



## TEOG (TEPSE) English Test: Content Validity and Teachers' Views

Ayşenur UZUN<sup>\*a</sup>, Ferit KILIÇKAYA<sup>b</sup>

### Article Info

DOI: 10.14686/buefad.732132

Article History:

Received: 04.05.2020

Accepted: 28.09.2020

Published: 05.10.2020

Keywords:

Content Validity,  
English Test,  
TEPSE.

Article Type: Research Article

### Abstract

One of the most important dynamics of the educational setting all over the world is examinations, and some of those are English tests. In Turkey, English tests for the students preparing for the high schools were included in the national exams with the Level Determination Exam (SBS) in 2008 for the first time, and Transition Examination from Primary to Secondary Education (TEPSE) replaced them in 2013. The effects of these exams have been investigated; however, to the best knowledge of the author, other principles of language assessment such as validity and reliability have not been investigated based on the analysis of documents. This study, therefore, aimed to investigate content validity of the English tests in TEPSE, which were conducted between 2016 and 2017. Data collection procedure included the analyses of TEPSE English tests, the coursebook, the curriculum of 8<sup>th</sup> grade and semi-structured interviews held with 21 English language teachers. The analyses of the frequently used items in the coursebook and table of specifications were used to examine the content validity. The results revealed that despite some factors, TEPSE English tests between 2016 and 2017 have content validity based on the alignment between the coursebook and TEPSE English tests.

## TEOG İngilizce Sınavı: İçerik Geçerliliği ve Öğretmen Görüşleri

### Makale Bilgisi

DOI: 10.14686/buefad.732132

Makale Geçmişi:

Geliş: 04.05.2020

Kabul: 28.09.2020

Yayın: 05.10.2020

Anahtar Kelimeler:

İçerik Geçerliliği,  
İngilizce Testi,  
TEOG.

Makale Türü: Araştırma  
Makalesi

### Öz

Tüm dünyada eğitimin en önemli dinamiklerinden bir tanesi sınavlardır ve bunların bazıları İngilizce sınavlarını da içermektedir. Türkiye’de, liseye hazırlanan öğrenciler için uygulanan İngilizce sınavlarına ilk kez 2008 yılında SBS’de yer verilmiş ve 2013 yılında bu sınav TEOG ile yer değiştirmiştir. Bu sınavların etkileri incelenmiştir fakat -bilindiği kadarıyla- geçerlilik ve güvenilirlik gibi dil değerlendirmesinin diğer ilkeleri araştırılmamıştır. Bu yüzden, bu çalışma 2016-2017 döneminde uygulanan TEOG İngilizce testlerinin içerik geçerliliğine ilişkin öğretmen görüşlerinin yanı sıra bu dönemlerde yapılan testlerin içerik geçerliliğini araştırmayı hedeflemiştir. Veri toplama yöntemi, 2016-2017 döneminde yapılan TEOG İngilizce sınavlarının analizini, ders kitabının analizini, 2016-2017 8. Sınıf müfredatını ve 21 İngilizce öğretmeniyle yapılmış olan yarı yapılandırılmış görüşmeyi içermektedir. Ders kitabında sık kullanılan yapıların analizi ve belirtke tablosu içerik geçerliliğini değerlendirmek için kullanılmıştır. Sonuçlar, bazı uyumsuzluklar, eşit olmayan dağılımlar ve dört dil becerisinin ihmeline rağmen 2016-2017 döneminde yapılan TEOG İngilizce sınavlarının ders kitabı ile uyumu açısından içerik geçerliliğinin olduğunu ortaya koymuştur.

\*Corresponding Author: aysenuruzunn@gmail.com

<sup>a</sup> Teacher, Ministry of National Education, <https://orcid.org/0000-0003-3701-7140>

<sup>b</sup> Prof. Dr., Mehmet Akif Ersoy University, Faculty of Education, Burdur, <https://orcid.org/0000-0002-3534-0924>

## Introduction

Language assessment is one of the dynamics of educational settings; therefore, it is a critical issue in language teaching to test the students' performance on the determined content. Brown (2004) voiced the traditional and alternative assessments as two ways of assessments, and traditional assessment might be summarized as standardized tests including generally multiple-choice format which focused on 'right' answers. Külekçi (2016) also revealed language proficiency exams in the world such as TOEFL, IELTS, PTE and GEPT; however, there were also language proficiency exams conducted in Turkey such as KPDS, UDS, and YDS, which is still in use. While the language exams 'KPDS', 'UDS', and 'YDS' were proficiency examinations, there were some English tests for the secondary school students, which were a must to attend a high school. These English tests started to be included in the national exams in 2008 in Turkey. Even though this system started with SBS (MoNE, 2011), it was followed by TEPSE (Transition Examination from Primary to Secondary Education) in 2013 which was also replaced by LGS (Transition Examination to High Schools) in 2018 (MoNE, 2018) in the course of conducting the current study.

Language assessment and its importance were emphasized by many scholars such as Brown (2004), Hughes (2003), Ekbatani (2011), and Solano-Flores (2016). One of the principles of language assessment "validity", and Brown and Abeywickrama (2010) also indicate that the sub-components of the validity are content-related evidence, criterion-related evidence, construct-related evidence, consequential validity, and face validity. Even though the other four types of validity are also important, content validity is the key concept of this study. Hughes (2003) states that "A test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc. with which it is meant to be concerned" (p. 26). Ekbatani (2011) asserts that content validity is consistence between objectives/functions of the test and the test itself. Hughes (2003) also indicates the importance of content validity by focusing on two important reasons. The first reason is that having a content validity provides an accurate measurement and guarantees the construct validity. The second reason is that the lack of content validity results in harmful washback effect. If the content of lesson does not match the content of the test, the learning and teaching are affected negatively. The importance of 'content validity', one of the basic principles of language assessment was also underscored by Brown and Abeywickrama (2010), Ekbatani (2011), and Gorsuch and Griffee (2018).

Content validity plays a crucial role in assessment as Hughes (2003) mentioned before. Based on the presence of language exams, content validity of such exams has been drawing researchers' attention for many years, and many researchers focused on this issue in their studies (Alderson & Kremmel, 2013; Al-Adawi & Al-Balushi, 2016; Haiyan & Fuqin 2005; Nicholson, 2015; Razmjoo & Tabrizi, 2010; Siddiek, 2010). For example, Siddiek (2010) investigated content validity of Sudan School Certificate English examination based on the alignment between the coursebook and the exam, and emphasized that this alignment increases content validity of the test. Nicholson (2015), who analyzed the TOEIC exam in Korea, found that content validity of the exam was weak because it did not test real communicative language skills. Even though the results of the several studies reviewed indicated low or lack of content validity of language exams, other studies (Ing et al., 2015; Jaturapitakkul, 2013; Kang & Chang, 2014; Külekçi, 2016) indicated high content validity in several other language tests. Ing et al. (2015), Kang and Chang (2014), and Sim (2015) focused on the exams which were content valid besides the using the table of specifications. Moreover, there were also some studies (Gömlüksiz & Aslan, 2017; Ökmen & Kılıç, 2016; Kılıçkaya, 2016; Vural, 2017) conducted on TEPSE English tests, which investigated teaching methods, students' views, and teachers' views from different perspectives. Vural (2017) investigated content validity of English tests in TEPSE in 2014 by only taking the views into consideration, while Gömlüksiz and Aslan (2017) investigated the students' perspectives towards TEPSE English tests conducted between 2016 and 2017.

It was claimed that content validity would be affected negatively if the tests were not able to measure communicative competence (Al-Adawi & Al-Balushi, 2016; Haiyan & Fuqin, 2005; Nicholson, 2015; Siddiek, 2010). Moreover, inconsistency between the coursebook and the test, and unequal distribution of the items also cause lack of content validity (Razmjoo & Tabrizi, 2010; Siddiek, 2010). However, if there is a consistency among the coursebook, curriculum, and tests, it means that these tests have content validity as it was emphasized in the studies of Aksan (2001), Jaturapitakkul (2013), and Kang and Chang (2014). Although there are several studies conducted on content validity of exams in other countries and also in Turkey, to the best knowledge of the author, there is not any detailed study conducted on content validity of TEPSE English tests between 2016

and 2017. As it was emphasized by many researchers, content validity of a test is very crucial to test students' performances on the intended area. Besides, the researcher was teaching English to 8<sup>th</sup> graders in those years and noticed some problems in the tests regarding content validity. Therefore, the researcher of the current study focused on content validity of TEPSE English tests conducted between 2016 and 2017. Moreover, content validity of TEPSE English tests in those years has not been investigated based both on the documents and on the teachers' views. For this purpose, this study aimed to investigate content validity of English tests in TEPSE between 2016 and 2017 by analyzing the items in the tests, the coursebook, and the curriculum on which the language tests were. Besides, the interviews held with teachers were also analyzed to reveal their views on content validity of TEPSE English tests.

In order to reach these aims, the research questions used in this study are listed as follows:

- 1- To what extent do the English tests in TEPSE conducted between 2016 and 2017 have content validity?
  - a) Do the English tests in TEPSE exactly focus on the frequently used items in the coursebook "Upturn in English"? If yes or no, which items are tested or not tested?
  - b) Is there an exact match between functions of the provided syllabus and the questions in the English language test in TEPSE?
- 2- What are English language teachers' views on content validity of English tests in TEPSE?

## **Method**

### **Research Design**

In order to obtain information about content validity of English tests in TEPSE, mixed method research design was utilized. The data collection methods of a qualitative study are observations and interviews (Creswell, 2009). In this study, semi-structured interview and documents which are TEPSE English tests and the coursebook "Upturn in English 8" were used to collect the data. Interview is a way of collecting qualitative data by asking questions to the interviewee, and can be conducted in many ways such as face to face, via phone, and internet (Christensen et al., 2015). The interviews in this study were held with 21 English language teachers teaching English to 8<sup>th</sup> grade students between 2016 and 2017. On the other hand, another way of collecting data is documents. Creswell (2009) stated that documents are beneficial for collecting data, and the public and private documents might be used. From this perspective, this study used the coursebook "Upturn in English" provided by MoNE and TEPSE test questions between 2016 and 2017 as documents, and the researcher benefited from quantitative analysis to compare the frequency of vocabulary items in the coursebook and the test questions. The reasons to analyze the frequency of vocabulary items in the coursebook rather than the curriculum are that in Turkey, the English curriculum is realized with the coursebooks, and neither the coursebook nor TEPSE English tests in those years included some of the vocabulary items suggested by the English curriculum. Furthermore, the table of specifications provided by Newman et al. (1973 as cited in Newman et al., 2013) was adapted and used to analyze TEPSE English tests questions based on the functions. After collecting the data from the interviews and the documents, the researcher compared the results and made interpretations on content validity of TEPSE English tests between 2016 and 2017.

### **Population and Sample/Study Group/Participants**

As one of the ways of sampling procedure, the purposeful sampling strategy was used to select the participants of the interviews, which intends to select participants based on specific criteria (Lochmiller & Lester, 2017). The participants of this study were 21 English language teachers teaching English to 8<sup>th</sup> grade students in Burdur, Afyonkarahisar, Ağrı, Istanbul, Ardahan, Antalya, Ankara, Hakkari, Şırnak, Elazığ, Konya, Denizli, Erzurum, Kilis, İzmir, Kocaeli, Samsun, and Van. The results of TEPSE conducted in 2016-2017 could not be taken into consideration while choosing the provinces because, to the best knowledge of the author, MoNE did not announce the whole list of each province's results in TEPSE. Therefore, the researcher asked for the volunteer teachers teaching English to 8<sup>th</sup> grade students through TEPSE groups on social media. Based on

this announcement, 21 English language teachers teaching English to 8<sup>th</sup> grade students from eighteen different provinces accepted being included in the current study, and required permissions were given by MoNE.

**Table 1.** The Ranks of Selected Provinces

Province	Rank	Province	Rank	Province	Rank
Burdur	3	İzmir	43	Ardahan	66
Denizli	9	Elazığ	50	Kilis	69
Antalya	11	Konya	51	Van	77
Ankara	14	İstanbul	55	Hakkari	78
Kocaeli	30	Afyon	58	Ağrı	79
Samsun	41	Erzurum	65	Şırnak	80

Table 1 presents the ranks of the selected provinces based on the average point of placement basic scores of the system for TEPSE (TUIK, 2016). The ranks of the determined provinces based on average placement scores for TEPSE (TUIK, 2016) were worthy of notice, which vary from the highest scores to lowest. When these provinces were considered, samples from seven regions of Turkey were included in the current study.

#### Data Collection Tools

The current study is a mixed method study, and benefited qualitative and quantitative data collection. Creswell (2012) listed the types of qualitative data as observations, interviews, and audio-visual materials. This study used semi-structured interviews and documents to collect data. The coursebook “Upturn in English” and TEPSE English tests between 2016 and 2017 were compared based on quantitative analysis, and interviews were held with 21 English language teachers teaching English to 8<sup>th</sup> grade students.

*Documents.* Documents are the way of collecting data, and they might be either personal or official (Christensen et al., 2015). In this study, the document were the official ones such as the coursebook “Upturn in English” provided by MoNE and TEPSE English test questions between 2016 and 2017. Table of specification was used to analyze content validity of TEPSE English tests, and the coursebook ‘Upturn in English’ which is provided by MoNE was analyzed in terms of frequency of the language use and vocabulary items by using quantitative analysis. In Turkey, the curricula are realized with coursebooks, and neither the coursebook nor TEPSE English tests in those years included some of the vocabulary items suggested by curriculum. Therefore, analyzing the frequency of vocabulary items in the coursebook and TEPSE English tests is another dimension of this study. Biemiller (2003) stated that vocabulary is important to determine the success in reading skill. As mentioned before, TEPSE is a standard multiple choice-test and based on reading skill; therefore, the researcher intended to analyze and compare the frequency of vocabulary items in the coursebook “Upturn in English” and in TEPSE English tests to determine the alignment between them. This alignment or misalignment is useful to decide on to what extent TEPSE English test has content validity.

*Semi-structured interviews.* Interview, a way of collecting qualitative data, is defined as asking questions to the people what they think about on a determined issue, and it helps researcher to check whether the previously gathered data are parallel to views of the participants of the study (Fraenkel, Wallen, & Hyun, 2012). Creswell (2012) states that one-on-one interview is the one which is also called as individual interview, and is a way in which the questions are asked to the only one person at a time. He also defines the telephone interviews as being a meaningful way of collecting data from the participants of the study who live in different or distant places; therefore, asking the questions via telephone might be possible to collect data (Creswell, 2012). Therefore, the researcher conducted face to face, one-on-one interviews and telephone interviews to collect data. Semi-structured interviews were conducted in this study since semi-structured interviews are more flexible according to Gas and Mackey (2012). The interviews which were held with the teachers in Turkish is a 3-question semi

structured interviews that might provide opportunities to gather detailed and wide range of information on the issue.

### Data Collection

**Table 2.** The Steps in Data Collection Procedure

Steps	Actions taken
Analysis of annual plan and syllabus	Determining the functions of the units
	Making the list of determined language use patterns
Analysis of coursebook	Making the list of language use patterns
	Making the frequency list of vocabulary items
Analysis of TEPSE English tests	Categorizing questions by using table of specifications
	Making the list of language use items
Interviews	Making the frequency list of vocabulary items
	Semi-structured interviews held with 21 English language teachers.

As it can be seen in Table 2, the annual plan and syllabus provided by MoNE were analyzed based on the functions of the units and determined language use patterns. Then, the coursebook ‘Upturn in English’ was analyzed to make a frequency list which included the frequency of vocabulary items and language use patterns that were used in the coursebook. After making the list, the provided syllabus of 8<sup>th</sup> grade English subject was taken into consideration to make a table of specifications. For the first term TEPSE English test, the functions of the first three units in the plan were used to make the table of specification while the functions of the first eight units were for the second term TEPSE English test. The reason of choosing functions of first three and first eight units was that students were responsible for the first three units in the first term TEPSE while they were responsible for the first eight units in TEPSE English test which were conducted in the second term (MoNE, 2016).

After making the table of specification, the English tests in TEPSE were analyzed based both on table of specifications and frequency list which included the numbers of vocabulary items and language use patterns. Moreover, the semi-structured interviews with 21 English language teachers were conducted. Before conducting these interviews, the coursebook, TEPSE English tests, plan provided by MoNE, and the main interview questions were sent via e-mail to the participants.

The following questions were asked in the interviews:

- 1- Do the English language tests in TEPSE exactly focus on the frequently used items in the coursebook? If yes or no, which items are tested or not tested?
- 2- Is there an exact match between functions of the provided syllabus and the questions in the English language test in TEPSE?
- 3- Do you have any comments?

### Data Analysis

The data collected from semi-structured interviews, analyses of exam questions based on table of specifications, and frequency list were used to determine content validity of English tests in TEPSE. Comparative analysis was utilized in order to analyze the coursebook and TEPSE English tests based on the vocabulary items and language patterns. Therefore, the word frequency lists were prepared by analyzing the coursebook and TEPSE English tests based on the vocabulary items and language use patterns. While preparing



these words frequency lists, the researcher used software called as 'Word Frequency Counter' (Pterneas, 2009). The coursebook "Upturn in English" was turned into a word document and modified to differentiate some words from each other such as writing the auxiliary verb "do" as "doyou" with a pronoun and writing action verb "do" as a separate item. After such modifications, documents were copied into the program 'Word Frequency Counter' (Pterneas, 2009), and this program provided the words with their frequency numbers. The lists provided by the program were revised by focusing on the content words and some function words such as conjunctions and some auxiliary verbs indicating tenses. The final version of the lists provided the frequency of the 'language use patterns' and vocabulary items that the coursebook included. The same steps were followed in the analyses of TEPSE English tests between 2016 and 2017, and the results of the frequently used items in the coursebook and in the exams were compared. In this analysis, 'language use patterns' in the coursebook focusing on the functions provided by MoNE (2016), the suggested annual plan provided by Antalya Provincial Directorate of National Education (2016), and the word frequency lists prepared by the researcher were used. Based on these documents, the coursebook was analyzed and 'language use patterns' were determined.

According to Davidson and Lynch (2002), table of specifications is a useful way to construct a test and includes great deal of information such as skills, subskills, number of items, desired score weighting, and special materials. Table of specification is a chart including the topics, objectives and the number of questions in the test. Moreover, Cheng and Fox (2017) emphasize the importance of developing a table of specifications, which is helpful for creating high quality tests. The reason of choosing table of specification to analyze the test questions was that a specification of the skills is a must to determine whether a test has a content validity or not (Hughes, 2003). The analysis of TEPSE English tests was conducted by considering the example of table of specification provided by Newman et al. (1973 as cited in Newman et al., 2013). The used table of specification based on the syllabus of 8<sup>th</sup> grade was developed by focusing on the functions of predetermined units which the students were responsible for in TEPSE English tests. The exam questions which were 40 in total were analyzed one by one by using the table of specifications, and the function of each question was determined with the help of table of specifications. Furthermore, the recorded interviews were transcribed by the researcher. In this study, descriptive codes which indicate main topics were used to analyze the interview data. Griffiee (2012) stated that reliability is comparing the consistency between two raters based on the assigned codes. Therefore, this classification and codes were revised once again and discussed with an experienced researcher to be sure about the reliability after the coding procedure. Miles and Huberman (1994 as cited in Griffiee, 2012) provided a formula to calculate the reliability, and this formula might be given as dividing the number of agreements into the number of the agreements plus disagreements. The codes assigned by the other rater were obtained and compared with the codes assigned by the researcher, and the reliability was calculated as 0.96. As a final analysis, findings on analysis of the coursebook, analysis of TEPSE English tests and interviews were compared, which might provide a better understanding on content validity of English test in TEPSE.

### **Research Ethics**

The required permission from Ministry of National Education was obtained for data collection which was conducted between October 2017 and June 2018.

## **Findings**

### **Content Validity of English Tests in TEPSE Conducted in 2016-2017**

In this section, the coursebook "Upturn in English" was analyzed based on the frequently used 'language use patterns' and vocabulary items. In Turkey, the curricula are realized with coursebooks, and neither the coursebook nor TEPSE English tests in those years included some of the vocabulary items suggested by curriculum. Therefore, it was considered crucial to compare the frequency of vocabulary items in the coursebook and TEPSE English tests. Moreover, this section also presents the results on whether there is an alignment between functions of the provided syllabus and the questions in the exam.

### **Alignment between the English Tests in TEPSE and the Coursebook "Upturn in English" Based on the Frequently Used Items**

In this section, TEPSE English test questions in 2016-2017 were analyzed based on the frequently used vocabulary items and language use patterns, and the details were presented in the following sub-sections.

### Analysis of the 2016-2017 TEPSE English Tests Based on Frequency of Vocabulary and Language Use Items

In the analysis of the 2016-2017 TEPSE English tests, the same methods which were used to analyze the coursebook ‘Upturn in English’ were followed by the researcher. With the help of the software ‘Word Frequency Counter’ (Pterneas, 2009), the following lists which provide the top 30 frequently used items were presented, and they helped researcher to decide to what extent TEPSE English tests had content validity. Top 30 frequently used items in the 2016-2017 1<sup>st</sup> term TEPSE English test and in the coursebook were determined. When the frequency of these items in the coursebook was taken into consideration, the following table (Table 3) might be beneficial to decide whether the frequently used items in the coursebook were presented in the exam or not.

**Table 3.** The Frequency List of Top-30 Items Based on TEPSE English Test (2016-2017 1<sup>st</sup> Term) and the Coursebook

Items	E1	C1	E2	C2	E3	C3	E.t	Ct	Items	E1	C1	E2	C2	E3	C3	E.t	C.t.
does...?	7	26	6	40	0	9	13	75	but	4	15	0	10	0	2	4	27
vegetable	0	0	0	0	9	14	9	14	can	2	5	0	5	2	5	4	15
eat	2	3	0	3	6	7	8	13	home	4	4	0	1	0	1	4	6
pizza	0	0	0	1	8	8	8	9	put	0	0	0	0	4	22	4	22
like	4	11	0	17	3	6	7	34	meat	0	0	0	0	4	3	4	3
watch / movie	7	14	0	2	0	0	7	16	music	3	4	1	6	0	0	4	10
friend	7	12	0	7	0	1	7	20	sure	3	3	0	0	1	0	4	3
movie	6	21	0	4	0	0	6	25	then	2	6	0	2	2	9	4	17
minutes	0	0	0	0	5	22	5	22	Where	2	5	2	3	0	0	4	8
would like to	4	39	1	4	0	0	1	43	always	2	2	0	5	2	1	4	8
action	5	2	0	0	0	0	5	2	fry	0	0	0	0	3	13	3	13
and	1	14	0	41	4	73	5	128	going to(t)	2	52	1	1	0	0	3	53
fix	0	0	5	3	0	0	5	3	refuse	3	8	0	0	0	0	3	8
prefer	2	0	1	11	2	0	5	11	study/ exam	3	10	0	2	0	0	3	12
dislike	3	1	0	0	1	0	4	1	What	2	19	0	16	1	6	3	41
favorite	2	5	1	7	1	5	4	17	Why	0	3	1	3	2	1	3	7

Table 3 provides the frequency list of top-30 items in the 2016-2017 1<sup>st</sup> term TEPSE English test and in the coursebook. The number of frequency of items in the coursebook can be seen based on the units as ‘C1, C2, etc.’, while the number of frequency of items in the exam can be seen based on the units as ‘E1, E2, etc.’. Table 3 presents the findings on the alignment between the coursebook and the test. When the analysis of the coursebook “Upturn in English” based on the frequently used items and the 2016-2017 2<sup>nd</sup> term TEPSE English test were compared, the results could be presented in Table 4.

**Table 4.** The Frequency List of Top-30 Items Based on TEPSE English Test (2016-2017 2<sup>nd</sup> Term) and the Coursebook

Items	E1	C1	E4	C4	E5	C5	E6	C6	E7	C7	E8	C8	E.t	C.t
friends	0	12	2	14	8	11	1	0	1	1	0	1	12	39
does ..?/do..?	2	26	4	17	3	50	0	33	0	21	0	10	9	157
extreme sports	0	0	0	0	0	0	8	16	0	0	0	0	8	16
net	0	0	1	0	7	8	0	0	0	0	0	0	8	8
use	0	1	1	9	7	28	0	2	0	0	0	2	8	42
and	0	14	0	20	3	30	5	34	0	42	0	65	8	205
Has/have	2	18	0	7	3	17	0	2	0	8	2	8	7	60
prefer	0	0	1	3	2	0	3	18	0	11	0	0	6	32
all	0	0	0	5	3	4	2	5	0	0	1	8	6	22
enjoy	1	4	0	0	0	1	5	6	0	1	0	3	6	15
I think	1	1	0	0	1	1	4	10	0	0	0	2	6	14
like	0	11	0	4	1	0	4	9	0	4	0	13	5	41
more/-er than	0	1	2	0	0	0	3	27	2	8	0	0	5	36
as	0	1	0	1	2	2	2	1	0	3	0	1	4	9
do	1	19	0	5	2	11	1	23	0	3	0	10	4	71
go	3	5	0	4	0	5	0	3	1	10	0	0	4	27
internet	0	0	1	0	3	47	0	0	0	0	0	0	4	47
located	0	0	0	0	0	0	0	0	4	2	0	0	4	2
most	0	1	1	2	1	6	2	1	0	5	0	0	4	15
sports	0	2	0	0	0	1	4	18	0	1	0	0	4	22
why	1	3	2	5	0	2	1	10	0	6	0	2	4	28
adrenalin seeker	0	0	0	0	0	0	4	6	0	0	0	0	4	6
never	0	1	0	4	3	4	0	0	0	0	0	0	3	9
usually	0	0	1	7	2	14	0	4	0	1	0	6	3	32
visit	1	8	0	1	0	2	0	0	2	14	0	2	3	27
Were/was	0	1	0	0	0	0	0	3	3	15	0	2	3	21
What time	2	1	0	2	1	1	0	0	0	0	0	0	3	4
always	0	2	0	2	3	3	0	0	0	0	0	4	3	11
because	0	4	0	5	1	6	2	13	0	5	0	4	3	37
come	1	14	0	12	0	1	0	0	1	3	1	3	3	33

As it can be seen in Table 4, 2016-2017 2<sup>nd</sup> term TEPSE English test does not include any questions on Unit-2 and Unit-3; therefore, the table above does not have any columns on these units. However, Table 4 indicates that except some of the items, most of the frequently used items in the test were also presented frequently in the coursebook.



**Table 5.** Assigned Numbers to the Functions

Units	Functions	Assigned Number to The Functions	Units	Functions	Assigned Number to The Functions
UNIT-1 Friendship	Accepting and refusing	1	UNIT-2 Teen life	Describing the frequency of actions	6
	Apologizing	2		Expressing likes and dislikes	7
	Giving explanations/ reason	3		Expressing preferences	8
	Making simple inq.	4		Making simple inq.	9
	Telling the time, days and dates	5		Stating person opinions	10
UNIT-3 Cooking	Describing simple processes	11	UNIT-4 Communication	Expressing concern and sympathy	15
	Expressing preferences	12		Handling phone conversation	16
	Making simple inquiries	13		Making simple inquiries	17
	Naming common objects	14		Talking about plans	18
UNIT-5 Internet	Accepting and refusing	19	UNIT-6 Adventure	Expressing preferences	26
	Giving explanations/ reason	20		Giving explanations/ reasons	27
	Making excuses	21		Making simple comparison	28
	Making simple requests	22		Making simple inquiries	29
	Making simple inquiries	23		Stating personal opinions	30
	Talking about plans	24		Talking about what people do regularly	31
Telling the time, days and dates	25	Talking about past events	32		
UNIT-7 Tourism	Describing places	33	UNIT -8 Chores	Expressing feelings	40
	Describing weather	34		Expressing likes/dislikes	41
	Expressing preferences	35		Expressing obligation	42
	Giving explanations/ reason	36		Giving explanations / reasons	43
	Making simple comparisons	37		Making simple inquiries	44
	Stating personal opinions	38		Making simple suggestions	45
	Talking about past events	39			

### Alignment between Functions of the Provided Syllabus and the Questions in the English Tests in TEPSE

In this section, English tests in TEPSE conducted in 2016-2017 were analyzed by adapting table of specification to decide on to what extent these exams had content validity or not.

### Content Validity of English Tests in TEPSE between 2016 and 2017

The questions of 2016-2017 1<sup>st</sup> term TEPSE English test were published by MoNE (2016b), and the questions of 2016-2017 2<sup>nd</sup> term TEPSE English test were published by MoNE (2017). The questions were revised and summarized based on the functions of each unit, which MoNE (2016) provided. The example of table of specification was provided by Newman et al. (1973 as cited in Newman et al., 2013). This provided table of specification was adapted and used in this study to analyze the test questions. The table of specification in this study includes not only the functions but also the numbers of the functions; therefore, numbers were assigned to each function of the units. The assigned numbers and the functions are listed in Table 5, and the table of specifications of 2016-2017 1<sup>st</sup> TEPSE English Test is given in Figure 1. As it can be clearly seen in Figure 1, there are three units that the students were responsible for the 2016-2017 1<sup>st</sup> term TEPSE English test. The functions of each unit are presented, and distributions of questions based on these functions can be seen in this table. Based on Figure 1, it can be stated that the questions were not distributed equally on the functions and the units. This table reveals that there were some items casting a doubt on content validity, and they were presented with the symbol '\*'. The details of this table of specification might be provided as follows:

Based on the functions of the units:

1. There are eight questions on the functions of Unit 1- Friendship, but one of these questions "question 2" casts a doubt on content validity because this question is on the topic of Unit 2.
2. There are six questions on the functions of Unit 2- Teen life, and four of these questions "question 1,4,17 and 18" cast a doubt on content validity because these questions are on the topic of Unit 1 and 3.
3. There are six questions on the functions of Unit 3- Cooking, and two of these questions "question 13 and 14" cast a doubt on content validity because they are on the topic of Unit 2.

Based on the topics of the units:

1. There are nine questions on the topic of Unit 1- Friendship, but one of these questions "question 17" is on the function "expressing likes and dislikes" of Unit 2.
2. There are five questions on the topic of Unit 2- Teen life; however, there are questions "questions 2, 13 and 14" that might be the focus of the functions "giving explanation reason" of Unit 1- Friendship and the function "naming common objects" of Unit 3- Cooking.
3. There are six questions on the topic of Unit 3- Cooking; however, two of these questions "question 1 and 18" are on the functions "expressing likes and dislikes" and "stating personal opinions" of Unit 2.

According to the table of specification, the number of the functions of each unit is different from each other. While Unit 1 and Unit 2 have five functions, Unit-3 has four functions. When the table of specification is examined, it can be noticed that some of the functions are common in some of the units. Moreover, there are eleven different functions in total in the first three units. The functions of first three units can be seen in Table 6.

**Table 6.** Functions of the First Three Units

FUNCTION	UNIT	FUNCTION	UNIT
Making simple inquires	1, 2, 3	Expressing likes and dislikes	2
Accepting and refusing	1	Expressing preferences	2, 3
Giving explanation/reason	1	Stating personal opinions	2
Apologizing	1	Describing simple process	3
Telling the time days and dates	1	Naming common objects	3
Describing the frequency of	2		

When Table 6 is considered, it can be noticed that the number of functions was eleven, and the number of questions in TEPSE English test was 20. Based on the table, two questions could have been asked on each function; however, the table of specification shows that there were more questions on some of the functions and units while there was not any question on the other functions. It can be said that the questions were not distributed equally to the functions and the units. The table of specification of 2016-2017 2<sup>nd</sup> term TEPSE English test is presented in Figure 2, Figure 3, and Figure 4.

According to Figure 2, Figure 3, Figure 4 and the plan that MoNE published, there were five units both in the first term and in the second term; however, the students were responsible for the first three units in the 1<sup>st</sup> term TEPSE English test while they were responsible for the first eight units in the 2<sup>nd</sup> term TEPSE English test. According to the table, which was presented based on the functions of the related units and the test questions, there were some items casting a doubt. The details of this table of specification might be provided as follows:

Based on the functions of the units:

1. There are three questions on the functions of Unit 1- Friendship but one of them “question 1” is on the topic of Unit 8.
2. There are two questions on the functions of Unit 2-Teenlife but these questions “question 15 and 19” cast a doubt on content validity because they are on the topic of Unit 5 and Unit 6.
3. There are three questions on the functions of Unit 3- Cooking and all of these questions “question 11, 13 and 14” cast a doubt because they are actually on the topic of Unit 4, Unit 6 and Unit 8.
4. There are two questions on the functions of Unit 4- Communication.
5. There are two questions on the functions of Unit 5- Internet.
6. There are four questions on the functions of Unit 6-Adventure, but one of these questions casts a doubt on content validity because it is on the topic of Unit 5.
7. There are three questions on the functions of Unit 7- Tourism.
8. There is one question on the functions of Unit 8- Chores.

Based on the topics of the units

1. There are two questions “question 5 and 6” on the topic of Unit 1- Friendship.
2. There is not any question on the topic of Unit 2- Teen life; however, there are two questions “question 15 and 19” that might be the focus of the functions “expressing likes and dislikes” and “expressing preferences” of Unit 2- Teen life.
3. There is not any question on the topic of Unit 3- Cooking; however, the number of questions on the function “naming common objects” of Unit 3 is three.
4. There are three questions “question 2, 4 and 11” on the topic of Unit 4- Communication, but one of them “question 11” casts a doubt because the function of this question is on “naming common objects” which belongs to Unit 3.
5. There are four questions “question 3, 18, 19 and 20” on the topic of Unit 5- Internet; however, “the questions 18 and 19” cast doubt because the function of ‘the question 18’ is “talking about what people do regularly” in Unit 6 and the function of ‘the question 19’ is “expressing preferences” in Unit 2.
6. There are five questions on the topic of Unit 6- Adventure; however, the functions of two questions “question 13 and 15” are “naming common objects” in Unit 3 and “expressing likes and dislikes” in Unit 2.
7. There are three questions “question 7, 9 and 12” on the topic of Unit 7- Tourism.
8. There are three questions “question 1, 10 and 14” on the topic of Unit 8- Chores; however, the function of ‘the question 1’ is “apologizing” in Unit-1 while the function of ‘the question 14’ is “naming common objects” in Unit-3.

As it can be clearly seen, the number of the functions of each unit is different from each other. When the table of specification is examined, it can be noticed that some of the functions are common in some of the units, and there are twenty-three different functions in total in the first eight units, which can be examined in Table 7.

Table of Specification of 2016-2017 1<sup>st</sup>Term TEPSE English Test  
The number at the top of the cells indicates the assigned number to the function, and the number at the bottom indicates the test item. The number at the top of the cell in the column 'totals' indicates the assigned number to the function, while the number at the bottom indicates the number of the test items)  
'\*' means that this item casts a doubt on content validity

UNITS	FUNCTIONS	KNOWLEDGE	COMPREHESION	APPLICATION	ANALYSIS	SYN./EVAL.	AFFECTIVE	PSYCHOMOTOR	TOTALS
UNIT-1 FRIENDSHIP	Accepting and refusing		1 5,12,16						1 3
	Apologizing								
	Giving explanations/ reason		3 2*, 15						3 2
	Making simple inquiries		4 7,10						4 2
	Telling the time, days and dates		5 9						5 1
UNIT-2 TEENLIFE	Describing the frequency of actions								
	Expressing likes and dislikes		7 1*, 17*						7 2
	Expressing preferences		8 4*						8 1
	Making simple inquiries		9 6,8						9 2
	Stating person opinions		10 18*						10 1
UNIT-3 COOKING	Describing simple processes		11 19,20						11 2
	Expressing preferences								
	Making simple inquiries		13 3						13 1
	Naming common objects	14 11, 13*,14*							14 3

**Figure 1.** Table of specification of 2016-2017 1<sup>st</sup> term TEPSE English test

Table of Specification of 2016-2017 2<sup>nd</sup> Term TEPSE English Test  
 The number at the top of the cells indicates the assigned number to the function, and the number at the bottom indicates the test item. The number at the top of the cell in the column 'totals' indicates the assigned number to the function, while the number at the bottom indicates the number of the test items)  
 '\* ' means that this item casts a doubt on content validity

UNITS	FUNCTIONS	KNOWLEDGE	COMPREHESION	APPLICATION	ANALYSIS	SYN./EVAL.	AFFECTIVE	PSYCHOMOTOR	TOTALS
UNIT-1 FRIENDSHIP	Accepting and refusing		1 5						1 1
	Apologizing		2 1*						2 1
	Giving explanations/ reason								
	Making simple inquiries		4 6						4 1
	Telling the time, days and dates								
UNIT-2 TEENLIFE	Describing the frequency of actions								
	Expressing likes and dislikes		7 15*						7 1
	Expressing preferences		8 19*						8 1
	Making simple inquiries								
	Stating person opinions								
UNIT-3 COOKING	Describing simple processes								
	Expressing preferences								
	Making simple inquiries								
	Naming common objects	14 11*,13*,14*							14 3

Figure 2. Table of specification of 2016-2017 2<sup>nd</sup> term TEPSE English test-1



Table of Specification of 2016-2017 2<sup>nd</sup> Term TEPSE English Test  
 The number at the top of the cells indicates the assigned number to the function, and the number at the bottom indicates the test item. The number at the top of the cell in the column 'totals' indicates the assigned number to the function, while the number at the bottom indicates the number of the test items)  
 '\*' means that this item casts a doubt on content validity

UNITS	FUNCTIONS	KNOWLEDGE	COMPREHENSION	APPLICATION	ANALYSIS	SYN./EVAL.	AFFECTIVE	PSYCHOMOTOR	TOTALS
UNIT-4 COMMUNICATION	Expressing concern and sympathy								
	Handling phone conversations								
	Making simple inquiries		2, 4						2 17
	Talking about plans								
UNIT-5 INTERNET	Accepting and refusing								
	Giving explanations/ reason		20						1 20
	Making excuses								
	Making simple requests								
	Making simple inquiries		3						1 23
	Talking about plans								
UNIT-6 ADVENTURE	Telling the time, days and dates								
	Expressing preferences		17						1 26
	Giving explanations/ reasons		8						1 27
	Making simple comparisons								
	Making simple inquiries								
	Stating personal opinions		16						1 30
	Talking about what people do regularly		18*						1 31
Talking about past events									

Figure 3. Table of specification of 2016-2017 2<sup>nd</sup> term TEPSE English test-2

Table of Specification of 2016-2017 2<sup>nd</sup>Term TEPSE English Test  
 The number at the top of the cells indicates the assigned number to the function, and the number at the bottom indicates the test item. The number at the top of the cell in the column 'totals' indicates the assigned number to the function, while the number at the bottom indicates the number of the test items.  
 '\*' means that this item casts a doubt on content validity.

UNITS	FUNCTIONS	KNOWLEDGE	COMPREHESION	APPLICATION	ANALYSIS	SYN./EVAL.	AFFECTIVE	PSYCHOMOTOR	TOTALS
UNIT-7 TOURISM	Describing places								
	Describing the weather		7						34 1
	Expressing preferences								
	Giving explanations/ reason								
	Making simple comparisons		12						37 1
	Stating personal opinions								
	Talking about past events		9						39 1
UNIT-8 CHORES	Expressing feelings								
	Expressing likes and dislikes								
	Expressing obligation		10						42 1
	Giving explanations/ reasons								
	Making simple inquiries								
	Making simple suggestions								

Figure 4. Table of specification of 2016-2017 2<sup>nd</sup> term TEPSE English test-3

**Table 7.** Functions of the First Eight Units

FUNCTION	UNITS	FUNCTION	UNITS
Making simple inquires	1,2,3,4,5,6,7	Handling phone conversation	4
Accepting and refusing	1,5	Talking about plans	4, 5
Giving explanation/reason	1, 5, 6, 7, 8	Making excuse	5
Apologizing	1	Making simple request	5
Telling the time days and dates	1, 5	Making simple comparisons	6, 7
Describing the frequency of actions	2	Talking about what people do regularly	6
Expressing likes and dislikes	2,8	Talking about past events	6, 7
Expressing preferences	2, 3, 6, 7	Describing places	7
Stating personal opinions	2, 6, 7	Describing the weather	7
Describing simple process	3	Expressing obligation	8
Naming common objects	3	Making simple suggestions	8
Expressing concern and sympathy	4		

Table 7 shows that the number of functions was twenty-three, and the number of questions in TEPSE was twenty. Each question of the test could have focused on only one function rather than asking more questions on some of the functions. The table of specification shows that there were more questions on some of the functions and units while there was not any question on the other functions and units.

### **The Teachers' Views toward Content Validity of TEPSE English Tests Conducted Between 2016 and 2017**

In the semi-structured interviews, three questions were addressed to the teachers to reveal their views on TEPSE English tests which were conducted between 2016 and 2017. The interview questions were parallel to the research questions that the researcher tried to find answers by analyzing the documents. The questions are:

1. Do the English language tests in TEPSE exactly focus on the frequently used items in the coursebook? If yes or no, which items are tested or not tested?
2. Is there an exact match between functions of the provided syllabus and the questions in the English language test in TEPSE?
3. Do you have any comments?

The semi-structured interview data collected from 21 English language teachers teaching English to 8<sup>th</sup> grade students, and they were coded and categorized after the transcribing process.

#### **Alignment between the English Tests in TEPSE and the Coursebook "Upturn in English" Based on the Frequently Used Items**

The responses of the teachers were analyzed and provided in the following sub-sections.

##### **Focused Both on the Vocabulary and Language Use Patterns.**

More than half of participants' view (n=14) about TEPSE English tests based on the frequently used items might be presented as:

*"I cannot say that the vocabulary items or language use patterns which were not presented in the coursebook were included in the exams. After answering the questions which were published by MoNE, I noticed that the items in the coursebook were used in the options or in the question itself"* (Participant 19, Age:31).

*"I think that, generally, the tests focused on the vocabulary items and language use patterns in the coursebook"* (Participant 15, Age:31).

##### **The Tested Items Even Though They Were Not the Focus of the Coursebook.**

Based on the responses, more than half of the participants (n=13) stated some tested items in TEPSE English tests even though they were not frequently used in the coursebook. According to their responses, these items were 'drum, frying pan, before, after, and, snowshoeing'. The following quotation might be useful:

*"Of course, there were some words, which we did not extremely focus in the exams. One of them was the word 'drum'. Although this word was not the focused one in the coursebook, it was included in the exam...And also, 'before' and 'after' were used in the exams, while the sequence words like 'first and second' were frequently used in the coursebook"* (Participant 13, Age:31).

### **The Tested Items Included in the Coursebook.**

The items tested in the exams were mentioned by some of the participants (n=9). Based on the responses, the following quotation summarizes the language use patterns and vocabulary items that TEPSE English test tested:

*"I have noted the conspicuous ones such as:*

*In the 2016-2017 1<sup>st</sup> TEPSE: Expressing opinion, responding to offers, present simple tense, expressing preference, be going to, cooking, expressing preference, and conjunctions.*

*In the 2016-2017 2<sup>nd</sup> TEPSE: Phone conversations, present simple tense, comparatives, expressing opinion, be going to, conjunctions, simple past tense, imperatives, extreme sports, and chore"* (Participant 3, Age:28).

### **Alignment between Functions of the Provided Syllabus and the Questions in the English Tests in TEPSE.**

The theme of the interview question-2 and most indicative quotations were presented in the following subsections.

#### **Alignment Problem between the Functions and the Exams/no Alignment.**

Based on the responses, two participants indicated the alignment problem between the functions and the exams. One of the quotations is provided as follows:

*"Listening, writing, and speaking cannot be exactly included because TEPSE is a multiple-choice exam. Therefore, the only skill out of four skills is reading. However, to me, TEPSE are also unable to assess reading skill since the questions in the tests do not align with each other when we consider on the functions"* (Participant 2, Age:25).

#### **The Exams Align with the Functions in General.**

Most of the participants (n=18) indicated that the questions in TEPSE generally aligned with the functions. The following statement is provided to emphasize the views of participants.

*"In fact, they align with each other...There are some functions which were frequently tested such as 'giving explanation and reason, expressing concern and sympathy, frequency'...Of course, there are some functions could not be tested because all of them could not be tested at the same time. However, the questions were based on nearly all of the functions"* (Participant 11, Age:25).

#### **Distribution of Questions Based on the Functions/units.**

More than half of the participants (n=12) made comments on the distribution of questions based on the units. Ten participants stated that the questions based on the units were not equally distributed in some of the English tests in TEPSE while two of them stated that the questions were distributed equally. The following two quotations present these two views:

*"When TEPSE between 2016-2017 were taken into consideration, the distribution of the questions based on the topics was not equal...For instance, the topics of Unit-2 and Unit-3 were not included in the 2016-2017 2<sup>nd</sup> term TEPSE English test and it could not test what it intended to test. The twenty questions could have been distributed equally to units"* (Participant 6, Age:32).

*"The questions were good and equally distributed"* (Participant 10, Age:31).

#### **Inconsistency between the Functions and the Units.**

Nearly half of the participants (n=9) indicated the inconsistency between the functions and the units. One of the participants stated as:

*"As I stated before, -the question on 'drum'- the function of this question does not belong to this unit. I mean it is a vocabulary question on the function 'naming object' but this function was not among the functions of this unit... Maybe some students had difficulties but they could tolerate and answer these questions because*

they were exposed to these functions in the previous units. However, I think that it would be better if the questions were asked based on the functions of the related units” (Participant 15, Age: 31).

### **The Functions/Items which were not Tested.**

Based on the responses (n=2), the following quotation might reveal the view towards the items which were not tested.

“For instance, we use a great variety of expressions to express idea, to make an offer etc.; however, just a few of them were included. The expressions such as ‘why don’t we’, ‘what about’, and ‘how about’ could have been used” (Participant 4, Age:25)

### **Functions Based on 4 Skills vs Exams.**

Five participants stated that four skills both the coursebook and the functions focused on could not be tested in the exams. One of the quotations was presented as follows:

“The functions were generally based on the communicative ones while the exams were multiple-choice tests. The coursebook focused on reading and listening... Of course, these activities were useful to improve students’ English; however, the students assume as if these activities were waste of time because of not being tested in the exams...” (Participant 12, Age: 24).

### **Suggestions and Comments of Participants.**

Interview question-3 examines the other comments and suggestions of the participants, and provided in the following sub-sections.

#### **Four Skills vs. Exams.**

Based on the responses, nearly half of the participants (n=9) emphasized that these exams could not test four language skills. The following quotation might summarize the views of the participants:

“The major problem in teaching foreign languages is that there are not any proper criteria to teach listening, writing, speaking, and even reading in our country. We have exams just to test grammar and vocabulary knowledge” (Participant 9, Age:27).

#### **The Coursebook vs TEPSE.**

Four participants mentioned the relation between the coursebook and TEPSE as a response to question-3. Based on these responses, two of them criticized the coursebook while the rests focused on the good sides of it.

“I think that the coursebook should be developed, and I demand that the coursebook should be more consistent with the exams. I used the coursebook provided by MoNE... and we do not have an opportunity to buy supplementary resources. That’s why the coursebook should be more consistent with the exams” (Participant 12, Age:24).

“Well, the coursebook was generally focusing on the items and using them frequently. From this perspective, I think it was much more effective and efficient” (Participant 19, Age:31).

#### **Pros of TEPSE.**

The pros of TEPSE such as content validity, number of questions and the matching between the coursebook and the exams were also emphasized by the participants (n=8):

“There were some good sides of the exams. I think that positive sides were conducting two exams in a year, including the first three units in the first term TEPSE, and including the following five units in the second term TEPSE. Also, I think that conducting make-up exams were also a positive side... (Participant 19, Age:31).

#### **TEPSE vs LGS.**

Based on the responses more than half of the participants (n=20) mentioned LGS exams focusing on the number of questions, content validity and the point value of English test questions, and criticized it. The responses indicated that more than half of the participants (n=15) compared TEPSE and LGS, and they stated that they were in favor of TEPSE exams. The following quotations might prove this inference:

*“I think, even though we criticized TEPSE in some aspects, it was better than the new system (LGS), and it could test more functions than the LGS can”* (Participant 11, Age:25).

*“As I stated before, in the LGS English tests, the number of questions and point value of English test questions were decreased. The decrease in the point value of the questions changed the views of the students toward English test in the exam. Also, when we focus on the exam, it seems that the decrease in the number of questions lowers content validity”* (Participant 16, Age:26).

### **Suggestions.**

Five participants provided some suggestions toward the language exams in Turkey by focusing on the four language skills and the way of presenting questions. These suggestions might be presented as:

*“But I wish that listening, speaking, and writing were also included, and we could assess them objectively. These types of things might be included if it is possible to conduct a new system”* (Participant 15, Age:31).

## **Discussion and Conclusion**

### **Content Validity of TEPSE English Test**

The current study aimed to investigate content validity of TEPSE English tests conducted between 2016 and 2017. As Wolf, Farnsworth and Herman (2008) emphasize that the purpose of a test is the first step of validation; thus, they also supported the idea of matching the content of the assessment with the intended construct. Therefore, based on the interviews and documents, the researcher tried to find answers to the research questions. The results show that the findings that the researcher obtained from the documents were triangulated with the responses of the participants. The items that the researcher obtained from the documents were also voiced by the participants.

### **Alignment between the Coursebook and TEPSE English Tests Based on the Frequently Used Items**

Language items used in the coursebook and the tests are crucial as Hughes (2003) states “A test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc. with which it is meant to be concerned” (p.26). For this purpose, the alignment gains importance considering the frequently used items between the coursebook and TEPSE English tests. The analyses of both the coursebook and the English tests in TEPSE show that most of the frequently used items in the tests were also the ones used frequently in the coursebook, which was also emphasized with a similar finding obtained in the study of Jaturapitakkul (2013) which revealed content validity of the traditional English language tests in Thailand by focusing on the alignment between the tests and the content that the students learnt in the classroom. Moreover, the consistency between the coursebook and TEPSE English tests affects content validity positively as it was also emphasized in the studies of Aksan (2001) and Kang and Chang (2014).

English test which was conducted in the 2016-2017 1<sup>st</sup> term included the frequently used items in the coursebook. The frequency list which was presented in Table 3 showed that the frequently used items in the test were included in the coursebook, and the frequency of these items was nearly parallel to the frequency of these items in the test. However, there were some items in the test which were not included in the coursebook. The number of the items in the test shows that there were 204 language items in the test while, 26 of the items were used once or none in the coursebook. This means that 87.26% of the language items in the test were used more than once in the coursebook. Moreover, Table 3 based on the frequency list of Top-30 items verifies that 2016-2017 1<sup>st</sup> term TEPSE English test mostly focused on the frequently used items in the coursebook. When the English test in TEPSE in 2016-2017 2<sup>nd</sup> term was taken into consideration, the frequency list in Table 4 shows that the most frequently used items in the test were also used frequently in the coursebook. However, the items which were not included in the coursebook were used in the test. When the number of these items was taken into



consideration, the number of the items used in the test was 220; however, 18 of them were used once or none in the coursebook. This means that 91.82% of the items in the test were also used in the coursebook more than once. Moreover, Table 4 based on the frequency list of Top-30 items verifies that 2016-2017 2<sup>nd</sup> term TEPSE English test mostly focused on the frequently used items in the coursebook

As a result, contrary to the study of Siddiek (2010), which emphasized the reason of lacking content validity of Sudan School Certificate English examinations as not having questions based on the textbook, and the study of Abella, Urrutia and Shneyderman (2003), which criticized language achievement tests because of not being valid measures of content area knowledge, the contents of TEPSE English tests were generally based on the coursebook. Most of the items in TEPSE English tests were also included in the coursebook and used frequently, which might indicate that there is an alignment between the coursebook and TEPSE English tests based on the frequency of the items. As it was also emphasized in the study of Külekçi (2016), providing representatives of the items intended to be assessed in the test proves content validity. Therefore, alignment between the coursebook and the test based on the frequently used items might increase content validity of TEPSE English tests. The study of Kang and Chang (2014) strengthens the idea that TEPSE English tests have content validity because they stated that because of being based on the textbook and curriculum, PECT had appropriate content to test learners' English skills. Moreover, the study of Aksan (2001) emphasized content validity of an exam based on the alignment between the exam and the content of the coursebook; therefore, the alignment between TEPSE English tests and the coursebook might prove content validity of TEPSE English tests.

#### **Alignment between the Functions and TEPSE English Tests**

The alignment between the functions and TEPSE English tests gains importance since as Ekbatani (2011) asserts, content validity is consistence between objectives/functions of the test and the test itself. For this purpose, table of specifications based on the tests were created by the researcher as it was also used in the studies of Sims (2015) and Newman et al. (2013) to analyze content validity of the tests. These researchers considered the table of specifications as a way of determining the alignment between the functions and the test items to analyze content validity.

In table of specification of TEPSE English test conducted in the 1<sup>st</sup> term of 2016 and 2017, Table 6 reveals that two questions could have been asked on each function to distribute the questions equally both on the functions and the units. However, based on the functions, there were eight questions on Unit-1, six questions on Unit-2 and Unit-3. When the topics were taken into consideration, there were nine questions on Unit-1, five questions on Unit-2, and six questions of Unit-3. Even though the distribution of the questions seems nearly equal, there were six questions casting doubt on content validity because of the inconsistencies between the topics and functions of these questions. However, most of the questions were on the intended functions provided by MoNE.

In addition to the 2016-2017 1<sup>st</sup> term TEPSE English test, 2016-2017 2<sup>nd</sup> term TEPSE English test was also examined. Figures 2, 3, and 4 show that based on the functions; there were three questions on Unit-1, Unit-3, Unit-6, and Unit-7, while there were two questions on Unit-2, Unit-4, and Unit-5, and only one question on Unit-8. However, based on the topics, there were two questions on Unit-1 and three questions on Unit-4, Unit-7 and Unit-8. Moreover, it was determined that there were four questions on Unit-5, and five questions on Unit-6, while there was not any question on Unit-2 and Unit-3. Each question of the test could have been focused on only one function rather than asking more questions on some of the functions. Moreover, the distribution of the questions was not equal as it can be seen in Figures 2, 3, and 4. These inconsistencies between the functions and the units cast a doubt on content validity, and there were seven questions emphasizing these inconsistencies even though most of the questions were on the intended functions.

As a conclusion, the results show that the distribution of the questions based on the functions in the 1<sup>st</sup> term TEPSE English test seems more equal than the 2<sup>nd</sup> term TEPSE English test. Besides, there were some questions casting doubt because of the inconsistency between the functions and units of these questions. However, more than half of the questions in each test were on the functions that they intended to test, which was also emphasized in the similar findings obtained in the studies of Vural (2017) and Fathony (2017), while the questions casting doubt might affect content validity of the tests negatively. This was also emphasized by Chakwera (2004) who stated that there was an alignment between content validity and curricular validity which

covers the functions. However, when the functions and the questions were examined thoroughly, it can be stated that the functions could be tested based on the units that the students were responsible for. The effect of the equal distribution of the questions on content validity was also emphasized in the study of Razmjoo and Tabrizi (2010) on TEFL M.A Entrance Examination. Similar to the results of TEPSE English tests in the current study, Razmjoo and Tabrizi's study indicated that there was not an equal distribution of items among the content categories and that TEFL M.A. Entrance Examination was not a valid one in terms of content validity. This might mean that the functions could be tested but the inequality in distribution of the functions might affect content validity negatively.

### **The Teachers' Views toward Content Validity of TEPSE English Tests Conducted between 2016 and 2017**

One of the research questions of the current study aimed to determine the teachers' views toward content validity of TEPSE English tests conducted between 2016 and 2017. For this purpose, semi-structured interviews were held with 21 English language teachers teaching English to 8<sup>th</sup> grade students.

#### **Alignment between the Coursebook and TEPSE English Tests Based on the Frequently Used Items**

As it was voiced by Hughes (2003), a test which has content validity is the test representing the items that will match the content. For this purpose, teachers' views were also taken into consideration in addition to the findings obtained from the analysis of documents.

Based on the responses, 23.80% of the participants (n=5) expressed that TEPSE English tests mostly focused on language use patterns in the coursebook rather than vocabulary items. Moreover, 66.6% of the participants (n=14) agreed that TEPSE English tests were generally in alignment with the coursebook based on the frequently used items. The following statement might clarify this assumption:

*"I think that, generally, the tests focused on the vocabulary items and language use patterns in the coursebook"* (Participant 15, Age:31).

This assumption might be strengthened with the findings obtained from the documents. When the percentages of the included items in the test were taken into consideration, 87.26% of the items in the 2016-2017 1<sup>st</sup> term TEPSE test and, 91.82% of the items in the 2016-2017 2<sup>nd</sup> term TEPSE test were used more than once in the coursebook. Even though there were many items which were frequently used both in the coursebook and the tests, 61.9% of the participants (n=13) indicated that the tests included some items as key words which were not frequently used in the coursebook. The mostly emphasized item which was not focused on the coursebook was presented as 'drum (n=6)'. The following response expresses this as follows:

*"Of course, there were some words, which we did not extremely focus in the exams. One them was the word 'drum'. Although this word was not the focused one in the coursebook, it was included in the exam"* (Participant 13, Age:31).

When the responses of participants were taken into consideration, the most conspicuous item which was tested even though it was not frequently used in the coursebook was 'drum'. Regarding the documents, 'drum' was used once in the coursebook in the coursebook. This might mean that including such items in the test might affect content validity negatively; however, when the number of the items was considered, most of the items in tests were frequently used in the coursebook. Contrary to the studies of Siddiek (2010) and Abella, Urrutia and Shneyderman, (2003) which attributed the lack of content validity in exams to their not having questions based on the textbook, Jaturapitakkulin (2013) emphasized content validity of traditional English language tests in Thailand by focusing the alignment between the test and the content that the students learnt in the classroom. In addition, Aksan (2001) examined content validity of English language exams at Niğde University, and the participants were asked the question 'to what extent is the content of coursebook represented in the exams'. The responses revealed that most of the teachers were positive on this issue, and this indicated that these exams had content validity based on teachers' views. Therefore, it can be inferred that the alignment between the coursebook and TEPSE English tests based on the frequency of items might prove content validity of TEPSE English tests.

### Alignment between the Functions and TEPSE English Tests

As it was voiced by Ekbatani (2011), alignment between the functions and the tests was crucial in terms of content validity. In line with this, the teachers' views were obtained on the second research question during the interviews. Based on the responses, 85.7% of the participants (n=18) agreed that there was an alignment between the functions and TEPSE English tests despite some inconsistencies and unequal distributions of questions across units. One of quotations on the alignment can be presented as follows:

*"In fact, they align with each other...There are some functions which were frequently tested...Of course, there are some functions could not be tested because all of them could not be tested at the same time. However, the questions were based on nearly all of the functions"* (Participant 11, Age:25).

47.61% of the participants (n=10) indicated that there was not an equal distribution of the questions based on the units/functions while 9.52% of the participants (n= 2) thought that the questions were distributed equally based on the units, especially the first three units. Moreover, 42.85% of the participants (n=9) emphasized the inconsistencies between the functions and the units. As in the study of Razmjoo and Tabrizi (2010) on TEFL M.A Entrance Examination which indicated that TEFL M.A. Entrance Examination was not a valid one in terms of content validity because there was not an equal distribution of items among the content categories, content validity of English tests in TEPSE might be also affected negatively. However, the participants of the current study also stated that these inconsistencies might not be a problem because of the exposure of the students to these functions in the previous units. The following statement might provide an insight into this issue:

*"As I stated before, in the question on 'drum', 'the function of this question does not belong to this unit -I mean it is a vocabulary question on the function 'naming object' but this function was not among the functions of this unit...As I stated before, even though the unit does not include this function, maybe some students had difficulties but they could tolerate and answer these questions because they were exposed to these functions in the previous units"* (Participant 15, Age:31).

Additionally, 57.14% of the participants (n=12) expressed that these exams could not assess four language skills as it was emphasized in the study of Akin (2016) which indicated that YDS tests grammar, vocabulary, and reading comprehension rather than four language skills. The following quotation on this issue is presented:

*"The major problem in teaching foreign languages is that there are not any proper criteria to teach listening, writing, speaking, even reading in our country. We have exams just to test grammar and vocabulary knowledge"* (Participant 9, Age:27).

To conclude, the findings regarding to the views of teachers toward content validity of TEPSE English test revealed that great majority of the participant agreed that TEPSE English tests had content validity despite some inconsistencies, unequal distributions, which were also emphasized in the study of Razmjoo and Tabrizi (2010) and the lack of assessing four language skills. These findings are also in alignment with those of the study conducted by Vural (2017), in which most of the teachers agreed that TEPSE test questions tested the functions in the coursebook, while acknowledging its failure in assessing listening and speaking skills, which could affect content validity negatively. Al- Adawi and Al-Balushi (2016) also obtained similar underscoring need to test listening and speaking in the exams. Moreover, the studies conducted by Weiping and Juan (2005), Haiyan and Fuqin (2005), and Nicholson (2015) emphasized the weak content validity of the exams due to the failure in reflecting the students' communicative competence. Therefore, from this perspective, content validity of TEPSE English tests can be stated to be affected negatively.

### Other Comments

Based on the responses, the comments of participants focused on four language skills, TEPSE, LGS, and the coursebook. Several participants also provided some suggestions to the exams conducted in Turkey. Similar to the study of Mart (2014), which also emphasized both the negative and positive perspectives of the teachers toward TEPSE, the views of the participants toward TEPSE tests were mostly positive.

As it was also emphasized in the study of Gömleksiz and Aslan (2017), 57.14% of the participants agreed that these exams could not assess four language skills. Similarly, the current study revealed 95.23% of the participants criticized LGS in some aspects. Moreover, the participants compared TEPSE and LGS, and 71.42%

of them stated that TEPSE was better than LGS considering the number of questions, content validity, and point value of test questions. The following quotation might summarize these perspectives:

*“As I stated before, in the LGS English tests, the number of questions and point value of English test questions were decreased. The decrease in the point value of questions changed the views of the students toward English test in the exam. Also, when we focus on the exam, it seems that the decrease in the number of questions lowers content validity”* (Participant 16, Age:26).

*“I think, even though we criticized TEPSE in some aspects, it was better than the new system (LGS), and it could test more functions than LGS can”* (Participant 11, Age:25).

The participants (23.8%) suggested improvements in the exams conducted in Turkey by indicating the need to focus on the four language skills, which was in alignment with the findings of the studies conducted by Vural (2017) and Al- Adawi and Al-Balushi (2016). Moreover, presenting the questions in different formats was also stated.

*“But, I wish that listening, speaking, and writing were also included, and we could assess them objectively. These types of things might be included if it is possible to conduct a new system”* (Participant 15, Age:31).

### **Overall Summary of the Study**

Language assessment has a crucial role in education, and the importance of assessment has been voiced by many researchers. Brown and Abeywickrama (2010) reflected the principles of language assessment as practicality, reliability, validity, authenticity, and washback. When the educational setting of Turkey is considered, exams which are generally multiple-choice are at the very heart of the educational system. Based on this fact, TEPSE English tests among the language tests in Turkey is the focus of the current study. This exam has been suddenly replaced by another exam ‘LGS’; however, the researcher had already started carrying out this study. There was not any study investigating content validity of TEPSE English tests conducted between 2016 and 2017. Also, to the best knowledge of the author, there was only one study conducted on content validity of English test in TEPSE (Vural, 2017). In her study, Vural (2017) focused on TEPSE English test in 2014 and the data were only based on the teachers’ views. The current study; therefore, aimed to find out to what extent English tests in TEPSE conducted between 2016 and 2017 have content validity based on the analysis of documents and teachers’ views. In order to obtain information about content validity of English tests in TEPSE, mixed research method was used. As Creswell (2009) stated that observations, interviews, and documents are the ways of data collection methods, and the current study benefited from the documents such as syllabus provided by MoNE, the coursebook ‘Upturn in English’ and TEPSE English tests conducted between 2016 and 2017. Besides, semi-structured interviews were held with 21 English language teachers teaching to 8<sup>th</sup> grade students in eighteen different provinces. As quantitative analysis, the coursebook ‘Upturn in English’ and TEPSE English tests were analyzed and compared based on the frequency of the items. In Turkey, curricula are realized with coursebooks, and neither the coursebook nor TEPSE English tests in those years included some of the vocabulary items suggested by curriculum. Therefore, analyzing the frequency of vocabulary items in the coursebook and TEPSE English tests is one of the aims of the current study. Moreover, the table of specification provided by Newman et al. (1973 as cited in Newman et al., 2013) was adapted and used to analyze TEPSE English tests based on content validity. The following main research questions were investigated.

1. To what extent do the English tests in TEPSE conducted between 2016 and 2017 have content validity?
2. What are English language teachers’ views on content validity of English test in TEPSE?

First of all, the alignment between the coursebook and TEPSE English tests based on the frequently used items was crucial as Hughes (2003) stated “A test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc. with which it is meant to be concerned” (p.26). The findings obtained from the documents and interviews revealed that most of the frequently used items in the tests were also used frequently in the coursebook. The details of the alignment between the coursebook and TEPSE English tests can be presented as:

1. 87.26% of the language items in the 2016-2017 1<sup>st</sup> term TEPSE English test were used more than once in the coursebook.

2. 91.82% of the items in the 2016-2017 2<sup>nd</sup> term TEPSE English test were also used in the coursebook more than once.

Also, the frequency lists of top-30 items based on TEPSE English tests and the coursebook (See Table 3 and 4) show similar results, which means that there is an alignment between the coursebook and TEPSE English tests regarding the frequently used items. Moreover, Kang and Chang (2014) state that a test has content validity if it is based on the textbook and curriculum; therefore, it can be stated that TEPSE English tests have content validity based on the representativeness of frequently used items. Moreover, Ekbatani (2011) claims that content validity is a consistence between objectives/functions of the test and the test itself. In regard, the alignment between the functions and TEPSE English tests was another focus of this study. The findings obtained from the table of specifications revealed that:

1. The distribution of the questions based on the topics, especially in the 2<sup>nd</sup> term TEPSE English test, was not exactly equal.
2. The distribution of the questions based on the functions in the 1<sup>st</sup> term TEPSE English test seems more equal than the 2<sup>nd</sup> term TEPSE English test.
3. There were some questions casting doubt on content validity because of the inconsistency between the functions and units of these questions.
4. More than half of the questions in each test were on the functions that they intended to test
5. The functions could be tested based on the units that the students were responsible for.

Razmjoo and Tabrizi (2010) emphasized the effect of the equal distribution of the items on content validity, and the impact of consistency between the functions and tests on content validity was voiced by Ekbatani (2011). Therefore, it can be inferred that the inconsistencies and unequal distributions in the tests affected content validity of TEPSE English tests negatively. However, it cannot be denied that more than half of the questions in each test could test what they intend to test, and the questions were on the predetermined functions provided by MoNE.

In response to the first research question, it can be claimed that TEPSE English tests between 2016 and 2017 seem to have content validity based on the alignment between the coursebook and the tests, while their content validities were affected negatively because of some inconsistencies and unequal distributions of the questions. In addition to the documents, teachers' views were the other focus of the current study. The participants' responses were investigated, and results might be presented as follows:

1. 66.6% of the participants agreed that there was an alignment between the coursebook and TEPSE English tests based on the frequently used items, while 23.80% of the participants expressed that TEPSE English tests focused on language use patterns in the coursebook rather than vocabulary items.
2. 61.9% of the participants indicated that the tests included a few items as key words which were not frequently used in the coursebook.

Based on the responses, it might be stated that there was an alignment between the coursebook and TEPSE English tests regarding the frequency of the items. However, including some items which were not frequently used might affect content validity negatively. Fortunately, most of the items in the tests were also used frequently in the coursebook. Aksan (2001) revealed that representativeness of the content in the tests proves content validity; therefore, it can be implied that the most of the participants agreed on content validity of TEPSE English tests based on the representativeness of frequently used items.

The current study has also focused on the views of teachers toward the alignment between the functions and TEPSE English tests. The responses of the participants can be presented as:

1. 85.7% of the participants agreed that there was an alignment between the functions and TEPSE English tests despite some inconsistencies and unequal distributions.
2. 47.61% of the participants indicated that there was not equal distribution of the questions based on the units/functions



3. 14.28% of the participants agreed that the questions were distributed equally based on the first three units.
4. 42.85% of the participants emphasized the inconsistencies between the functions and the units. However, they also stated that these inconsistencies might not be a problem.
5. 57.14% of the participants expressed that these exams could not assess four language skills
6. 95.23% of the participants criticized LGS in some aspects such as the number of questions, point value of English test questions, and content validity.
7. 71.42% of them stated that TEPSE was better than LGS considering the number of questions, content validity, and point value of test questions.

The responses indicate that the participants agreed on content validity of TEPSE English tests regarding the alignment between the functions and tests despite some inconsistencies and neglecting assessing four language skills. Moreover, the participants compared LGS and TEPSE, and criticized the LGS for the decrease in the number of the questions and in the point value of English test questions. They also stated that the number of the questions in LGS English test which is 10, might affect content validity negatively when compared to the number of questions in TEPSE English test which was 20. Therefore, the participants were in favor of conducting TEPSE English test rather than LGS English test. Moreover, the participants of the current study also demand a test which can assess four language skills because they believe that language learning mean more than vocabulary and grammar, and students should be tested based on these skills, which might also improve content validity if it is implemented successfully. As a conclusion, it can be put forward that TEPSE English tests conducted between 2016 and 2017 seem to have high content validity based on the alignment among the coursebook, functions, and the tests; however, content validity of these tests was also affected negatively because of some inconsistencies, unequal distributions, and lacking of assessing four language skills. The reason for this might be attributed to the frequent changes in the educational system in Turkey. TEPSE English tests are one of the exams which were also prone to the changes based on the system and could not test what intend to test in some aspects. To improve content validity of these exams, it might be suggested that these exams include the neglected features such as assessing four language skills as mentioned in the studies of many researchers like Aslan and Gömleksiz (2017), and Vural (2017). Moreover, equal distribution of the topics/functions in the exams plays a crucial role in content validity as it was emphasized in the studies of Razmjoo and Tabrizi (2010); therefore, the questions should be distributed equally based on the topics/functions.

### **Acknowledgments**

This study is based on the M.A. thesis of the first author, which was submitted to the Institute of Educational Sciences at Mehmet Akif Ersoy University in 2019 under the supervision of Prof. Dr. Ferit Kılıçkaya.

### **Statement of Publication Ethics**

The authors of this article declare that the research has no unethical problem, and required permission to collect data was obtained from Ministry of National Education.

### **Researchers' Contribution Rate**

The first author collected data, analyzed and prepared necessary reports while the second author supervised, guided and finalized the research process.

### **Conflict of Interest**

The authors of this article declare that there is not conflict of interest.



## References

- Abella, R., Urrutia, J., & Shneyderman, A. (2003, February). *An examination of the validity of English language assessment test scores in a LEP student population*. Paper presented at the annual meeting of the National Association for Bilingual Education, New Orleans, LA. Retrieved from <http://oer.dadeschools.net/nabe03paper.pdf>
- Akın, G. (2016). Evaluation of national foreign language test in Turkey. *Asian Journal of Educational Research*, 4(3), 11-21. Retrieved from <http://www.multidisciplinaryjournals.com/wp-content/uploads/2016/04/FULL-PAPER-EVALUATION-OF-NATIONAL-FOREIGN-LANGUAGE-TEST-IN-TURKEY.pdf>
- Aksan, M. M. (2001). *Instructors' perceptions of the content validity of the English language exams at Niğde university* (Unpublished Master's Thesis). Bilkent University, Ankara, Turkey. Retrieved from <http://www.thesis.bilkent.edu.tr/0001838.pdf>.
- Al-Adawi, S. S. A., & Al-Balushi, A. A. K. (2016). Investigating content and face validity of English language placement test designed by Colleges of Applied Sciences. *English Language Teaching*, 9(1), 107-121. doi:10.5539/elt.v9n1p107
- Alderson, C. J., & Kremmel, B. (2013). Re-examining the content validation of a grammar test: The (im)possibility of distinguishing vocabulary and structural knowledge. *Language Testing*, 30(4), 535-556. doi:10.1177/0265532213489568.
- Antalya Provincial Directorate of National Education (2016). *2016-2017 eğitim-öğretim yılı örnek yıllık plan [2016-2017 educational year sample annual plan]*. Retrieved from [http://ilbap.meb.k12.tr/icerikler/2016-2017-egitim-ogretim-yili-ornek-yillik-planlari\\_2705897.html](http://ilbap.meb.k12.tr/icerikler/2016-2017-egitim-ogretim-yili-ornek-yillik-planlari_2705897.html)
- Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. *Reading Psychology*, 24(3-4), 323-335. doi: 10.1080/02702710390227297
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Education.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson Education.
- Chakwera, E. W. J. (2004). *Content validity of independently constructed curriculum-based examinations* (Unpublished Doctoral Dissertation). Graduate School of the University of Massachusetts, Amherst, U.S.A.
- Cheng, L., & Fox, J. (2017). *Assessment in language classroom: Teachers supporting student learning*. Palgrave.
- Christensen, L. B., Johnson, B. R., & Turner, A. L. (2015). *Research methods, design, and analysis* (12th ed.). Pearson.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). SAGE Publications.
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson Education.
- Davidson, F., & Lynch, B. K. (2002). *Testcraft- A teacher's guide to writing and using language test specifications*. Yale University Press.
- Ekbatani, G. (2011). *Measurement and evaluation in post-secondary ESL*. Routledge.
- Fathony, F. (2017). *An analysis on content validity of English summative test items of second grade students at Mtsn Malijambe in the academic year 2015/2016* (Unpublished Master's Thesis). The State Islamic Institute, Surakarta, Indonesia.
- Fraenkel, R. J., Wallen, E. N., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.

- Griffiee, T. D. (2012). *An introduction to second language research methods: Design and data*. California, CA: TESL-EJ Publications.
- Gorsuch, G., & Griffiee, D. T. (2018). *Second language testing for student evaluation and classroom research*. Information Age Publishing Inc.
- Gömlüksiz, M. N., & Aslan, S. (2017). Secondary school students' views about the effectiveness of TEOG exam on English language learning: A qualitative study. *International Journal of Social Science*, 11(64), 1-14. doi: 10.9761/JASSS7375.
- Haiyan, W., & Fuqin, Z. (2005). A validation of CET for testing candidates' communicative competence and a proposal of a university-based communicative test. *CELEA Journal*, 28(2), 26-31. Retrieved from <http://www.celea.org.cn/teic/60/60-26.pdf>
- Hughes, A. (2003). *Testing for language teachers (2nd ed.)*. Cambridge University Press
- Ing, L. M., Musah, M. B., Al-Hudawi, V. H. S., Tahir, L. M., & Kamil, M. N. (2015). Validity of teacher made assessment: A table of specification approach. *Asian Social Science*, 11(5), 193-200. doi:10.5539/ass.v11n5p193.
- Jaturapitakkul, N. (2013). Students' perceptions of traditional English language testing in Thailand. *Academic Journal of Interdisciplinary Studies*. 2(3), 445-452. doi:10.5901/ajis.2013.v2n3p445.
- Kang, M., & Chang, H. (2014). Ensuring validity of practical English certification of local office of education in Korea. *Pan-Pacific Association of Applied Linguistics*, 18(2), 111-122. Retrieved from [http://paal.kr/download/18\\_2/18-2-7\\_Mun-koo\\_Kang\\_and\\_Hyung-ji\\_Chang.pdf](http://paal.kr/download/18_2/18-2-7_Mun-koo_Kang_and_Hyung-ji_Chang.pdf)
- Kılıçkaya, F. (2016). Washback effects of a high-stakes exam on lower secondary school English teachers' practices in the classroom. *Lublin Studies in Modern Languages and Literature*, 40(1), 116-134. doi:10.17951/lsml.2016.40.1.116.
- Külekçi, E. (2016). A concise analysis of the Foreign Language Examination (YDS) in Turkey and its possible washback effects. *International Online Journal of Education and Teaching (IOJET)*, 3(4), 303-315. Retrieved from: <http://iojet.org/index.php/IOJET/article/view/141/143>
- Lochmiller, C. R., & Lester, J. N. (2017). *An introduction to educational research: Connecting methods to practice*. Sage Publications.
- Mackey, A., & Gas, M. S. (Eds.). (2012). *Research methods in second language acquisition: A practical guide*. Blackwell Publishing.
- Mart, M. T. (2014). *Determination of tutors' views on the common exams applied in secondary schools within the student passing process from primary education to high school education (TEOG)* (Unpublished Master's Thesis). Zirve University, Gaziantep, Turkey. Available from Council of Higher Education Thesis Centre.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis. (2nd ed.)*. Sage.
- MoNE. (2011). *Sınavla öğrenci alacak ortaöğretim kurumlarına ilişkin merkezi sınav başvuru ve uygulama kılavuzu [Guidelines for applicants regarding to secondary education institutions which admits the students based on the central exams]*. Retrieved from [http://mevzuat.meb.gov.tr/html/608\\_0.html](http://mevzuat.meb.gov.tr/html/608_0.html)
- MoNE. (2016). *2016-2017 eğitim ve öğretim yılı merkezi sistem ortak sınavları [TEPSE in 2016-2017 educational year]*. Retrieved from: <http://ttkb.meb.gov.tr/www/2016-2017-egitim-ve-ogretim-yili-merkezi-sistem-ortak-sinavleri/icerik/281>.
- MoNE. (2016b). *8. Sınıf 1. dönem İngilizce dersi merkezi ortak sınavı [English exam questions for grade 8 in the first semester]*. Retrieved from [https://odsgm.meb.gov.tr/meb\\_iys\\_dosyalar/2016\\_11/24014749\\_ingilizce\\_a.pdf](https://odsgm.meb.gov.tr/meb_iys_dosyalar/2016_11/24014749_ingilizce_a.pdf).
- MoNE. (2017). *8. Sınıf 2. dönem İngilizce dersi merkezi ortak sınavı [English exam questions for grade 8 in the second semester]*. Retrieved from [http://odsgm.meb.gov.tr/meb\\_iys\\_dosyalar/2017\\_04/27131529\\_YNGYLYZCE\\_A.pdf](http://odsgm.meb.gov.tr/meb_iys_dosyalar/2017_04/27131529_YNGYLYZCE_A.pdf)

- MoNE (2018). *Sınavla öğrenci alacak ortaöğretim kurumlarına ilişkin merkezi sınav başvuru ve uygulama kılavuzu [Guidelines for applicants regarding to secondary education institutions which admits the students based on the central exams]*. Retrieved from: [https://www.meb.gov.tr/sinavlar/dokumanlar/2018/MERKEZI\\_SINAV\\_BASVURU\\_VE\\_UYGULAMA\\_KILAVUZU.pdf](https://www.meb.gov.tr/sinavlar/dokumanlar/2018/MERKEZI_SINAV_BASVURU_VE_UYGULAMA_KILAVUZU.pdf)
- Newman, I., Frye, B., Blumenfeld, G., & Newman, C. (1973). *An introduction to the basic concepts of measurement and evaluation*. University of Akron.
- Newman, I., Lim, J., & Pineda, F. (2013). Content validity using a mixed method approach: Its application and development through the use of a table of specifications methodology. *Journal of Mixed Method Research*, 7(3), 243-260. doi: 10.1177/1558689813476922.
- Nicholson, S. J. (2015). Evaluating the TOEIC® in South Korea: Practicality, reliability and validity. *International Journal of Education*, 7(1), 221-233. doi: 10.5296/ije.v7i1.7148.
- Ökmen, B., & Kılıç, A. (2016). The effect of language teaching methods on academic success in Turkey. *Journal of Education and Training Studies*, 4(10), 193-199. doi: 10.11114/jets.v4i10.1767.
- Pterneas, V. (2009). Word frequency counter [Software]. Available from <https://www.softpedia.com/get/Office-tools/Other-Office-Tools/Word-Frequency-Counter.shtml>
- Razmjoo, S., & Tabrizi, H. (2010). A content analysis of the TEFL M.A entrance examinations (case study: Major courses). *Journal of Pan-Pacific Association of Applied Linguistics*, 14(1), 159-170. Retrieved from <http://files.eric.ed.gov/fulltext/EJ920511.pdf>
- Siddiek, G. A. (2010). The impact of test content validity on language teaching and learning. *Asian Social Science*, 6(12), 133-143. Retrieved from <http://www.ccsenet.org/journal/index.php/ass/article/view/6817/6305>
- Sims, M. J. (2015). A valid and reliable English proficiency exam: A model from a university language program in Taiwan. *English as a Global Language Education*, 1(2), 91-125. doi: 10.6294/EaGLE.2015.0102.04.
- Solano-Flores, G. (2016). *Assessing English language learners: Theory and practice*. Routledge.
- TUIK. (2016). *2015 well-being index for provinces*. Retrieved from: <http://www.tuik.gov.tr/PreHaberBultenleri.do?id=24561>.
- Vural, A. (2017). *The assessment of consistency of 8th grade English course curriculum to the questions in TEOG English exam according to teachers' opinions* (Unpublished Master's Thesis). Gaziantep University, Gaziantep, Turkey. Retrieved from Council of Higher Education Thesis Centre.
- Weiping, G., & Juan, L. (2005). Test analysis of college students' communicative competence in English. *Asian EFL Journal*, 7(2). Retrieved from [http://www.asian-efl-journal.com/June\\_05\\_wg.pdf](http://www.asian-efl-journal.com/June_05_wg.pdf)
- Wolf, M. K., Farnsworth, T., & Herman, J. (2008). Validity issues in assessing English language learners' language proficiency. *Educational Assessment*, 13(2-3), 80-107. doi:10.1080/10627190802394222.