.

Anahtar kelimeler: EFL, Program değerlendirme, Uzaktan eğitim, ölçek geliştirme, Açıklayıcı ve doğrulayıcı faktör analizi

1. Introduction

After the Covid-19 pandemic that affected the whole system of the world, education has also transformed from face-to-face to distance education (DE). Distance learning supplies huge potential innovations to instructional designs and procedures while it also carries many challenges in order to succeed the program aims (Moore & Fodrey, 2017). However, the benefits of conducting distance education can be listed that distance education is not only for students but also for everyone who are still inquisitive about new subjects and topics (Petsuvan et al., 2019), and it is so flexible that it gives both synchronous and asynchronous teaching / learning opportunities to the learners (Clark, 2020). The use of distance learning is not only for students in remote areas, but it is also used by everyone interested in learning. As all levels of education have been exposed to transformation, teaching English language which became the common language of international communication has also experienced change on its own behalf (Jenkins et al., 2017) With this massive change in English language teaching (ELT), implementation of the program evaluation has required new ways. Therefore, in this study, in order to keep up with the new evaluation process, we aimed to develop a scale to investigate the systemic aspect of ELT which can also be adopted to other fields in education.

1.1. Distance Learning

Distance learning is defined as a way of delivering instruction to both individual and group basis to the learners who are not present physically in a classroom or in other settings using technological devices (Rao & Krishen, 2015). Babori et al. (2016) asserted that various devices and platforms were introduced in order to develop distance learning and education. These tools are investigated in technical, managerial, and organizational, informational, and educational aspects (Vasilevska et al. 2017). Evaluation, on the other hand is employed to decide on whether the program aims have met the desired outcomes. While implementing a curriculum, it is important to evaluate and give feedback on the program both during and at the end (Usun, 2016). Rovai (2003) pointed out that program evaluations that result in changes and regulations in the curriculum or instructional procedures are crucial when quality is the objective. Moreover, language educators, program managers and policy makers have become aware of the program evaluation due to the need of understanding the accountability in testing, quality assurance and outcome assessment (Norris, 2009). Thompson and Irele (2003) stated that distance education programs must be evaluated due to four reasons: to affirm the investment for the sources, to find out whether the objectives have been achieved, to support the progress of goals, and to make decisions for continuing, expanding, or inactivate the program.

1.2. Program Evaluation

Program evaluation is directly related to curriculum design because it is necessary to conduct a program evaluation study to see whether the curriculum results are effective and efficient, and it should be

checked whether the desired results and goals are achieved. In order to understand the program evaluation processes, we first need to look into the components of curriculum design. McNulty (2013) lists five components of curriculum design as educational objectives, content, materials, teaching methods, and assessment. These components vary in different studies such as Sand et al. in 1960, which is one of the earliest studies in curriculum design and four components of curriculum are listed. These are "(a) objectives, including both behavioral and content components; (b) types and quality of opportunities for learning, including organizing centers for learning; (c) organizing threads and patterns of organization; and (d) evaluation procedures" (p. 266). Lunenberg (2011) on the other hand emphasized three different components as objectives, content or subject matter, and learning experiences.

While these components are important before designing the curriculum, the other aspect of education is to test the outcomes and results of the teaching / learning process during and at the end of education and this procedure is called program evaluation. Chen and Chen (2005) define program evaluation as "the application of evaluation approaches, techniques, and knowledge to systematically assess and improve the planning, implementation, and effectiveness" (p.3). Also, Murphy (2000) describes evaluation as the tool to decide how much a program is effective and efficient in terms of its aims. While these definitions may tell us the scope of program evaluation, Brown (1995) detailed it by describing three irrevocable dimensions of program evaluation as time of evaluation (formative vs. summative), focus of evaluation (process vs. product) and data type (qualitative vs. quantitative). In this study, we aimed to create a data collection tool that can be used for the dimensions above.

When we look at the literature, a great number of research has been conducted on program evaluation in ELT (Saito & Ebsworth, 2004; Coskun & Daloglu, 2010; Tunc, 2010; Tom-Lawyer,2014; Hsu, 2014; Khorunnisa, 2018;). In such studies, face to face English language education has been evaluated. Moreover, Chan (2001) compared the expectations of students and instructors at a university setting. Also, in another study conducted by Peacock (2009), course content and technology integration to the English programs were discussed. Efeoglu et al. (2018) also studied the program of an English program and found that all parties must be included to the process. There are also other studies conducted on perceptions of the stakeholders in English language programs (Le & Tran, 2021; Kiely&Rea-Dickens,2005), and by looking at the results of these studies, the required changes have been implemented after the program evaluation processes.

On the other hand, as the distance learning has expanded recently, literature about program evaluation during distance education in ELT is not as comprehensive as the program evaluation process done after the face-to-face education; therefore, it is pointed out that evaluating the effects of distance education curriculum has become a necessity (Dorrega, 2016). Studies on distance education emphasize different aspects of program evaluation. For instance, Davie (1995) investigated the computer-mediated information technologies, and Swart (2016) emphasized the role of student feedback to assess the leaning outcomes at the end of distance learning process. Moreover, Berge and Muilberg (2002) stated the role of student readiness for distance education. Regarding distance education program evaluation, Gunawardena et al. (2000) focus on an evaluation model of the distance education. On the other hand, Sae-Khow (2014) investigated the e-learning applications and implementations.

It can be concluded that, while distance education and learning has been widely used, the evaluation has gained upmost importance. In order for a systemic evaluation process, the researchers need well-designed and structured evaluation tools as Lynch (1996) expresses that one of the most common data

methods for data collection in program evaluation is questionnaires. Therefore, in order to gather data to evaluate students' thoughts on distance education system at an English Language school at a university, we aimed to create a new inventory in order to fill the gap in the current literature. This inventory tests students' thoughts on the whole distance education system including language skills, communication, content evaluation, instructors, and assessment. Also, we compared these components considering independent variables such as academic level, academic achievement scores and attendance. As we look at the scales used as a data collection tool in program evaluation, studies consist of some common categories such as such as skills, materials, textbooks, and assessment while the communication category is missing in all studies. However, during distance education, the importance of communication cannot be denied (Cunningham, 2017). Also, these questionnaires were designed in order to evaluate the face-to-face programs. On the other hand, the developed scale in this study that is called 'Distance education system evaluation scale (DESES)', which can also be adapted to other fields of education, is specifically designed to evaluate English language taught via distance education and consist of the communication category. Therefore, the scale that is developed in this study would help to provide researchers with an appropriate evaluation instrument in distance education. For this purpose, in order to present a valid and reliable distance education program evaluation tool, the following research questions have been investigated.

1. What are the exploratory factor and reliability analyses results of Distance Education System Evaluation Scale (DESES)?

2. What are the confirmatory factor analysis results of Distance Education System Evaluation Scale (DESES)?

2. Method

In this section, the setting, selection of participants, data collection and analysis, pilot study and ethics committee approval will be explained in detail.

2.1. Setting

This study was carried out in an intensive foreign language education institution of a state university located in the northern part of Turkey. The main purpose of the school is to teach general English. Approximately, 900 students receive education at the school every year. While the school provided face-to-face education until March 2020, after this date, it switched to distance education due to the Covid-19 pandemic. The school has been providing intensive English education at the university for about 30 years for an average of 720 hours in an academic year consisting of two semesters. There are students from nearly 30 different departments. In each class, there are approximately 25 students.

2.2. Participants

Purposeful sampling (Denzin & Lincoln, 2000) in which the researcher selects participants in order for the suitable representation of the whole group is used in order to select the participants. The questionnaire was sent to the whole population which was nearly 800 students; however, 370 university level English preparatory school students who took one-year English language education during distance education in 2020-2021 academic year participated in this study. These students were asked to fill in the questionnaire that was written in google forms, and it was sent to the participants via WhatsApp groups and e-mails.

There are two types of students in the institution. The first is optional and the other one is compulsory students. If thirty percent or more of the courses are given in English in major departments, students who will study in these departments have to take English preparatory education compulsorily and they must successfully complete the preparatory education in order to move on to the major departments. If thirty percent or less of the courses are taught in English in major departments, these students do not need to take a compulsory preparatory program and these students are in the status of optional preparatory students. Optional preparatory students can take one year of optional preparatory language education before going to their major, and unlike compulsory group, even if students are not successful, they can move on to their major.

2.3. Data collection and analysis

In accordance with the purpose of this study, first a pilot study of DESES was done with the online survey method and the data obtained in the Excel was transferred to the SPSS 26 package program. The factor structure of the developed scale was examined with the help of Exploratory Factor Analysis (EFA), and the scale was updated again by removing items with low factor loadings or items that were collected under more than one factor. Then, the renewed scale was applied to the sample number again with the online survey method in accordance with the universe and the data obtained were transferred to the SPSS 26 package program from the Excel environment. First, EFA was performed on the data, and the factor structure of the scale was examined and divided into sub-factors. In order to ensure the validity of the sub-factors obtained as a result of EFA, Confirmatory Factor Analysis (CFA) was performed with the help of AMOS 23 package program. The Cronbach Alpha value was examined to determine the reliability level of the validated scale. In order to ensure the validity and reliability of the scale, firstly, with the help of SPSS 26 package program. In the interpretation of the results obtained, the level of statistical significance was accepted as 0.05.

2.4. Pilot study

In this summative evaluation study whose focus was the product of the system, in order to measure the effects of the distance education received by the students in the English preparatory program, the literature was examined in detail and five-point Likert-type items were created to measure the program development and evaluation of the new system. The created scale items were analyzed by two faculty members who are experts in their fields and the items were confirmed. Scale items were created in the native language of the participants that is Turkish, and then data were collected from the participants in their own language. Afterwards, the items were translated by eight language experts. Then five items on which no agreement could be reached, reviewed by a final faculty member, who was an expert in English language and English items were decided on. As a result, a 40-item scale was prepared to measure the effects of the distance education system of the students studying in the English preparatory program. Afterwards, the created measurement tool was applied as a pilot application to 121 participants studying in the English preparatory program of a Turkish university with the online questionnaire method and the data obtained in the Excel format were transferred to the SPSS 26 package program. The factor structure of the developed scale was examined with the help of EFA and the 1st, 4th, 5th, 9th, 19th, 20th, 22nd, 23rd, 24th factors with low factor load or gathered under more than one factor were examined. The scale was updated again with the elimination of items and 30 items were included in the scale.

2.5. Ethics

As this research involves human participants, approval from the ethics committee of Bolu Abant Izzet Baysal University situated in Turkey is taken officially on 01/06/2021 with the approval protocol number 2021/254.

3. Results

In this study, it is aimed to develop the "Distance Education System Evaluation Scale (DESES)". Other evaluation scales in the literature were examined, and a 5-point Likert-type scale consisting of 40 questions was prepared as a draft after the 5-item demographic questions. After the pilot study, the scale, which was renewed by removing 10 items, was applied to 370 participants with the online questionnaire method and the data obtained were transferred to the SPSS 26 package program from Excel. In the following sections, EFA, CFA and reliability test results of DESES are presented.

3.1. EFA and reliability analyses results of Distance Education System Evaluation Scale (DESES).

In order to determine the construct validity of the developed DESES, EFA was conducted using principal component analysis with varimax rotation. In the analysis, factor loads were determined as at least 0.30 (Büyüköztürk, 2006) and the factor structure of the scale was examined and divided into sub-factors. Cronbach Alpha coefficient was calculated for the sub-dimensions and total reliability of the scale. In order to ensure the validity of the sub-factors obtained as a result of EFA and to test the accuracy of the structure, CFA was performed with the help of AMOS 23 package program.

After EFA was applied to the developed 30-item DESES, 5-factor structure was obtained which can be seen in Table 1 below. The variance explained by the first factor was 16.55%, the variance explained by the second factor was 16.47%, the variance explained by the third factor was 13.86%, the variance explained by the fourth factor was 11.35%, and the variance explained by the fifth factor was 8.72%. The total variance explained is 66.95%. The total variance explained is sufficient as it exceeds 50% (Büyüköztürk, 2006).

FACTORS	VARIABLES	$\bar{X}$±SS	Factor Loads	Explained Variance	Cronbach 's Alpha
Language Skills	1. I can understand what I read in English in accordance with the current level.	3,98±0,88	0,788	16,55	0,906
	2. I can speak English in accordance with the current level.	3,49±0,99	0,803		
	3. I can write in English in accordance with the current level.	3,74±0,96	0,789		
	4.I can understand what I listen in English in accordance with the current level.	3,84±0,89	0,743		
	5. I know the appropriate vocabulary for the current level.	3,51±0,96	0,749		

	6. I can pronounce English words appropriate to the current level.	3,81±0,87	0,617		
	7. I know the grammar rules appropriate to the current level.	3,70±0,92	0,634		
Communication	8. I could easily reach the responsible people when I faced any problem.	3,98±0,96	0,726	16,47	0,912
	9. I was able to communicate easily with the university during distance education.	3,57±1,14	0,749		
	10.I easily communicated with the department administration during distance education.	3,69±1,04	0,746		
	11. The problems I encountered in distance education were solved quickly and effectively by the responsible people.	3,82±0,95	0,796		
	12. Sufficient information was given about the Preparatory Education in distance education.	3,99±0,90	0,565		
	13. The documents on the website in distance education are sufficient.	3,72±1,00	0,567		
	14. Adequate information was given about extracurricular practices.	4,00±0,87	0,578		
Content Evaluation	15. I am satisfied with the distance education system.	2,90±1,33	0,820	13,86	0,911
	16. I easily adapted to distance education.	2,97±1,30	0,795		
	17. I think I can learn English through distance education.	3,27±1,16	0,695		
	18. I am satisfied with my learning and course performance in distance education.	3,29±1,19	0,699		
	19. I am satisfied with the application used for distance education (Teams, Zoom etc.).	3,86±0,95	0,493		
	20. The distance education system is suitable for active participation of students such as discussion, asking questions and group work.	3,39±1,15	0,693		
	21. I effectively used extracurricular applications.	3,69±0,94	0,485		
	22. I think that the English curriculum is prepared appropriately to distance education.	3,69±1,01	0,636		
Instructors	23. I can easily communicate with my instructors.	4,41±0,73	0,775	11,35	0,733
	24. I can easily ask questions to my instructors during the lessons.	4,43±0,74	0,785		
	25. Courses in distance education are taught effectively and efficiently.	3,74±1,10	0,486		
Assessment	26. I am satisfied with the exam frequency.	3,27±1,17	0,789	8,72	0,850
	27. In the exams, questions were asked about the subjects we studied.	3,61±1,10	0,822		

	28. In the exams, the difficulty level of the questions was appropriate.	3,37±1,13	0,704
	29.The distribution of points and overall grade evaluation required for year-end pass-fail in distance education are fair and accurate.	3,48±1,09	0,641
	30. I am satisfied with the use of different measurement tools (online exams, projects, writing assignments, etc.).	3,85±1,00	0,576
Evaluation Criteria	Kaiser-Meyer-Olkin Measure of Sampling Adequacy: ,944 Barlett's Test of Sphericity; Approx. Chi-Square: 8213,180 Sig. : 0,000 Extraction Method: Principal Components Rotation Method: Varimax Total Explained Variance : 66,95 Cronbach's Alpha: 0,952		

Table 1. Examining the Sub-Dimensions of the Developed Scale (EFA And Reliability Analysis)

Since the Kaiser-Meyer-Olkin (KMO) coefficient was 0.944 as a result of the KMO test, which was used to determine whether the sample size used in the study was sufficient, it can be said that the sample size in the study was quite sufficient. Since the significance value (p-value) obtained as a result of the Bartlett Test (Bartlett Test of Sphericity) was less than 0.05 (0.000 < 0.05), the data provided the assumption of multiple normal distribution (Hair et al. 1998; Akgül-Çevik, 2003; Coşkun & Mutlu, 2017) and confirms the feasibility of factor analysis. In other words, since the Bartlett test is significant, it is possible to say that there are high correlations between the variables, so the data set is suitable for factor analysis (Kalaycı, 2010; Karagöz et al., 2019).

In order for a factor to be very stable, it must have at least 3 items (Velicer & Fava, 1998). Therefore, the ratio of the number of items to the number of factors (n:p) has gained importance. According to Cattell (1978), this ratio should be between 3 and 6 (MacCallum et al. 1999). Gorsuch (2008) stated that this ratio should be at least 5. In order to apply CFA, there must be at least three variables that measure each latent variable. For this reason, attention was paid to have at least three variables under any factor. In addition, the factor weight should be ±0,30 and above (Kalaycı, 2010).

Reliability analysis of the 30-item scale directed to the participants was made in terms of both the overall and sub-factors, and the internal consistency coefficient (Cronbach Alpha coefficient) was 0.952 for the overall, 0.906 for the first factor, "Language Skills", 0.912 for the second factor, "Communication", and the third factor. It was obtained as 0.911 for "Content Evaluation", 0.733 for the fourth factor "Instructors" and 0.850 for the fifth factor "Assessment". Since the coefficient value obtained in terms of the overall scale and the first, second, third and fifth factors is greater than 0.80, the scale used is a highly reliable scale (Kalaycı, 2010). In addition, the internal consistency coefficient (Cronbach's Alpha coefficient) obtained for the fourth factor of the scale also shows that this sub-factor has sufficient reliability. In the analysis results obtained, it is seen that the scale has construct validity.

3.1.1. Naming the factors

Since the main reason for performing EFA is to reduce a large number of variables to a smaller number of factors, these factors should be named. This naming process is done according to the common

characteristics of the variables in the factor (Nakip, 2006). Items belonging to 5 factors obtained from EFA and appropriate names for these items were given (see table 2). The first factor consisting of 7 items was called "Language Skills", the second factor consisting of 7 items was called "Communication", the third factor consisting of 8 items was called "Content Evaluation", the fourth factor consisting of 3 items was called "Instructors" and the fifth factor consisting of 5 items was called "Assessment".

When the average scores of the answers given by the students participating in the study to the DESES and its sub-dimensions are examined, it is seen that the 'Instructors' sub-dimension has the highest average, and the 'Content Evaluation' sub-dimension has the lowest average. While the participants responded as "agree" to the 'Instructors', 'Communication', 'Language Skills' and 'Assessment' sub-dimensions on average, they responded as "undecided" to the 'Content Evaluation' sub-dimension on average.

Factors	**N**	**Minimum**	**Maximum**	$\bar{X}$**±s.d.**
Language Skills	370	7,00	35,00	26,07±5,17
Communication	370	7,00	35,00	26,77±5,56
Content Evaluation	370	8,00	40,00	27,05±7,14
Instructors	370	3,00	15,00	12,57±2,12
Assessment	370	5,00	25,00	17,58±4,35

Table 2. The total score obtained by the participants from the factors

The minimum, maximum and average values of the participants from DESES sub-dimensions are given below. According to the results obtained the highest score from Language Skill sub-dimension was 35, the lowest score was 7, and the mean score was 26.07±5.17; the highest score from the Communication sub-dimension was 35, the lowest score was 7, and the mean score was 26.77±5.56; the highest score from the Content Evaluation sub-dimension was 40, the lowest score was 8, and the mean score was 27.05±7.14; the highest score from the Instructors sub-dimension was 15, the lowest score was 3, and the mean score was 12.57±2.12; the highest score from the Assessment sub-dimension was 25, the lowest score was 5, and the mean score was 17.58±4.35.

3.2. The CFA results of Distance Education System Evaluation Scale (DESES).

CFA was performed to ensure the validity of the sub-factors obtained as a result of EFA and to test the accuracy of the structure. The diagram of the model fit was obtained as follows in Figure 1 and the fit values related to the created model are given in Tables 4-5 below.

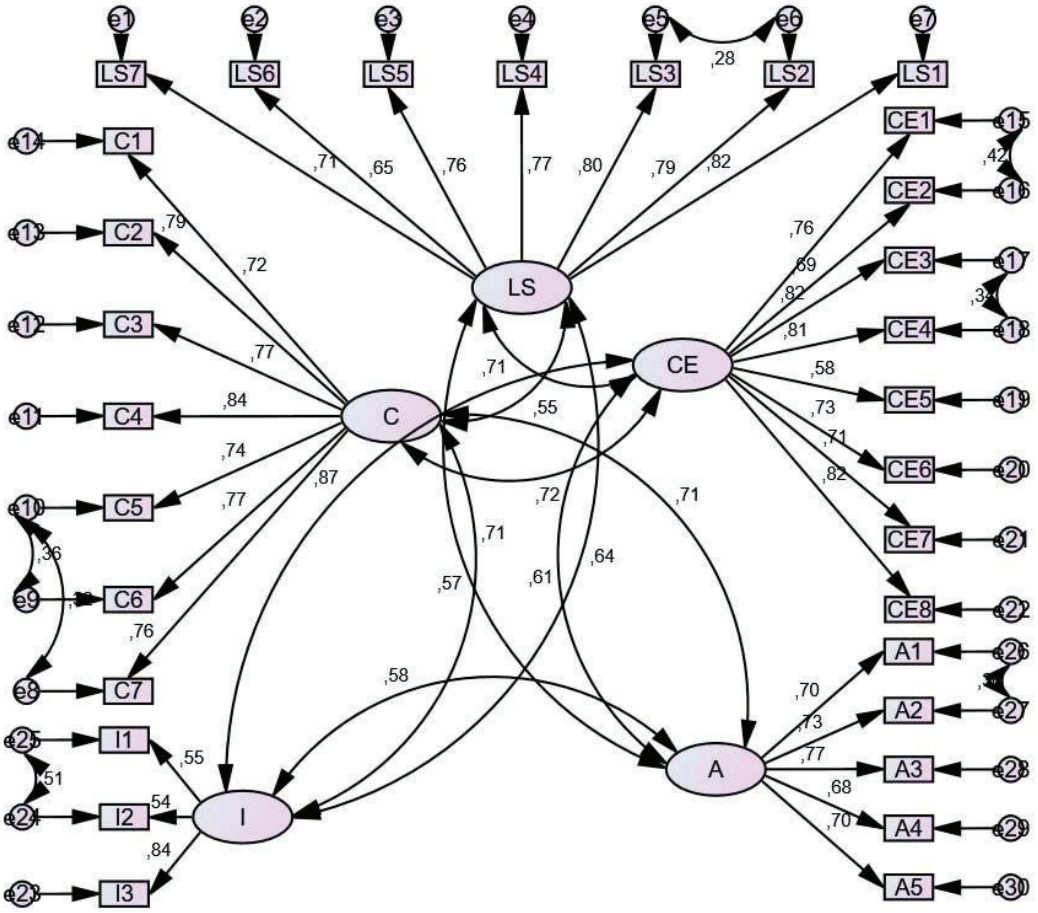

Figure 1. Model fit of DESES

Analysis results should be evaluated according to different fit indices. Researchers emphasize that there is no consensus on which fit indices the model should be evaluated (İlhan & Çetin, 2014). Information on the intervals in which the goodness of fit values explained above show good and acceptable fit is given in Table 3 (Gürbüz, 2019; p. 34).

Model fit criteria	**Good fit**	**Acceptable fit**
x^2/df	$x^2/df \leq 3$	$x^2/df \leq 5$
NFI	$0.95 \leq NFI \leq 1$	$0.90 \leq NFI < 0.95$
TLI	$0.95 \leq TLI \leq 1$	$0.90 \leq TLI < 0.95$
CFI	$0.95 \leq CFI \leq 1$	$0.90 \leq CFI < 0.95$
RMSEA	$0 < RMSEA \leq 0{,}05$	$0.05 < RMSEA \leq 0{,}08$
GFI	$0.90 \leq GFI \leq 1$	$0.85 \leq GFI < 0.90$
PGFI	~ 1	Min. 0,50
RMR	$0 < RMR \leq 0{,}05$	$0 < RMR \leq 0{,}08$

Table 3. Goodness-of-fit indices value

The goodness-of-fit values shared above indicate the goodness-of-fit intervals that the researchers compiled based on different studies (Gürbüz & Şahin, 2018).

Model	Default model	Saturated model	Independence model
RMR	,056	,000	,433
GFI	,844	1,000	,162
AGFI	,813		,104
PGFI	,704		,152
NFI Delta1	,870	1,000	,000
RFI rho1	,855		,000
IFI Delta2	,912	1,000	,000
TLI rho2	,901		,000
CFI	,912	1,000	,000
RMSEA	,067		,214
LO 90	,063		,210
HI 90	,072		,218
PCLOSE	,000		,000

Table 4. RMR, GFI, Baseline Comparisons, RMSEA

When the Table 4 given above is examined, the obtained fit values show that the model fit is achieved. There is no limitation on the values to be looked at in model fit. The reported values may vary according to the values that the researcher wants to draw attention to. The fit values examined below for the current study show that the data fit the model in Table 3 well. The default model is Standardized RMR =,0564; CMIN/DF=2,822<5; RMR=0,056<0,08; 0,90≤IFI=0,912; 0,90≤TLI=0,901; 0,90≤CFI=0,912; RMSEA=0,067<0,08; SRMR=0,0564<0,08.

			Estimate	Standardized Estimate	S.E.	C.R.	P
LS7	<---	LS	1,000	,715			
LS6	<---	LS	,859	,651	,069	12,466	***
LS5	<---	LS	1,103	,762	,076	14,586	***
LS4	<---	LS	1,042	,772	,071	14,770	***
LS3	<---	LS	1,164	,802	,076	15,280	***
LS2	<---	LS	1,185	,791	,079	15,059	***
LS1	<---	LS	1,094	,822	,070	15,716	***
C7	<---	C	1,000	,757			
C6	<---	C	1,178	,775	,074	15,947	***
C5	<---	C	,996	,737	,054	18,309	***
C4	<---	C	1,216	,844	,069	17,588	***
C3	<---	C	1,222	,768	,077	15,792	***

			Estimate	Standardized Estimate	S.E.	C.R.	P
C2	<---	C	1,378	,795	,084	16,426	***
C1	<---	C	1,054	,718	,072	14,652	***
CE1	<---	CE	1,000	,761			
CE2	<---	CE	,892	,693	,048	18,545	***
CE3	<---	CE	,947	,825	,055	17,268	***
CE4	<---	CE	,960	,812	,057	16,937	***
CE5	<---	CE	,543	,580	,047	11,651	***
CE6	<---	CE	,837	,731	,055	15,086	***
CE7	<---	CE	,659	,708	,045	14,532	***
CE8	<---	CE	,820	,821	,048	17,247	***
I3	<---	I	1,000	,840			
I2	<---	I	,431	,537	,042	10,237	***
I1	<---	I	,434	,553	,041	10,555	***
A1	<---	A	1,000	,705			
A2	<---	A	,971	,730	,059	16,415	***
A3	<---	A	1,042	,766	,078	13,321	***
A4	<---	A	,899	,682	,074	12,087	***
A5	<---	A	,847	,698	,069	12,334	***

Table 5. Regression Weights: (Group number 1 - Default model)

Regression values in Table 5 above show the power of observed variables to predict latent variables, that is, factor loadings. Factor loadings are important as the "p" values for each binary relationship above are less than 0.001. The significant p values indicate that the items were loaded correctly on the factors. In addition, the fact that the standardized regression coefficients are 0.537 and larger indicates that the power to predict the latent variables, that is, the factor loadings of each item is high. Since the p values of the covariance, correlation and variance values above were less than 0.01, it was determined that all covariance, correlation, and variance values were statistically significant. As a result of the CFA, it was seen that the validity of the sub-factors that emerged with the EFA was ensured.

4. Discussion

When we investigate the relevant literature, various scales have been used for program evaluation data collection purposes. For instance, In the Vietnamese context, Do Le and Tran (2021) investigated the students' perception of English language training curriculum at a university. One of the data collection tools was a self-designed 40-items Likert type questionnaire. The questionnaire asked questions under six categories. These categories are content, teaching methods, materials, educational objectives, assessment, and expectations from the curriculum. Another program evaluation study (Solihati & Rayahu, 2020) was conducted in Indonesia investigating students' perception about English program. A 23-item questionnaire was used to collect the data and the questionnaire consist of two sections. Out

of 23, 12 questions asked about the course content, and the 11 items questioned the learning process during the semester. Moreover, Canaran et al (2020) investigated an ESP course at three different faculties. In order to collect data, a self-designed 25-items questionnaire was used. The questionnaire was divided into two sections: one contains demographic information, the second section consists of three sub-sections called the program, content of the course and assessment methods. Tunc (2012), in a Master's thesis, used a scale in order to collect data from the English preparatory department. The first section of the scale consists of some demographic questions asking students' department, sex, age, class, average scores, high school type, and parents' educational background. After the demographic section, the researcher divided the rest of the scales into five.

In the first part, there are several statements asking how much emphasis given to each language skills such as reading, writing, listening, speaking, grammar and vocabulary. In the second section, the researcher listed twenty-four statements asking how adequate students feel themselves regarding language skills separately. In the third section, the author investigated the method used in the classes asking eight questions to the participants. Next, assessment is investigated, and six items were listed. In the last section, communication is examined with five items. Moreover, Kirmizi (2011) designed a program evaluation scale in order to evaluate the Master of Arts (MA) programs. After several adaptations, this scale was also used in Bilican (2014) in order to evaluate another MA program. The questionnaire aimed to evaluate the content and program instruction to see whether the program aims fulfills the outcomes. The questionnaire was designed in five sections. These are program outcomes, program content and program instruction such as planning, implementation, and assessment, evaluation of courses in relation to their contribution to learning outcomes, and finally courses that must be included into the program. Totally, under five sections, there are 55items asking the participants' opinions about the program. On the other hand, Karatas and Fer (2007) developed a scale in order to evaluate an English language program at the university level. The questionnaire consists of 46 items and the questions were grouped under context, input, process, and product evaluation of the instruction program. In another study, Tasci (2020) investigated the program outcomes of and English for Specific Purposes (ESP) at a Turkish university. A self-designed 21-item Likert type questionnaire was used in the study. 14 questions were grouped under course objectives sections and 7 questions were grouped under textbooks and other materials section. Another English language program evaluation study was conducted by Nur (2020). The study aimed to find out students' thoughts on the English language courses. In order to collect data, the researcher used a 20-item questionnaire. This questionnaire was categorized into five sections which are teaching techniques, time allocation, materials, communication activity, and importance of the program. Also, Gokdemir (2005) designed a questionnaire in order to find out the problems that English language learners faced during their intensive courses. A 34-items Likert type questionnaire was prepared by the researcher. This questionnaire categorized into three main sections: students' perceptions about intensive English program, teachers' approach to language teaching, and the system of institutional practices.

5. Conclusion

Scales are the most common instruments that are used for program evaluations. Some of these questionnaires are created for the general education purposes while others are designed specifically for the ELT contexts. Moreover, some researchers who are working on ELT adapt the scales designed for the general education purposes into their own fields. By looking at the measurement tools mentioned above, we understand that program evaluation scales consist different dimensions, and these scales are mostly designed in order to evaluate face-to-face education.

DESES is a measurement tool consisting of five sub-dimensions developed to measure the effects of the education received by the students studying in the English preparatory program with the distance education system. The items belonging to the 5 factors obtained from the EFA and appropriate names for these items were given. The first factor consisting of 7 items was called "Language Skill", the second factor consisting of 7 items was called "Communication", the third factor consisting of 8 items was called "Content Evaluation", the fourth factor consisting of 3 items was called "Instructors" and the fifth factor consisting of 5 items was called "Assessment". The Language Skills dimension measures the language skill levels of students. There are 7 items in this sub-dimension. The highest score that can be obtained from this sub-dimension is 35, and the lowest score is 7. A high score indicates a high level of language skill. The Communication dimension measures the communication levels of students. There are 7 items in this sub-dimension. The highest score that can be obtained from this sub-dimension is 35, and the lowest score is 7. A high score indicates a high level of communication. Content Evaluation dimension measures students' content evaluation levels. There are 8 items in this sub-dimension. The highest score that can be obtained from this sub-dimension is 40, and the lowest score is 8. A high score indicates a high level of content evaluation. The Instructors dimension, on the other hand, measures the level of students' perspective towards instructors. There are 3 items in this sub-dimension. The highest score that can be obtained from this sub-dimension is 15, and the lowest score is 3. A high score indicates that students have a high level of perspective towards instructors. Finally, the Assessment dimension measures the general assessment levels of the students. There are 5 items in this sub-dimension. The highest score that can be obtained from this sub-dimension is 25, and the lowest score is 5. A high score indicates that the students' overall assessment level is high. The high Alpha coefficients for the sub-dimensions of the scale (Language Skill=0.906, Communication=0.912, Content Evaluation=0.911, Instructors=0.733, and Assessment=0.850) indicate that the items in the sub-dimensions are consistent with each other. EFA and CFA results also confirmed the validity of the scale. As a result, based on the validity and reliability studies, it can be said that this scale is applicable in measuring the effects of the distance education system of the students studying in the English preparatory program.

During the Covid-19 pandemic, face-to-face education has been turned into online distance education and education programs have been adjusted accordingly. The evaluation phase has gained even more importance in terms of seeing the one-year results of distance education. In this way, distance education programs can be improved in the following years. For this reason, in this study, in order to determine the effects of the distance education system of the students studying in the English preparatory program; a 30-item scale was developed, consisting of 5 sub-dimensions: Language Skill, Communication, Content Evaluation, Instructors, and Assessment. As a result of this research, it is seen that the validity and reliability of the scale were ensured. Also, the scale obtained at the end of this study can also be adapted to the programs that provide education at other levels and subjects and can be used as a tool for evaluation purposes.

References

Akgül A. & Çevik O. (2003). İstatistiksel Analiz Teknikleri, Ankara: Emek Ofset.

Babori, A., Fassi, H. F., Hariri, A., & Bideq, M. (2016). An E-Learning Environment for Algorithmic: Toward an Active Construction of Skills. *World Journal on Educational Technology:Current Issues*, *8*(2), 82-90.

Berge, Z. L., Muilenburg, L.Y., &Haneghan, J.V. (2002). Barriers to distance education and training: Surveyresults. TheQuarterlyReview of DistanceEducation, 3(4), 409-418.

Bilican, R. (2014). Impact of a PhD ELT program on academic development of students at a Turkish state university. *ELT Research Journa,l* 3(3), 111-139.

Brown, J. D. (1995). The elements of language curriculum. Boston: Heinle and Heinle.

Büyüköztürk, Ş.(2006). Sosyal Bilimler İçin Veri Analizi: İstatistik, Araştırma Deseni SPSS Uygulamaları ve Yorum. Ankara: PegemA Yayıncılık.

Canaran, Ö., Bayram, İ., & Altuğ, Ç. (2020). English for Specific Purposes (ESP) Program Evaluation: Perspectives from Three Faculties. *Başkent University Journal of Education*, *7*(1), 20-28.

Chan, V. (2001). Determining students' language needs in a tertiary setting. English Teaching Forum July, 16-27.

Chen, H. T., & Chen, H. T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Sage.

Clark, J. T. (2020). Distanceeducation. In *Clinical engineering handbook* (pp. 410-415). Academic Press.

Coskun, A., & Daloglu, A. (2010). Evaluating an English language teacher education program throughPeacock's model. *Australian Journal of Teacher Education (Online)*, *35*(6), 24-42.

Coşkun, A. & Mutlu H.T. (2017). Investigating High School Students' Use of Extramural English: A Scale Development Study, İnsan ve Toplum Bilimleri Araştırmaları Dergisi 6(1), 571-590.

Cunningham, P. D. (2017). Bridging the distance: Using interactive communication tools to make online education more social. *Library trends*, *65*(4), 589-613.

Davie, L. (1995). Times of turbulence and transition: needed research in computermediated-communication (CMC) course design. In E. Wagner & M. Koble (Eds.), Distance education symposium 3: coursedesign. Selected papers presented at the thirdi nternational distance education research symposium. ACSDE research monograph #14. University Park, PA: The Pennsylvania State University.

Denzin, N. K., & Lincoln, Y. S. (2000). Handbook of qualitative research. London: Sage Publications.

Dorrego, E. (2016), Distance Education and Learning Evaluation, *Revista De Educacion a Distancia*, 50,1-18

Do Le, T., & Tran, T. N. T. (2021). English Language Training Curriculum: An Evaluation From Learner's Perceptions. *Open Journal of Social Sciences*, 9(3), 40-57.

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

Gorsuch, R. L. (2008). *Factor analysis (Second Edition)*. New York: Psychology Press.

Gökdemir, C. V. (2005). Üniversitelerimizde verilen yabancı dil öğretimindeki başarı durumumuz. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *6*(2), 251-264.

Gunawardena,C ., Lowe , C., & Carabajal,K. (2000). Evaluating online learning :models and methods. Paper included in the proceedings of the eleventh society for Information Technology and Teacher Education International Conference. SanDiego, CA.

Gürbüz, S. (2019). *Aracı, Düzenleyici ve Durumsal Etki Analizleri*. Seçkin Yayınevi, Ankara.

Gürbüz, S. ve Şahin, F. (2018). *Sosyal Bilimlerde Araştırma Yöntemleri – Felsefe – Yöntem – Analiz*. 5. Baskı. Seçkin Yayınları, Ankara.

Hair J. F.,Tatham R. L., Anderson R. E. & Black W. (1998). *Multivariate Data Analysis*, New York: PrenticeHall.

Hsu, L. (2014). Effectiveness of English for specific purposes courses for non-English speaking students of hospitality and tourism: a latent growth curve analysis. *Journal of Hospitality, Leisure, Sport & Tourism Education, 15,* 50-57.

İlhan, M., & Çetin, B. (2014). LISREL ve AMOS programları kullanılarak gerçekleştirilen yapısal eşitlik modeli (yem) analizlerine ilişkin sonuçların karşılaştırılması. *Journal of Measurement and Evaluation in Education and Psychology*, *5*(2), 26-42.

Jenkins, J., Baker, W., &Dewey, M. (Eds.). (2017). *The Routledge handbook of English as a lingua franca*. Routledge.

Kalaycı, Ş. (2010). *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri,* Ankara: Asil Yayınları.

Karagöz, Y., Mutlu, H. T., Sağır, S., & Celil, M. (2019). Kargo Şirketi Seçimini Etkileyen Faktörlerin Belirlenmesine Yönelik Ölçek Geliştirilmesi: Sivas Örneği. *Ekonomik ve Sosyal Araştırmalar Dergisi*, *15*(2), 235-247.

Karataş H. & Fer S. (2009). Evaluation of English Curriculum at Yıldız Technical University Using CIPP Model. *Egitim ve Bilim*, 34(153), 47-60.

Khoirunnisa, S. & Supriyadi, S. (2018). ESP teacher's ands tudent's perceptions on teaching speaking for a tourism program. *Dinamika Ilmu, 18*(1), 37-48.

Kieyle, R., & Rea-Dickens, P. (2005). *Program Evaluation in Language Education*. Basingstoke: Palgrave Macmillan

Kırmızı, Ö. (2011). An evaluation of M.A. ELT programs in Turkey. Unpublished Doctoral Dissertation, Hacettepe University, Ankara, Turkey.

Kroll, R., Schoen, D. (2015), A Holistic Evaluation Approach for Degree Courses in Engineering Sciences For Distance Learning Universities, 16th International Conference on Research and Education in Mechatronics (REM), 1 – 5.

Lunenburg, F. C. (2011). Keycomponents of a curriculum plan: Objectives, content, andlearning experiences. *Schooling*, *2*(1), 1-4.

Lynch, B. K. (1996). *Language Program Evaluation*: Theory and Practice. Cambridge: CUP.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods,* 4(1), 84-99.

McNulty, T. (2013). Standardized Clients in Case-Based Simulation: Five Years of Development in an Occupation-BasedCurriculum. The American Occupational Therapy Education Summit, Atlanta, October 2013.

Moore, R. L., & Fodrey, B. P. (2018). Distance education and technology infrastructure: Strategies and opportunities. In *Leading and managing e-learning* (pp. 87-100). Springer, Cham.

Murphy, D. F. (2000). Key Concepts in ELT, Evaluation. *ELT Journal*, 54(2), 210- 211.

Nakip, M. (2006). *Pazarlama Araştırmaları: Teknikler ve SPSS destekli uygulamalar*. Ankara: Seçkin Yayıncılık.

Norris, J. M. (2009). Understanding and improving language education through program evaluation: Introduction to the special issue. *Language Teaching Research,*13, (1), 7–13

Nur, R. (2020). Students' Perception On TheI mplementation Of Intensive English Course For Freshmen At Universitas Indonesia Timur: A Program Evaluation. *JISIP (Jurnal IlmuSosial dan Pendidikan)*, *4*(3),545-558.

Peacock, M. (2009). Theevaluation of foreign language teacher education programmes. Language Teaching Research, 13(3), 259-278.

Petsuwan, S., Pimdee, P., & Pupat, P. (2019). Impacts of DistanceEducationSystem on Teacher Competency of Remote Schools in Lower Northern Thailand. *Mediterranean Journal of SocialSciences*, *10*(3), 41-47.

Rao, S. M., & Krishnan, V. (2015). Distance education. Fourth International Conference on Higher Education: Special Emphasis on Management Education, Nitte University, December 29-30, 2014.

Rovai, A. P. (2003). A practical framework for evaluating online distance education programs. *The Internet and Higher Education*, *6*(2), 109-124.

Sae-Khow, J. (2014). Developing of Indicators of an E-Learning Benchmarking Model for Higher Education Institutions. *Turkish Online Journal of Educational Technology-TOJET*, *13*(2), 35-43.

Saito, H., &Ebsworth, M. E. (2004). Seeing English Language Teaching and Learning through the Eyes of Japanese EFL and ESL Students. Foreign Language Annals, 37, 111- 124.

Sand, O., Davis, D., Lammel, R., & Stone, T. (1960). Chapter III: Components of the Curriculum. *Review of Educational Research*, *30*(3), 226-245.

Swart, A. J. (2016), Distance Learning Engineering Students Languish Under Project-Based Learning, But Thrive in Case Studies and Practical Workshops, IEEE Transactions On Education, 59(2), 98–104.

Tom-Lawyer, O. (2014). An Evaluation of the English Language Curriculum of the Nigeria Certificate in Education: A Case Study of a College of Education. *Open Journal of Social Sciences,* 2, 69-79.

Tunç, F. (2010). *Evaluation of an ELT Program at a Public University Using CIPP Model*. Saarbrucken: Lambert Academic Publishing.

Velicer, W. F. & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. Psychological Methods, 3(2), 231-251.