Kuram ve Uygulamada
SOSYAL BİLİMLER DERGİSİ
*Social Sciences: Theory & Practice*
**ISSN:** 2619-9408

***Geliş/Received***: 03.05.2024 ***Kabul/Accepted:*** 20.06.2024
***Makale Türü*:** Araştırma

# Development of Turkish Listening and Reading Test as a Foreign Language: The Case of Gazi University TÖMER

*Nezir TEMÜR*[*]
*Haluk GÜNGÖR*[**]

**ABSTRACT**
In this study, the validity and reliability of the listening and reading exams prepared by Gazi University TÖMER to be used in the Turkish as a foreign language C1 certification exam were examined. The sample of the research consists of reading and listening test data of 250 participants who came to TÖMER between January and October 2023 and took the face-to-face exam. The content validity of the tests was ensured by the table of specifications created by the researchers. The test items were prepared based on CEFR (2020) qualifications and achievements in the MoNE Teaching Turkish as a Foreign Language Program (2020). The prepared test forms were presented to expert opinion and necessary arrangements were made in line with the suggestions. After the test preparation phase was completed, the administration of the exams started. After the sample size reached a sufficient number, data analysis began. In addition to descriptive analyzes such as percentage and frequency, factor analysis, item difficulty indices, item discrimination indices, reliability coefficient, average difficulty and average discrimination indices were calculated on the data. Factor analyzes of the tests were performed in the R-based Shiny application. One-dimensional tests were obtained by deleting 5 items from the reading test and 10 items from the listening test with factor load values below .40. In terms of reliability and discrimination, the reading test generally has high reliability and discrimination; It was determined that the listening test had good reliability and discrimination and it was an easy test. When the item analyzes of the reading test were examined, it was understood that 15 items were good, and when the item analyzes of the listening test were examined, 9 items were good and 1 items could be corrected and included in the test. Based on the findings, it was concluded that the items of the reading exam gave better response than the items of the listening exam and that a significant part of the reading exam could be preserved and used.

**Keywords:** Turkish Teaching to Foreigners, Measurement and Evaluation, Measurement and Evaluation in Teaching Turkish to Foreigners, Test Development.

# Yabancı Dil Olarak Türkçe Okuma ve Dinleme Testi Geliştirme: Gazi Üniversitesi TÖMER Örneği

**ÖZ**
Bu araştırmada, yabancı dil olarak Türkçe C1 sertifika sınavında sınavında kullanılmak üzere Gazi Üniversitesi TÖMER tarafından hazırlanan dinleme ve okuma sınavlarının geçerliği ve güvenirliği incelenmiştir. Araştırmanın örneklemi, 2023 yılının Ocak-Ekim ayları arasında TÖMER'e gelerek yüz yüze sınava giren 250 katılımcının okuma ve dinleme sınav verilerinden oluşmaktadır. Testlerin kapsam geçerliği araştırmacılar tarafından oluşturulan belirtke tablosu ile sağlanmıştır. Test maddeleri, CEFR (2020) yeterlikleri ve Türkçenin MEB Yabancı Dil Olarak Öğretimi

***Citation:*** Temür, N. & Güngör, H. (2024). Development of Turkish listening and reading test as a foreign language: The case of Gazi University TÖMER. *Social Sciences Theory and Practice*, *8*(1), 246-263. Doi: 10.48066/kusob.1478059
[*] Prof. Dr., Gazi Üniversitesi, Gazi Eğitim Fakültesi, Türkçe ve Sosyal Bilimler Eğitimi Bölümü, ntemur@gazi.edu.tr, ORCID: 0000-0002-8052-1927.
[**] Doç. Dr., Gazi Üniversitesi, Gazi Eğitim Fakültesi, Türkçe ve Sosyal Bilimler Eğitimi Bölümü, halukgungor@gazi.edu.tr, ORCID: 0000-0002-4111-4106

Programı'ndaki (2020) kazanımlar esas alınarak hazırlanmıştır. Hazırlanan test formları uzman görüşüne sunulmuş ve gelen öneriler doğrultusunda gerekli düzenlemeler yapılmıştır. Test hazırlama aşaması tamamlandıktan sonra sınavların uygulamasına başlanmıştır. Örneklem büyüklüğü, yeterli sayıya ulaştıktan sonra verilerin analizine geçilmiştir. Veriler üzerinde yüzde, frekans gibi betimleyici analizlerin yanı sıra faktör analizi, madde güçlük indeksleri, madde ayırt edicilik indeksleri, güvenirlik katsayısı ve ortalama güçlük ile ortalama ayırt edicilik indeksleri hesaplanmıştır. Testlerin faktör analizleri R tabanlı Shiny uygulamasında yapılmıştır. Faktör yük değerleri .40'ın altında olan 5 madde okuma testinden, 10 madde dinleme testinden silinerek tek boyutlu testler elde edilmiştir. Güvenirlik ve ayırtedicilik açısından ise, okuma testinin genel olarak güvenirliği ve ayırt ediciliği yüksek, kolay; dinleme testinin ise güvenirlik ve ayırt ediciliğinin iyi düzeyde ve kolay bir sınav olduğu belirlenmiştir. Okuma testinin madde analizleri incelendiğinde 15 maddenin iyi, dinleme testinin madde analizleri incelendiğinde ise 9 maddenin iyi, 1 maddenin ise düzeltilerek teste alınabileceği anlaşılmıştır. Ulaşılan bulgulardan hareketle, okuma sınavını oluşturan maddelerin dinleme sınavının maddelerine göre daha iyi tepki verdiği, okuma sınavının önemli bir kısmının korunarak kullanılabileceği sonucuna ulaşılmıştır.

**Anahtar Kelimeler:** Yabancılara Türkçe öğretimi, ölçme ve değerlendirme, yabancılara Türkçe öğretiminde ölçme ve değerlendirme, Test geliştirme.

## Introduction

The problems in the field of assessment and evaluation in teaching Turkish as a foreign language have been discussed for many years. However, a significant part of these problems are still valid. There are issues with writing questions that are in line with cognitive taxonomy, teaching objectives, and learning outcomes. There are also worries about how valid and reliable the placement, passing, and proficiency exams are, which are used a lot in the field. Writing questions in accordance with the teaching objectives, learning outcomes and cognitive taxonomy and the concerns about the extent to which the placement, passing and proficiency exams, which are frequently used in the field, are prepared in a valid and reliable manner are among these problems. As in other fields, in teaching Turkish as a foreign language, the inability to make qualified measurement and evaluation, which constitutes the last step of the teaching process, leads to wrong and erroneous decisions. The observation that some learners whose Turkish proficiency is certified as "advanced level" according to the results of the "Turkish Proficiency Exam", "Certificate Exam" or "Diploma Exam" cannot demonstrate the language proficiency expected from them in accordance with their level in the communication environment or in the academic undergraduate and graduate education processes they start after the Turkish preparatory programme is a concrete indicator of the wrong decisions made as a result of faulty measurement. It is of great importance for the future academic success of individuals and institutions that measurements are made with measurement tools whose validity and reliability have been proven and which are appropriate for the levels and purposes. This is only possible through the development of standardised measurement tools.

Although standard or standardised tests may be perceived as difficult and unpleasant for both test preparers and practitioners, they are indispensable for healthy assessment. In its simplest form, the term "standardised" means that *"the content of the test is equivalent in all applications"* and that *"the conditions under which the test is administered are the same for all test participants"* (Sireci, 2005). While the measurement results made with instruments with these features provide correct decisions, the measurement results using non-standardised instruments are open to discussion in many respects. At this point, the question *"How can standardisation be achieved in tests?"* comes to mind. Sireci (2005) states that the logic in standardisation stems from the scientific method. In other words, a standardised test is an exam that is prepared, administered and analysed in accordance with scientific methods. The source of the differences in the measurement results of standardised tests is the differences in knowledge, skills and competences between individuals. These differences do not mean that the tests are not

standardised. However, the fact that the scoring is equal for everyone, that it is applied to all participants under the same conditions and that the scoring of the test in different applications is statistically and qualitatively equivalent means that the test applied is a standard test.

*Standardised tests* are tests in which students' scores are evaluated by comparing their scores with predetermined performance standards. When *standardised tests* are developed, they are administered to large samples, called *norm groups,* and the scores of this group provide standards for interpreting the scores of all other students taking the test. Exams in which student scores are interpreted based on the scores in the *norm group* are *norm-based* exams, and exams in which a cut-off score is determined and evaluated by test developers are *criterion-based* exams (Caldwell, 2008). Since *norm-based* exams compare student achievement or scoring with the scores of the group to which the exam is administered, participants in such exams are ranked according to the group to which the exam is administered. *Criterion-based* exams are exams that aim to measure the skill in a certain subject, with a defined passing score or acceptability level defined without depending on any norm group (Flippo, Armstrong & Schumm, 2018). *Criterion-based* exams are exams in which individuals are judged according to predetermined score ranges, such as *successful-failed; pass-fail; basic-intermediate-advanced* level. Since decisions are made according to predetermined score ranges based on the scores of the participants, placement tests, Turkish Proficiency Exam, Diploma Exam, and course exams applied by the relevant institutions in the field of teaching Turkish to foreigners are *criterion-based* exams. In these exams, which have "*extremely important consequences for students, teachers and schools*" and are therefore characterised as "*high-stake tests*" (Afflerbach, 2005, p.151), it is imperative that each item and each stage of the exam is valid and reliable in order for the decisions made as a result of the assessments to be correct.

*Test scores reliability* is generally expressed by the reporting of calculated reliability coefficients and the consistency of measurements (Flippo, Armstrong & Schumm, 2018; Gregory, 2014). One of the most important proofs of reliability is that a test administered to different groups with similar characteristics yields consistent results. In addition, as a result of such a test, it can be assumed that test practitioners and developers have accurately measured student performance, knowledge or behaviour. In order to prove test reliability, methods such as *test-retest, parallel (equivalent*) *forms* and *internal consistency coefficient,* which calculates the relationship between items and internal consistency coefficients between items, have been developed. KR-20, KR-21 and Cronbach Alpha coefficients are used to prove the reliability of a measurement tool. These coefficients take values between 0-1 and if the results are close to 0, reliability is *low*; if close to 1, reliability is interpreted as *high* (Güler, 2023).

No matter which method is used to prove reliability, the main point to be focussed on is that it has no value if it does not measure students' skills, strategies, knowledge and content knowledge related to the subject area. One of the criteria that must be proved in order for the prepared tests to be used is the *validity of the test*. The concept of *validity*, in general, is to prove that the instrument is capable of measuring the characteristic in terms of scope, structure and the relationship (criterion) between the scores obtained after the application. In other words, a measurement tool should not measure any other attribute other than the one intended to be measured. For example, a placement test administered to an international student who wants to learn Turkish should only determine his/her level of Turkish and a proficiency test administered to an international student who wants to learn Turkish should only determine his/her competences. In short, *validity* can be seen as the process of creating and evaluating positive or negative evidence for the intended interpretation of test scores and their relevance to the proposed use (APA, 2014). The first step in obtaining a highly valid test is to ensure content validity. Content validity is providing evidence as to whether the items that make up the test measure the entire target subject/feature. Obtaining expert opinion and creating a table of specifications are commonly used

to ensure the content validity of a test (Özgüven, 2011). In addition to being a clear plan that guides the preparation of the test, the specification table is also very useful in providing a framework for test development and mapping the processes associated with each stage of the scope (objectives/objectives, cognitive level, percentage of the question in the test) (Kaya, 2017). After ensuring that the scope of the test is ensured, the prepared test is applied to the target audience and the validity of the test is tested again with the analysis to be made on the quantitative data obtained. Presenting the evidence numerically usually provides more robust data. In order to determine whether a test is valid, *the validity coefficient* is obtained by calculating the relationship between the scores obtained from the measurement tool and the criteria/measures determined in accordance with the purpose. This coefficient takes a value between -1/+1 and the closer it is to +1, the more valid the measurement tool is and the more it serves its purpose (Bilican Demir, 2023). This refers to the construct validity of the test. Construct validity is checked by factor analysis in a suitable statistical program. With factor analysis, the relevant factor and the feature that measures it and its meaning are revealed.

When the literature on teaching Turkish as a foreign language is examined, it is determined that there are a few studies that can be considered within the scope of valid and reliable test development attempts, applied in small samples, and only basic analyses are performed (Tarı Yardımcı & Elmalı, 2021; Eke, 2023). The tests created in these studies are aimed at measuring listening skills at B1 level and reading skills at A2 level. Apart from these, there is no study prepared for proficiency or certification exams with item analyses or for proving C1 level Turkish knowledge. In this respect, the study is a first in the related literature. The study aims to prove the usability of the listening and reading tests of the C1 level proficiency exam prepared by Gazi University TÖMER by calculating the validity and reliability of the listening and reading tests and the difficulty and discrimaniton index of the items in the tests. Thus, the study is expected to fill an important gap in the literature. On the other hand, Gazi University TÖMER organises proficiency exams for those who want to certify their Turkish proficiency within the institution or outside the institution for the institutions with which it has a protocol. These exams are prepared and administered by academicians who are experts in their fields. Determining the reliability, item difficulty index and discrimaniton indexes of the entire exam and each question that constitutes the exam will provide an evidence-based understanding of the quality of the questions. Necessary precautions can be taken by revealing the quality of the exam. Thus, the present study also aims to reveal and interpret the coefficients of the analysis parameters related to the exam and to guide the determination of exam improvement policies from an institutional perspective. For these purposes, answers to the following sub-problems will be sought:

1. Is the reading test of Gazi University TÖMER proficiency exam valid?
2. Is the reading test of Gazi University TÖMER proficiency exam reliable?
3. What is the difficulty index of the reading test of Gazi University TÖMER proficiency exam?
4. What is the discrimination index of the reading test of Gazi University TÖMER proficiency exam?
5. Is the listening test of Gazi University TÖMER proficiency exam valid?
6. Is the listening test of Gazi University TÖMER proficiency exam reliable?
7. What is the difficulty index of the listening test of Gazi University TÖMER proficiency exam?
8. What is the discrimination index of the listening test of Gazi University TÖMER proficiency exam?

## Method

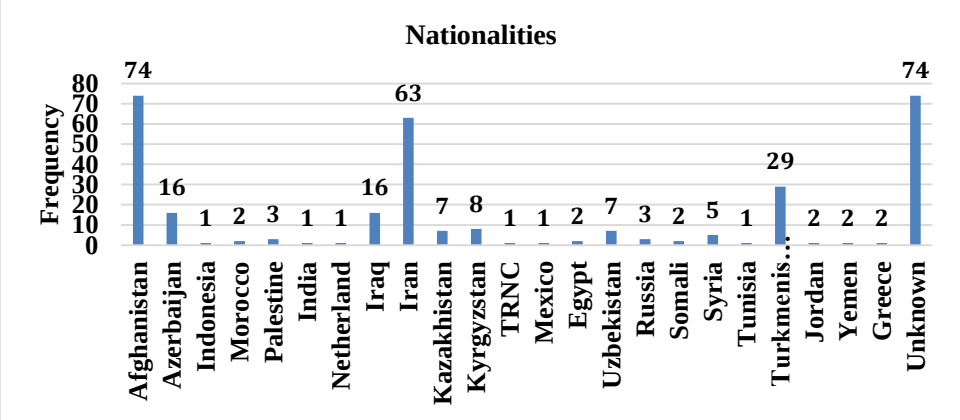
### Research Model

The research aims to reveal the validity and reliability of the reading and listening tests of the C1 certificate exam prepared by Gazi University TÖMER and used in teaching Turkish as a foreign language through test statistics and item analyses. The study was conducted in accordance with the survey model, one of the quantitative research methods. Survey researches aim to "*describe the existing situation generally related to the research subject by taking a picture of the existing situation*" (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz & Demirel, 2016). Since no experimental method was applied and there was no aim such as theory development and hypothesis verification, the survey model was used in the current study.

### Study Group of The Research

The study group of the research consists of the reading and listening test data of the C1 certificate exam administered face-to-face at Gazi University TÖMER between January 2023 and October 2023. Demographic information of the exam participants (candidates) is given below.

**Table 1.** Information about the nationalities of the participants




Based on Table 1, it is seen that Gazi University TÖMER C1 certificate exam was held with 250 students from 24 different countries in the relevant period. Iranian learners (63) participated in the exam the most. The other exam participants were from Turkmenistan (29), Iraq and Azerbaijan (16) in order from the highest to the lowest nationalities. 74 participants did not declare their nationality. The sample of the study was determined according to the convenience sample method, which is one of the non-random sampling methods. This method is a non-probabilistic, easily accessible and low-cost method (Schonlau, Fricker & Elliot, 2002).

### Data Collection Tools

The reading and listening skills tests in the C1 *Certificate Examination* of Gazi University TÖMER were used as data collection tools in the study. The tests consist of 20 questions each. In the preparation of the questions, the reading and listening competences in the "Common European Framework of Reference for Languages" (2020) and the level-appropriate learning outcomes for the relevant skills in the "Turkish as a Foreign Language Teaching Programme" were taken as basis. The stages in Figure 1 were followed in the development of the reading and listening tests.

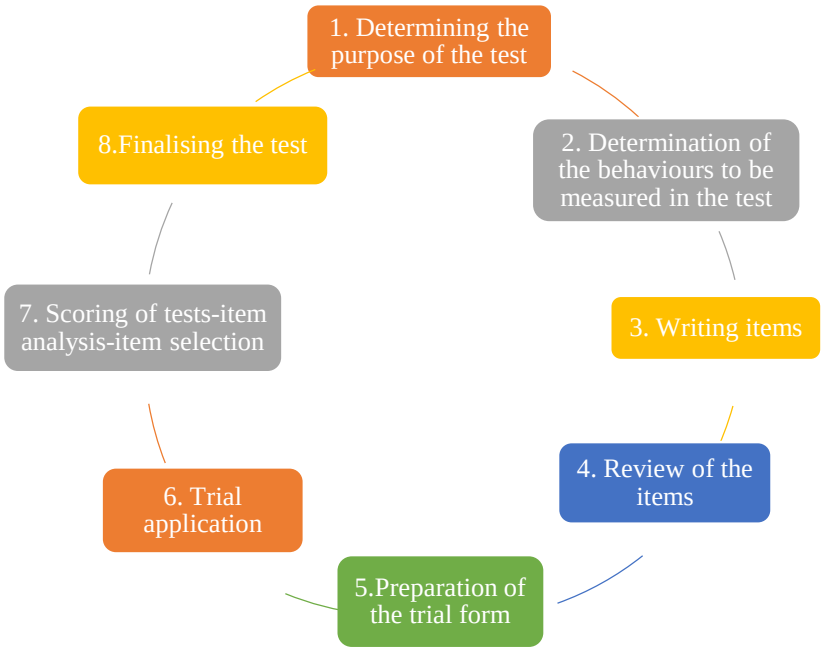



**Figure 1.** Test development stages (Turgut & Baykul, 2021)

### Analysing the Data

KR-20 coefficient was used to calculate the reliability indexes of the tests. Exploratory Factor Analysis (EFA) was conducted to provide evidence for construct validity. At this stage, Kaiser Meyer Olkin (KMO) and Barlett Sphericity test were used to determine the suitability of the data for factor analysis. Exploratory Factor Analyses of the tests were carried out using the Shiny application developed by Kılıç (2023), which is a programme that uses packages such as psych, lavaan, which is also included in the R programme in the background. Through this application, it was examined whether the items provided the factor predicted by the researchers. In order to ensure the content validity of the tests, specification tables were prepared and submitted to expert opinion.

### Validity and Reliability

Before the validity and reliability analyses, the data belonging to the reading and listening tests of the exam conducted for those who wanted to certify their Turkish proficiency at Gazi University TÖMER in a one-year period between January 2023 and October 2023 were entered into the Excel programme by scoring the correct answers as '1' and the wrong answers as '0'. Then, the data were transferred to SPSS 25.0 programme and reliability, item difficulty and item discrimaniton indexes were calculated for each test item and for the whole exam. While the discrimaniton index of the items was determined by the *corrected item-total correlation*, the discrimaniton index was obtained by calculating the correct answers given by each participant to the questions separately and dividing by the number of participants after the data set was transposed. Kuder Richardson-20 (KR-20) reliability coefficient was calculated to determine the internal consistency of the tests. The KR-20 reliability coefficient is sensitive to the degree of representativeness of the test scope and the homogeneity of the feature measured in the test, such as the two-half method. This formula is suitable for calculating reliability for scores obtained from tests consisting of items scored 1-0 (true-false) (Çıkrıkçı, 2022, p.80). In the study, since each test item was scored as 1-0 (true-false), the internal consistency coefficients of the tests were determined by the KR-20 formula. According to Pallant (2020), a reliability coefficient of .70 and above indicates that the items in the measurement tool are reliable. In addition to the KR-20 coefficient, McDonald Omega score was also calculated to prove reliability. After the items that were decided to be removed as a result of the factor analysis were deleted from the data set, the analyses were performed again.

**Content Validity**

Tables of specification were created to prove the content validity of the tests. The objectives measured by the questions were determined by the Turkish as a Foreign Language Teaching Programme (MoNE, 2020). The table of specifications was prepared in accordance with Bloom's taxonomy. The specification tables were sent to two PhD-level field experts who are experts in the field of teaching Turkish to foreigners and have field experience in Turkey and abroad. For the items that the field experts had difficulty in deciding on the cognitive level, an online meeting was held with the ZOOM programme. In order to ensure the face validity of the tests, the options were ordered from short to long, the negative expressions were underlined, the numbering was made in the same format and the verbs in the question stem were adapted to the cognitive level. The tables of specification for the tests are given in Table 2.

**Table 2**. Table of specification of the reading test

| Learning outcome of MoNE | Question number | Learning Outcome | Taxonomic Level |
|---|---|---|---|
| C1.O.30 | 1, 2, 3, 4, 5, 6, 16, 17, 18, 19 | Selects the desired information from a text. | Remember |
| C1.O.12 | 7,8 | Makes intratextual and/or intertextual comparisons. | Analyze |
| C1.O.33 | 9 | Determines the subject and main message of the text. | Synthesis |
| C1.O.21 | 10 | Identifies justified opinions and suggestions. | Analyze |
| C1.O.1 | 11,12 | Makes sense of the elements of vocabulary based on the context. | Comprehension |
| C1.O.17 | 13,14,15 | Understands texts (interviews, questionnaires, etc.) that request/report personal information and opinions. | Comprehension |
| C1.O.34 | 20 | Determines the main idea and auxiliary ideas of texts related to a field of specialisation. | Synthesis |

In the MoNE Turkish as a Foreign Language Teaching Programme (2020), there are 54 learning outcomes belonging to C1 level reading skill. In this study, the objectives in the programme were matched with the questions and presented to the expert opinion. The final version of the acquisition-cognitive level association reached as a result of expert opinions is given in the Table. After the table of specification of the reading test was completed, the specification table of the listening test was prepared to ensure the content validity of the listening test. Listening test specification table is given in Table 3.

**Table 3.** Table of specification of the listening test

| Learning outcome of MoNE | Question number | Learning Outcome | Taxonomic Level |
|---|---|---|---|
| C1.D.25 | 1,17,18,19,20 | Makes inferences about what he/she listens/watches. | Analyze |
| C1.D.27 | 2,14 | Identifies justified opinions and suggestions. | Comprehension |
| C1.D.8 | 3,4,5,6,7,8,9,10,11,12 | Selects the information he/she needs from audio and/or video news. | Comprehension |
| C1.D.13 | 13 | Determines the subject and main idea of narrative/informative texts. | Analyze |
| C1.D.19 | 15 | Compares what he/she listens/watches in terms of content. | Comprehension |

| | | | |
|---|---|---|---|
| C1.D.1 | 16 | Makes sense of the elements of vocabulary based on context. | Comprehension |

In the MoNE Turkish as a Foreign Language Teaching Programme (2020), there are 45 learning outcomes belonging to C1 level listening skill. It is seen that the listening test questions are distributed in 6 learning outcomes and 2 different cognitive levels.

**Construct Validity**

Construct validity is the degree to which the feature to be measured by the measurement tool can be measured without interfering with other features (Güler, 2023). In scale development studies, factor analysis is commonly used to prove construct validity. Factor analysis is used to determine whether the items that make up a measurement tool are suitable for a predetermined structure. This method of analysis is defined as *exploratory factor analysis* (Yurdabakan, 2022). In the study, *exploratory factor analysis* was performed since it was aimed to determine the extent to which the items that make up the tests measure the feature that the measurement tool wants to measure and to define the structures of the tests. The suitability of the data for factor analysis is decided by Keiser Meier Olkin and Barlett test. In order to perform factor analysis, KMO value should be greater than .05 and Barlett test should be significant ($p<.05$). Factor analysis can be performed after the conformity of both assumptions is proved.

**Difficulty and discrimaniton indexes**

The development of a valid and reliable test also depends on the quality of the items (questions) that make up the test. The properties of the items are determined by item analyses. When the related literature is examined, it is seen that the most common methods used in the calculation of item statistics are *the simple method and Henryson method*. In *the simple method*, test scores are ordered from highest to lowest; item analyses are performed based on the data in the 27% with the highest score and the data in the 27% with the lowest score. The 46% section in the middle of the sorted data set is not included in the calculation. In *the Henryson method*, all data are included in the analysis without any distinction on the data set. Therefore, this method is considered to be more reliable than *simple method* (Başol, 2019; Hasançebi, Terzi & Küçük, 2020). The statistical analyses conducted in the current study were carried out based on the Henryson method because it is practical and economical in terms of time.

Item analysis is a process in which the responses of test participants to each item in the test are measured and index scores are determined to make interpretations about the items and the test as a whole. The index traditionally calculated in item analyses are item difficulty index and item discrimaniton index (de Gruijter & van der Kamp, 2008). The item difficulty index is defined not according to the perceived difficulty of the items or the effort required to answer them, but according to the probability of the correct answer (DeMars, 2010). Difficulty index takes a value between 1.00 and 0. If an item is answered correctly by few people, it indicates that the item is difficult; if it is answered correctly by many people, it indicates that the item is easy. Difficulty indexes and their interpretation are given in Table 4.

**Table 4.** Item difficulty indexes ($p_j$) and comments

| Value | Comment |
|---|---|
| .85 – 1.00 | Very easy item (should be removed from the test) |
| .61 – .84 | Easy item (can be made more difficult according to need) |
| .40 - .60 | Medium item (ideal item) |
| .39 – .16 | Difficult item (can be facilitated according to need) |
| .15 - .00 | Very difficult item (should definitely be removed from the test) |

(Başol, 2019: 247)

The discrimination index of the test items is a value calculated to determine the level of differentiation between the participants with different levels of constructs; in other words, to distinguish between those who know and those who do not know. Therefore, it is always desired by researchers that items in tests have a high level of discrimaniton index (DeMars, 2010). The values of the discrimination indexes and their interpretations are presented in Table 5.

**Table 5.** Item discrimination indexes ($r_{jx}$) and comments

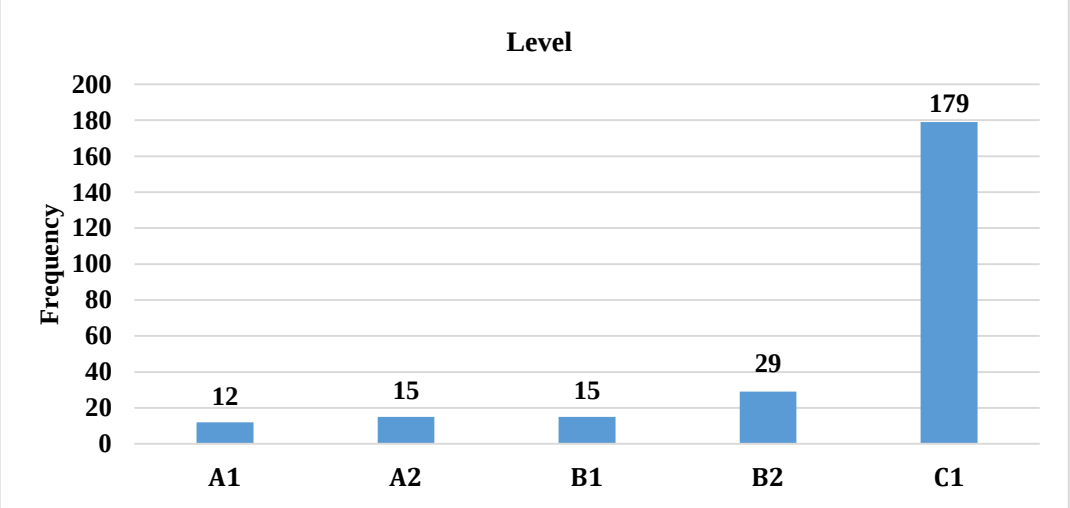
| Value | Comment |
|---|---|
| 0 - .19 | The item has low discrimaniton; it should not be included in the test. |
| .20-.29 | The item can be corrected and included in the test. |
| .30 ve üzeri | Item discrimaniton is high; it should be included in the test. |

*(Turgut & Baykul, 2021: 228)*

## Findings

Before proceeding to the item analyses in the tests, it was determined at which levels the test participants received appropriate scores as a result of the application. The findings related to this are given below.

**Table 6.** Information about the level of the test participants




According to the descriptive analyses, based on the exam scores, it was determined that there were 12 participants at A1 level, 15 at A2 level, 15 at B1 level, 29 at B2 level and 179 at C1 level. However, certification was not made for all levels, but for C1 level and students enrolled in departments that accept B2 level teaching in a foreign language. After the descriptive findings related to the exam results were determined, the normal distribution of the data was analysed. The findings related to this are given below.

**Table 7.** Information on the normal distribution of the data

| | skewness | skewness error | z_skewness | kurtosis | kurtosis error | z_ kurtosis |
|---|---|---|---|---|---|---|
| Reading test | -.532 | .154 | -3.45 | -.600 | .307 | -1.95 |
| Listening test | -.398 | .154 | -2.58 | -.022 | .307 | -.07 |

Field (2009) states that when one of the skewness and kurtosis Z scores is greater than 1.96, the data do not show normal distribution at the level of .05. When Table 5 is analysed, it is seen that the kurtosis z score of the reading test is -1,95 and the kurtosis z score of the listening test is -.07. It is proved that the kurtosis z scores of both tests are less than 1.96 and the data are normally distributed.

**Findings for the Reading Test**

***Construct Validity: Exploratory Factor Analysis***

When the related literature is examined, there are different opinions about the sample size that should be reached for factor analysis. However, it is accepted that the sample size should be 5-10 times the number of items (Kass & Tinsley, 1979; Kline, 1994; Pett, Lackey & Sullivan, 2003; Tavşancıl, 2005). In addition for tests consisting of a small number of items (e.g. 20 or less items), a sample size in the range of 100-200 is considered sufficient for basic item and factor analyses (Netemeyer, Bearden & Sharma, 2003). Since the size of the data set in which the analyses were performed consisted of 250 participants and the exam form consisted of 20 questions, it was decided to conduct the analyses considering that the required sample size was reached. KMO coefficient and Barlett Sphericity Test results were used to prove the suitability of the sample size for factor analysis. The related findings are given in Table 8.

**Table 8.** KMO and Barlett test results

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin Test | | 0.821 |
| Barlett Sphericity Test | chi-square | 1918.29 |
| | Sd | 91.00 |
| | p | 0.00 |

According to Kaiser (1974), KMO value should be greater than 0.5 and according to Pallant (2000), this value should be greater than 0.6 for factor analysis. When Table 8 is analysed, it is seen that the KMO value of the data is higher than the value predicted in the sources (0.821). When the Barlett Sphericity Test results in the same table are analysed, it is proved that Barlett's test is significant ($x2 = 1918.29$; $p = 0.00$); that is, the data show multivariate normal distribution. These analyses proved that the data were suitable for factor analysis.

After the suitability of the test data for factor analysis was proved, a structure explaining 57% of the variance, with a KMO value of 0.79 and Barlett's Test ($p<.05$) significant, was obtained as a result of the factor analysis performed in the Shiny (Kılıç, 2023) application of the test created in a single dimension. The item factor loadings and eigenvalues that emerged after the factor analysis of the first version of the draft test are given in Table 9.

**Table 9.** Factor loading values and common factor variance

| **items** | **factor loading** | **common variance** |
|---|---|---|
| item 1 | .57 | .81 |
| item 2 | .73 | .81 |
| item 3 | .57 | .79 |
| item 4 | .73 | .82 |
| item 5 | .34 | .73 |
| item 6 | .65 | .81 |
| item 7 | .52 | .81 |
| item 8 | .36 | .78 |
| item 9 | .17 | .64 |
| item 10 | .43 | .71 |
| item 11 | .19 | .54 |
| item 12 | .59 | .82 |
| item 13 | .68 | .83 |
| item 14 | .54 | .79 |
| item 15 | .52 | .83 |

| | | |
|---|---|---|
| item 16 | .46 | .74 |
| item 17 | .62 | .81 |
| item 18 | .29 | .59 |
| item 19 | .54 | .81 |
| item 20 | .68 | 84 |

*Eigenvalue 5,70*
*Explained variance: 29.0%*

In order for an item to be represented in a factor, factor loading values should be .40 and above (DeVellis, 2003; Field, 2005). When Table 9 is analysed, it is seen that the factor loading values of item 5, item 8, item 9, item 11 and item 18 are less than .40. Therefore, these items were removed from the data set and exploratory factor analysis was performed on the remaining items. The findings related to this analysis are given in Table 10.

**Table 10.** Factor analysis results after deleted items

| **items** | **factor loading** | **common variance** |
|---|---|---|
| item 1 | .62 | .84 |
| item 2 | .77 | .82 |
| item 3 | .59 | .80 |
| item 4 | .77 | .83 |
| item 6 | .68 | .80 |
| item 7 | .51 | .83 |
| item 12 | .57 | .84 |
| item 13 | .67 | .81 |
| item 14 | .57 | .82 |
| item 15 | .51 | .84 |
| item 16 | .43 | .80 |
| item 17 | .59 | .83 |
| item 19 | .52 | .83 |
| item 20 | .64 | .82 |

*Eigenvalue: 5.21*
*Explained variance: 37%*

The values obtained as a result of the factor analysis conducted after the deletion of the related items in the data set are given in Table 10. According to this, the factor loading values of the final version of the test vary between .43 and ,77. This shows that the items were gathered in a single dimension and the validity of the construct was proved. It is seen that the common variance values are between 0-1. A common variance between 0-1 indicates that the items measure the same construct and that the items are related to each other (Tabachnick & Fidell, 2007). Other analyses of the reading test were conducted on the remaining items.

### Findings Related to the Reliability of the Reading Test

The reliability of the test was tested on the form formed as a result of the deleted items after the factor analysis. When the literature is analysed, it is seen that the reliability coefficients should be .70 and above (Nunnally, 1978; Pallant, 2020; Fraenkel & Wallen, 2009). As a result of the calculation, the KR-20 coefficient of the test was calculated as .74. In addition to the KR-20 coefficient, the McDonald Omega coefficient was calculated as .79. These findings reveal that the reading test is reliable.

### Findings Related to the Item Discrimination Index (rjx) of the Reading Test

Cronbach alpha internal consistency coefficient was used to determine the discrimaniton index of the reading test items. The item-total correlations Cra reliability coefficients for each item in the test form are given in Table 11.

**Table 11.** Item discrimination indexes and reliability coefficients of the reading test

| Items | Item discrimaniton coefficient ($r_{jx}$) | Cra Reliability Coefficient when the item is removed | Comment on the item (according to the index of discrimaniton) | Item difficulty index ($p_j$) | Comment on the item (according to difficulty index) |
|---|---|---|---|---|---|
| item 1 | .316 | .783 | Good item | .84 | Easy |
| item 2 | .484 | .773 | Good item | .67 | Easy |
| item 3 | .320 | .783 | Good item | .86 | Very easy |
| item 4 | .481 | .773 | Good item | .73 | Easy |
| item 6 | .429 | .776 | Good item | .74 | Easy |
| item 7 | .337 | .782 | Good item | .74 | Easy |
| item 10 | .361 | .781 | Good item | .59 | Medium |
| item 12 | .433 | .776 | Good item | .56 | Medium |
| item 13 | .475 | .773 | Good item | .60 | Medium |
| item 14 | .359 | .781 | Good item | .53 | Medium |
| item 15 | .373 | .780 | Good item | .55 | Medium |
| item 16 | .354 | .781 | Good item | .62 | Easy |
| item 17 | .443 | .775 | Good item | .72 | Easy |
| item 19 | .385 | .779 | Good item | .68 | Easy |
| item 20 | .494 | .772 | Good item | .61 | Easy |
| Average difficulty index of the reading test | | | | .67 | Easy |

When the discrimination indexes in Table 11 are interpreted according to the values in Table 5, all of the items can be included in the test. According to these findings, 15 items of the 20 items reading test can be used in the test form according to the discrimination index coefficient.

**Findings Related to the Item Difficulty Index (pj) of the Reading Exam**

There are various methods to calculate item statistics of tests. The most commonly used of these methods are *Henryson Method and Simple Method*. The main point where these two item analysis methods differ from each other is the number of measurements used in the calculations, that is, the number of samples included in the calculation. In the Henryson Method, all respondents included in the measurement are used, whereas in the simple method 54% of the population is used by determining the top 27% most successful and 27% least successful subgroups from the total scores of all respondents. Henryson Method gives reliable results even in small samples. According to Henryson Method, item difficulty index is the ratio of the number of correct answers to the number of all respondents. When the ratio is made, it is seen what percentage of the class answered the question correctly. This index can take values between 0 and 1. As the difficulty index approaches 0, it can be interpreted that the item is a difficult item, and as it approaches 1, it can be interpreted that the item is an easy item. Information about the interpretation of the items according to the item difficulty index is given in Table 4.

When the difficulty index scores of the items in Table 11 are interpreted according to Table 4, it is seen that item 1, item 2, item 4, item 6, item 7, item 16, item 17, item 19 and item 20 are **easy** items; item 10, item 12, item 13, item 14 and item 15 are of **medium** ; item 3 is very easy. It is seen that the test form generally consists of **easy** questions.

### Findings Related to Listening Test

#### *Construct Validity: Exploratory Factor Analysis*

Before the factor analysis of the listening test data, as in the reading test, it was first examined whether it met the prerequisites for factor analysis. For this reason, KMO and Barlett Sphericity Test analyses were performed first. The findings related to this are given in Table 12.

**Table 12.** Listening test KMO and Barlett test results

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin Test | | 0.77 |
| Barlett Sphericity Test | chi-square | 13437.74 |
| | Sd | 190.00 |
| | p | 0.00 |

According to Kaiser (1974), KMO value should be greater than 0.5 and according to Pallant (2001), this value should be greater than 0.6 for factor analysis. When Table 6 is analysed, it is seen that the KMO value of the data is higher than the value predicted in the sources (0.77). When the Barlett Sphericity Test results in the same table are analysed, it is proved that Barlett's test is significant (x2 = 13437.74; p = 0.00) and the data show multivariate normal distribution. These analyses proved that the data were suitable for factor analysis. Then, the construct validity was tested by factor analysis. The factor analysis results of the first version of the test are given in Table 13.

**Table 13.** Listening test Factor loadings and common factor variance

| **items** | **factor loading** | **common variance** |
|---|---|---|
| item 1 | -.34 | .77 |
| item 2 | -.14 | .61 |
| item 3 | .87 | .84 |
| item 4 | .93 | .86 |
| item 5 | .83 | .85 |
| item 6 | .40 | .65 |
| item 7 | -.18 | .59 |
| item 8 | .64 | .74 |
| item 9 | -.34 | .70 |
| item 10 | .26 | .59 |
| item 11 | .63 | .81 |
| item 12 | .45 | .72 |
| item 13 | .24 | .72 |
| item 14 | -.30 | .67 |
| item 15 | -.17 | .59 |
| item 16 | -.70 | .88 |
| item 17 | .27 | .71 |
| item 18 | .82 | .83 |
| item 19 | -.30 | .78 |
| item 20 | .43 | .76 |

*Eigenvalue: 5,52*
*Explained variance: 28 %*

In order for an item to be represented in a factor, factor loading values should be .40 and above (DeVellis, 2003; Field, 2005). When Table 13 is examined, since the factor loading values of item 1, item 2, item 7, item 9, item 10, item 13, item 14, item 15, item 17 and item 19 were less than .40, these items were removed from the data set and exploratory factor analysis was performed again. The results of the analyses after the items were removed are given in Table 14.

**Table 14.** Factor analysis results after deleted items

| **items** | **factor loading** | **common variance** |
|---|---|---|
| item 3 | .85 | .85 |
| item 4 | .88 | .84 |
| item 5 | .80 | .87 |
| item 6 | .41 | .76 |

| | | |
|---|---|---|
| item 8 | .69 | .78 |
| item 11 | .65 | .91 |
| item 12 | .50 | .76 |
| item 16 | -.61 | .88 |
| item 18 | .87 | .84 |
| item 20 | .50 | .86 |

*Eigenvalue: 5,11*
*Explained variance: 43%*

The values obtained as a result of the factor analysis conducted after deleting the related items in the data set are given in Table 14. Accordingly, the factor loadings of the final version of the test varied between .40 and ,88. As a result of the analyses, a one-dimensional listening test consisting of 12 questions, explaining 43% of the variance, with a KMO value of 0.83 and a significant Barlett's test ($p<.05$), was obtained. In addition, it was proved that the common variance values were between 0-1, the items measured the same construct and were related to each other.

**Findings Related to the Reliability of the Listening Test**

The reliability of the listening test was analysed with the 12 items data set obtained as a result of factor analysis. As a result of the calculation, the KR-20 coefficient of the test was calculated as .70. When the literature is examined, it is the opinion that reliability coefficients should be .70 and above (Nunnally, 1978; Pallant, 2020). The McDonald Omega coefficient of the test, which has sufficient reliability according to the KR-20 coefficient, was calculated as .73. This finding shows that the reliability of the listening test is at a sufficient level.

**Findings Related to Item Discrimination (rjx) and Difficulty Index (pjx) of Listening Test**

Item-total correlation coefficients were used to calculate the discrimaniton index of the items of the listening test. The item-total correlations and Cra reliability coefficients analysed for each item in the test form are given in Table 15.

**Table 15.** Item discrimination index and reliability coefficients of the listening test

| **Items** | **Item discrimination coefficient (rjx)** | **Cra Reliability Coefficient when the item is removed** | **Comment on the item (according to the index of discrimination)** | **Item difficulty index (pj)** | **Comment on the item** |
|---|---|---|---|---|---|
| item 3 | .60 | .64 | Good item | .56 | Medium |
| item 4 | .56 | .64 | Good item | .59 | Medium |
| item 5 | .52 | .65 | Good item | .53 | Medium |
| item 6 | .26 | .69 | Need to be corrected | .34 | Difficult |
| item 8 | .39 | .67 | Good item | .81 | Easy |
| item 11 | .43 | .67 | Good item | .55 | Medium |
| item 12 | .30 | .69 | Good item | .89 | Very easy |
| item 16 | -.38 | -.38 | Good item | .66 | Easy |
| item 18 | .58 | .64 | Good item | .70 | Easy |
| item 20 | .37 | .68 | Good item | .66 | Easy |
| Average difficulty index of the listening test | | | | .66 | Easy |

When the discrimination indexes in Table 15 are interpreted according to Table 5, item 3, item 4, item 5, item 8, item 11, item 12, item 16, item 18 and item 20 are **good items**; and item 6 is **need to be corrected.** When the item difficulty index scores of the items in the listening test in Table 15 are interpreted according to Table 4; item 6 is **difficult**, item 3, item 4, item 5, item 11 are **medium**, m8,

item 16, item 18, item 20 are **easy** and item 12 is **very easy**. The average difficulty index of the listening test was calculated as .66. In general, it can be said that the listening test is easy.

## Discussion and Result

In this study, the validity and reliability analyses of the reading and listening tests used in Gazi University TÖMER C1 Certificate Examination, one of the institutions teaching Turkish to foreigners, were conducted. In order to prove the validity of the tests, expert opinion was consulted, and questions were prepared based on the competencies and achievements in CEFR (2020) and MoNE Turkish as a Foreign Language Teaching Programme (2020). The answers given to the test items by the participants of the practices carried out at Gazi University TÖMER on different dates between January-October 2023 constituted the data of the study.

As a result of the item analysis of the reading test consisting of 20 questions, the KR-20 coefficient calculated to determine the reliability of the test in question was .74; McDonald Omega coefficient was determined as .79. These values reveal that the reading test is reliable. The construct validity of the test was tested with factor analysis. As a result of the analysis, m5, m8, m9, m11 and m18, whose factor load values were below .40, were removed from the data set. According to the discrimination index made on the remaining items, all of the items are good; item 3 is very easy, item1, item2, item4, item6, item7, item16, item17, item19, item20 are easy; item10, item12, item13, item14 and item15 are medium difficulty items. The average difficulty of the reading test is .67; was determined to be easy. The ideal approach is to test items with a difficulty index value of .40-.60 and items with a discrimination index of .30 and above. However, it is not always possible for all items to be of medium difficulty and high discrimination. Easy and difficult items should be taken in accordance with the purpose of the test, provided that their discrimination is high (Karaca, 2022). In general, the reading test of the C1 level certification exam is a useful test in distinguishing between participants with and without C1 level. It is an exam that distinguishes those who know and those who do not, or those whose level is C1 and those who do not.

As a result of the item analysis of the listening test consisting of 20 questions, the KR-20 coefficient calculated to determine the reliability of the test was .70; McDonald Omega coefficient was calculated as .73. These values reveal that the listening test is reliable. The construct validity of the test was tested with factor analysis. As a result of the analysis, item1, item2, item7, item9, item10, item13, item14, item15, item17 and item19, whose factor load values were below .40, were deleted from the data set. According to the discrimination index made on the remaining items, item6 should be corrected, and the remaining items are good items. According to the difficulty index, item12 is very easy; item8, item16, item18, item20 easy; item3, item4, item5, item11 are medium difficulty items and item6 are difficult items. The average difficulty level of the listening test is .66. According to this finding, it was concluded that the listening test was easy. According to the factor analysis performed on the test form consisting of items other than the deleted items, a valid, single-dimensional listening test was obtained. The substances that make up the test can be included in the tests according to their stated characteristics.

Based on the findings analysed and discussed for use in certificate exams at the C1 level in teaching Turkish to foreigners, it is possible to say that the reading test has high reliability and discrimination, while the listening test has low reliability, weak discrimaniton and is an easy test. While creating the final test form of the reading test, easy questions with high discrimaniton can be selected. However, the overall listening test needs to be revised and corrected. While the reasons such as the length of the texts used in the listening test, the number of words and syllables are problems arising from the test, the fact that the Turkish listening skills of the test participants are developed is also a reason for

the participants to evaluate the test as easy. When the literature was reviewed, no test development study was found to be used in teaching Turkish as a foreign language, except for the studies of Eke (2023) and Yardımcı Tarı & Elmalı (2021). This situation, which stands out as a major deficiency in the literature, causes passing, placement, proficiency and certificate exams to be conducted with exams whose validity and reliability are controversial and whose difficulty and discrimination levels are unclear. The biggest reason why teaching Turkish as a foreign language is not recognised internationally is the lack of measurement and evaluation tools in accordance with international criteria and the lack of an appropriate teaching process. Institutions teaching Turkish as a foreign language should organise in-service trainings for question-exam preparers, and feedback should be provided by analysing the exams applied. In this way, the deficiencies in the process can be determined precisely and clearly and the necessary arrangements can be made quickly.

**Research and Publications Ethics**

In this study, all rules specified in the Higher Education Institutions Scientific Research and Publication Ethics Directive were followed. None of the actions described under the title of Actions Contrary to Scientific Research and Publication Ethics in the Directive have been carried out.

**Ethics Commitee Permission**

This study was carried out within the scope of the permission of Gazi University Ethics Committee, document number E-77082166-604.01-919774, dated 26.03.2024.

**Contribution Rate of Authors**

The authors contributed equally to the study.

**Conflict of Interest**

There is no conflict of interest.

## References

Afflerbach, P. (2005). National reading conference policy brief: High stakes testing and reading assessment. *Journal of Literacy Research, 37*(2), 151-162.

American Educational Research Association & American Psychological Association & National Council On Measurement in Education (2014). *Standarts for Educational and Psychological Testing.* Washington: Americak Educational Research Association.

Başol, G. (2019). *Eğitimde ölçme ve değerlendirme. [Measurement and evaluation in education].* Ankara: Pegem Academy.

Bilican Demir, S. (2023). Ölçmede geçerlilik. [Validity in measurement] *Eğitimde ölçme ve değerlendirme.* (p. 53-68) içinde. Ankara: Anı Publication.

Büyüköztürk, Ş., Çakmak Kılıç, E., Akgün, Ö. E., Karadeniz, Ş. & Demirel, F. (2016). *Bilimsel araştırma yöntemleri. [Scientific research methods].* Ankara: Pegem Academy.

Caldwell, J. S. (2008). *Reading assessment: A primer for teachers and coaches* (2nd edition). New York: The Guilford Press.

Council of Europe (2020). *Common European framework of reference for languages: Learning, teaching, assessment-Companion volume.* Council of Europe Publishing, Strasbourg.

Çıkrıkçı, N. (2022). Ölçmede güvenilirlik. *[Measurement and evaluation in education]* In (69-90). Ankara: Anı Publication.

De Gruijter de, D. N. M. & van der Kamp, Leo, J. Th. (2008). *Statistical test theory for he behavioral sciences.* Cahpman & Hall/CRC, Taylor & Francis Group.

DeMars, C. (2010). *Item response theory: Understanding statistics measurement*. New York: Oxford University Press.

DeVellis, R. F. (2003). *Scale development: Theory and applications.* CA: Sage.

Eke, H. (2023). Developing a reading exam for the A2 level of teaching Turkish as a foreign language. (Unpublished MA Thesis). Bartın University Graduate School.

Field, A. (2005). *Discovering statistics using SPSS (2nd edition).* London: Sage Publications.

Field, A. (2009). *Discovering statistics using SPSS (3rd. Edition).* London: Sage Publications.

Flippo, Rona F., Armstrong, Sonya L. & Schumm, Jeanne S. (2018). *Reading Tests.* Handbook of College Reading And Study Strategy Research (3th edition), Eds. (Rona F. Flippo & Thomas W. Bean) p.340-366. Routledge: New York.

Fraenkel, J. R. & Wallen, N. E. (2009). *How to design and evaluate research in education.* (7th edition). McGraw-Hill Higher Education.

Gregory, R. J. (2014) *Psychological testing: History, principles and applications* (7th global edition). Pearson.

Güler, N. (2023). *Eğitimde ölçme ve değerlendirme. [Measurement and evaluation in education].* Ankara: Pegem Academy.

Hasançebi, B., Terzi, Y. & Küçük, Z. (2020). Distractor analysis based on item difficulty index and item discrimination index. *Gümüşhane University Journal of Science and Technology Institute, 10*(1), 224-240.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31.36.

Karaca, E. (2020). *Test ve madde analizi.* [Testing and item analysis]. Eğitimde Ölçme ve Değerlendirme (Eds. S. Erkan & M. Gömleksiz). Ankara: Nobel Publication.

Kass, R. A. & Tinsley, H. E. A. (1979). Factor analysis. *Journal of Leisure Resrach, 11*(2), 120-138.

Thorndike, R. M. & Christ-Thorndike, T. (2017). *Ölçme sürecinde aranan özellikler: Geçerlik.* M. Otrar (Çev. Ed.) and M. Kaya (Çev.) *Measurement and evaluation in psychology and education* (8. Edition). (154-198) in. Nobel Publication (Original work published 2010).

Kline, P. (1994). *An easy guide to factor analysis.* New York: Routledge.

Kılıç, A. F. (2023). *Factor analysis for all (FAFA) [Software].* https://afarukkilic.shinyapps.io/Factor_Analysis_For_All_FAFA/

Ministry of National Education (2020). *Türkçenin Yabancı dil olarak öğretimi programı. [Teaching Turkish as a Foreign Language Program]* Ankara.

Netemeyer, Richard, G., Bearden, Willim, O. & Sharma, S. (2003). *Scaling procedures: Issues and applications.* California: Sage Publications.

Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw Hill.

Özgüven, İ. E. (2011). *Psikolojik testler. [Psychological tests].* Ankara: PDREM Yayınları.

Pallant, J. (2000). *SPSS survival manual.* Buckingham: Open University Press.

Pett, M. A., Lackey, N. R. & Sullivan, J. J. (2003). *Making sense of analysis: The use of factor analysis for instrument development in health care research.* Sage Publications.

Schonlau, M., Ronald, D. F. & Marc, N. E. (2002). *Conducting research surveys via e-mail and the web.* Santa Monica, CA: Rand Corporation.

Sireci, Stephen G. (2005). The most frequently unasked questions about testing. *Defending standardized testing* (Ed. Richard P. Phelps) içinde s. 111-123, Lawrence Erlbaum Associate: New Jersey.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics (5th Edition).* Pearson.

Tavşancıl, E. (2005). *Tutumların ölçülmesi ve SPSS ile veri analizi. [Measuring attitudes and data analysis with SPSS]*. Ankara: Nobel Publishing.

Turgut, M. Fuat & Baykul, Y. (2021). *Eğitimde ölçme ve değerlendirme. [Measurement and evaluation in education]*. Ankara: Pegem Academy.

Yardımcı Tarı, B. & Elmalı, M. (2021). Development of a B1 listening test for learners of Turkish as a foreign language. *Journal of Linguistics, 36,* 1-21.

Yurdabakan, İ. (2022). Eğitimde Kullanılan Ölçme Araçlarının Nitelikleri. [Qualifications of measurement tools used in education]. S. Erkan & M. Gömleksiz (Eds.), *Eğitimde Ölçme ve Değerlendirme* In (37-66). Ankara: Nobel Publication.