



Yabancı Dil Olarak Türkçe Öğretiminde B1 Düzeyinde Dinleme Sınavı Geliştirme¹

Development of a B1 Listening Test for Learners of Turkish as a Foreign Language

Berna TARI YARDIMCI* , Murat ELMALI** 



¹Bu çalışma 2021 yılında İstanbul Üniversitesi Sosyal Bilimler Enstitüsü Yabancı Dil Olarak Türkçe Öğretimi Anabilim Dalında, Doç. Dr. Murat Elmali danışmanlığında Berna Tari Yardımcı tarafından hazırlanan "Yabancı Dil Olarak Türkçe Öğretiminde B1 Düzeyinde Dinleme Sınavı Geliştirme" adlı yüksek lisans tezinden türetilmiştir.

* (Öğr. Gör.), Fatih Sultan Mehmet Vakıf Üniversitesi, Yabancı Diller Bölümü, İstanbul, Türkiye

** (Doç. Dr.), İstanbul Üniversitesi, Edebiyat Fakültesi, Dilbilimi Bölümü, İstanbul, Türkiye

ORCID: B.T.Y. 0000-0002-4422-7026;
M.E. 0000-0001-7097-3738

Sorumlu yazar/Corresponding author:
Berna TARI YARDIMCI,

Fatih Sultan Mehmet Vakıf Üniversitesi, Yabancı Diller Bölümü, İstanbul, Türkiye
E-posta: berna.tari@hotmail.com

Başvuru/Submitted: 10.05.2021
Kabul/Accepted: 24.05.2021

Atf/Citation: Tari Yardımcı, B. Ve Elmali, M. (2021). Yabancı dil olarak türkçe öğretiminde B1 düzeyinde dinleme sınavı geliştirme. *Dilbilim Dergisi - Journal of Linguistics*, 36, 1-21.
<https://doi.org/10.26650/jol.2021.935593>

ÖZ

Türkçenin yabancı dil olarak öğrenimi yurtiçi ve yurtdışında gittikçe çoğalmaktadır. Bu talep artışı beraberinde öğretmen ve materyal arayışını da getirmiştir. Materyal hazırlamada kitaplara ve eğitim programlamasına odaklanılmış, sınav hazırlama konusunda yürütülen çalışmalar sınırlı kalmıştır. Günümüzde uluslararası geniş kitleye uygulanabilecek tek bir Türkçe yeterlik sınavı bulunmamaktadır. Bu nedenle bu çalışmada yabancı dil olarak Türkçe öğretiminde uluslararası geçerlikte kullanılabilir B1 dinleme sınavı geliştirmek hedeflenmiştir. Bu amaçla, Avrupa Ortak Başvuru Metni B1 düzeyi, Weir (2005) sosyal-bilişsel çerçevesi, güvenilirliği ve geçerliği kanıtlanmış, yaygın olarak kullanılmakta olan diğer dil sınavları incelenerek tüm bir sınav hazırlanmıştır. Çalışmanın ilk bölümünde dinleme tanımı, yabancı dil olarak Türkçe'de dinleme, dinleme becerisinde karşılaşılan problemler ve sınav geliştirme konuları üzerinde durulmuştur. İkinci bölümde nasıl sınav hazırlanacağı üzerine yoğunlaşmış ve farklı dillerde yaygın olarak kullanılan yeterlik sınavları incelenmiştir. Hazırlanan sınav 31 katılımcıya uygulanmış ve üçüncü bölümde katılımcıların yanıtları doğrultusunda sınav incelenmiştir. İncelemede uygulamanın ardından katılımcılara ve sınavı gerçekleştiren öğretmenlere sınavla ilgili yönlendirilen değerlendirme formu da kullanılmıştır. Sınav ve form sonuçları doğrultusunda sınavın geçerliği ve güvenilirliği tartışılmıştır. Son olarak sınavın sınırlılıkları ve ileride kullanılabilir versiyonları için öneriler sunulmuştur. Uygulamada kullanılan sınav harici bu formatta hazırlanan diğer örnek sınavlar ekte öneri olarak sunulmuştur.

Anahtar kelimeler: Yabancı dil olarak Türkçe, dinleme becerisi, sınav geliştirme, B1 düzeyi, dinleme sınavı

ABSTRACT

The number of learners of Turkish as a foreign language is increasing, which emphasizes the need for teachers and materials. However, the preparation of course books and curriculum development are emphasized while preparing materials. The number of studies relevant to test development is quite limited. There is no single Turkish examination that can be used internationally. This study aims to develop an internationally valid B1-level Turkish listening examination for learners of Turkish as a foreign language. For this purpose, the Common European Framework's (CEFR) B1 level, Weir's (2005) socio-cognitive framework, and other language examinations that are being used internationally, which



were approved as being valid and reliable were examined, and an entire examination was developed with reference to the aforementioned tests. The first section of this paper focuses on the definition of listening, listening in Turkish as a foreign language, problems related to listening skills and examination development. In the second section, the ways to develop an examination as well as commonly used examinations in other languages are discussed. The third section details the administering of the listening examination created for this study to 31 participants. The section then continues to discuss and analyze the examination based on participants' responses. An examination evaluation form that was administered after the examination to the participants and the teachers who conducted the examination is also discussed, as well as the examination's validity and reliability based on its results and the evaluation form. Finally, limitations and suggestions for further versions of the examination are discussed. In addition to the examination developed for this study, samples of other examinations using the same format are provided in the addendum.

Keywords: Turkish as a foreign language, listening skills, test development, B1 level, listening test

EXTENDED ABSTRACT

Language tests can be applied for different purposes, ranging from placement to proficiency determination. Developing a valid and reliable test helps students and curriculum developers to achieve their language-learning goals. For learners of Turkish as a foreign language, the examinations that are currently being used are commonly prepared within language-learning centers or schools. In addition, the question formats and the listening texts are not standardized. Further, listening can sometimes be taught by the teachers in an environment fraught with distractions, which can cause difficulties in hearing and following the recordings for the test-takers. For these reasons, the Yunus Emre Institute developed a Turkish proficiency examination; this is the most common examination for Turkish administered in other countries. However, this examination is still not considered an international benchmark. This study aims to provide a format for listening examinations that can be used in Turkey and other countries.

In order to achieve this, valid and reliable examinations, such as the International English Language Test (IELTS), the *Diplomas de Español como Lengua Extranjera* (Diplomas of Spanish as a Foreign Language, DELE), the Japanese Language-Proficiency Test (JLPT) and the Test of Proficiency in Korean (TOPIK), were examined to gain a better understanding of the format of language tests. In addition, B1 listening-level criteria and Field's (2008) listening framework were taken into consideration. During the examination preparation process, the aim of the examination, its question types and listening materials were clearly determined before conducting the examination. In addition to the CEFR B1-level, Weir's (2005) socio-cognitive framework was also followed to provide texts and questions of an appropriate level. The questions tested the identification of main ideas, deduction, reason-result and important details as stated in the B1 level. Monologues and dialogues were also employed. For the texts, different types, such as television programs, lectures, and phone calls, were preferred as it is stated in CEFR that successful candidates should be able to follow daily conversations, television programs, radio shows, news, and conversations, provided that they do not include technical terms. To decide on the number of texts and questions, several widely used language examinations were taken into consideration. This resulted in providing a minimum of 3 texts

and approximately 30–40 questions. As a result of the number of questions, there were two sections in the examination. The first included the 3 texts and 24 questions. Multiple-choice, matching and short-answer type questions were preferred for the sake of reliability. For short-answer type questions, content words were chosen and the answer key was clearly provided to prevent differences in rating. In all the investigated international examinations except DELE, the audio was played once. However, when this matter was discussed with teachers, they strongly suggested playing the audio twice or even that everyone should be able to listen as many times as they need by asking for a repetition of the audio or using technology to effect this repetition. For these reasons, participants could listen to the audio of the texts twice in this section. There were eight multiple-choice items in the second section. Shorter texts were used, as is the case in the JLPT and TOPIK examinations. The texts were approximately one minute long. Multiple-choice questions with four options were used. Test-takers listened to them once and received instructions and information for every section before they started. The questions generally tested understanding, deduction, and finding reasons or results. After the examinations, the teachers and students were asked to fill out examination evaluation forms and comment on the examination.

Convenience sampling was used for this study. Thirty-one students from two different institutions took the test. To facilitate an effective evaluation, the students chosen were from the B2 level and had just started. The answers of 26 students was analyzed (the remaining students skipped one section, rendering their data unusable). For the evaluation forms, replies from 29 students and 5 teachers were collected. Validation arguments for the examination drew upon Weir's (2005) socio-cognitive framework. Other sources used to assess the validity of the examination were Field's (2008) listening framework and the CEFR B1 level descriptors. Based on this aforementioned validation, it was found that the cognitive and contextual requirements of the test were appropriate for this level. To examine the reliability of the test, item analysis was conducted with the examination results of 26 students. Item difficulty and discrimination revealed that matching questions were not applicable at this level as both low- and high-level groups answered the same questions correctly or incorrectly. In their exam evaluation forms, teachers stated that this type of question could be useful for the A2 level. Moreover, the distractors in the multiple-choice questions were not effective. It was understood that the question and text type for Section 2 was helpful to discriminate students. However, the text types and distractors were not appropriate. As a result, the number of distractors in the shorter texts in the second section should be reduced and matching-type questions can be eliminated to improve reliability.

1. Giriş

Dinleme becerisini gözlemek diğer becerilerle kıyaslandığında oldukça zordur. Konuşma ve yazma becerisi üretici beceriler olduğu için dil öğrencisinin dili yazılı ve sözlü kullanımına bakılarak hedeflenen becerileri edinip edinmediği anlaşılabilir. Dinleme becerisi ise ancak dolaylı bir şekilde yazma, konuşma ve okuma becerilerine dayanarak ölçülebilmektedir. Öğrenci dinlediğini anlamış olsa dahi üretici becerileri yetersizse yanlış değerlendirilebilir. Öğrencinin var olan becerisini gösterebilmesine olanak sağlamak için zihninde gerçekleşen aşamalar anlaşılmaya çalışılmıştır. Zihindeki aşamaların bilinmesi soru ve görevlerin hazırlanırken istenilen hedeflere ulaşmada yardımcı olmaktadır. Dinleme becerisinin aşamalarının anlaşılması, hedeflenen yeterli dil kullanıcısının tanımlanmasına da katkı sağlamıştır.

Alan yazında yer alan çalışmalara göre dinleyici öncelikle “konuşma metnini algılar”. Algılama aşaması seslerin zihinde şekillendirilmesi, duyma anlamına gelmektedir. İkinci aşama ise öğrencinin zihninde gerçekleşen “anlamlandırma” işlemidir. Gözlemlenmesi ve değerlendirilmesi zor olan aşamadır. Burada dinleyici sadece sözcükleri kodlamaz, bağlamdaki anlamlarını takip eder, metin içerisinde gidip gelerek sözcükleri anlamlandırır (Field, 2013:89). Field (2008) dinleme becerisi çerçevesinde konuşma metnini algılamayı düşük seviye, anlamlandırma sürecini ise ileri seviye beceriler olarak ifade etmiştir. İleri seviyedeki süreçlerde dinleyici sözcüklerin tek başına sahip olduklarını anlamların ötesine geçerek boşlukları doldurur, kültürel bilgisini kullanarak anlamlandırır. Gereksiz sözcüğü belirleyerek, kendi anladıklarını takip ederek, konuyla ilgili-ilgisiz bilgileri ayırt ederek ve bilgileri önem sırasına göre dizerek söylem oluştururlar. İşitme ve anlamlandırma aşamaları sırayla değil eş zamanlı olarak gerçekleşmektedir. Buck (2002) dinlemeye devam ederken konuşmanın başında söylenenlerle bağlantı kurulduğunu, sözcüklerin konuşmanın sonraki aşamalarında anlamlandırıldığını ifade etmektedir. Sonuç olarak dinleme tek yönlü değil gidiş-gelişlerle gerçekleşmektedir. Özetle, başlangıç seviyesindeki bir dil öğrencisinden daha çok metni algılama, bir araya getirerek tümceleri oluşturması beklenmektedir. Orta ve ileri düzeydeki becerilerde ise öğrencilerden bunların ötesinde anlamlandırma aşamalarını gerçekleştirebiliyor olmaları hedeflenir. Böyle bir durumda sözcük anlamının ötesine geçildiği için duyduklarını aynı şekilde ifade etmek yerine dinleyici kendi ifadeleriyle aktarabilmektedir.

Öğrencilerin dil yeterlik düzeyleri göz önünde bulundurulduğunda A1-A2 seviyesi çerçevede düşük düzey, B1 ve üzerindeki seviyeler ileri düzey olarak değerlendirilebilir. Araştırmada B1 düzeyindeki öğrencilerin dinleme becerisi hedeflediği için göstermeleri beklenen becerilerin kapsamı detaylandırılırken Field (2008) dinleme çerçevesiyle birlikte Avrupa Dilleri Ortak Çerçeve Metni’ne başvurulmaktadır. AOÇM’ne göre B1 eşik düzeyidir. B1 seviyesindeki bir dil öğrencisi dili artık bağımsız bir şekilde kullanmaya başlamaktadır. Alan bilgisi gerektirmeyen konuları, günlük konuşmaları, kurmacaları, radyo, televizyon programlarını ve haberleri açık ve anlaşılır olması kaydıyla takip edebilmektedir. Konuşmaların konusuna ilişkin çıkarımlarda bulunur, neden-sonuç ilişkisi kurar, günlük hayattaki cihazların kullanım

yönergelerini anlar. Field (2008) çerçevesi ve AOÇM B1 düzeyi beraber değerlendirildiğinde dinleyicinin duyduklarını kodlamanın ötesinde anlamlandırma, metin içerisindeki başka bilgilerle ve kendi bilgileriyle bağlantılar kurarak çıkarımlarda bulunma, ana fikri kavrama gibi beceriler sergilediği görülmektedir. B1 düzeyinde bir sınav hazırlanırken zihinlerinde gerçekleşen aşamaların bilincinde olmak ve bu düzeyde göstermeleri beklenen becerilerini ortaya çıkaracak metinler, sorular ve etkinlikler yer almalıdır.

Dil öğrenim sürecinde uygulanan sınavların genel amacı dil öğrencisinin hedeflenen beceriyi sergileyip sergilemediğini görmektir. Sınavların sonuçları eğitim programını düzenleyip öğrencinin eksik yönlerini tamamlamada, yeterlik düzeyini belirlemede ya da seviye belirlemede kullanılabilir. Ancak sınavın seviyeye uygun hazırlanmaması, öğrencinin becerilerini sergilemesine olanak sunmaması öğrencinin yanlış değerlendirilmesine ve eğitim sürecinin olumsuz etkilenmesine sebep olabilir. Araştırmada kullanılan dinleme sınavı hazırlanırken çerçeveler haricinde sınav geliştirmeye yönelik yaklaşımlar göz önünde bulundurularak sınavın hazırlama ve uygulama aşamaları detaylı olarak çizilmiştir.

Sınav hazırlarken 3 yaklaşımın ön plana çıktığı görülmektedir. Sınav hazırlama süreçleri de tarih boyunca değişiklik gösteren dil öğretim yaklaşımlarından etkilenmiştir. Dil öğretiminde ilk zamanlar dilbilgisi-çeviri yöntemleri kullanılmıştır. Sonrasında dilbilgisinin yanında diğer becerilerin de kullanımı önem kazanmış böylece doğal ve doğrudan dil öğretim yöntemleri kullanılmıştır. Ardından dilin iletişimsel kullanımı ön plana çıkmış ve görev odaklı, iletişimsel yöntemler yaygınlaşmıştır. Günümüzde ise seçmece yöntem yaygın olarak kullanılmaktadır. (Larsen-Freeman, 2000: Richards-Rodgers, 2002) Bu yaklaşımlar sınav hazırlama sürecine de etki etmiş ve parça-odaklı (discrete-point), bütünlükçü (integrative) ve iletişimsel yaklaşımlar ortaya çıkmıştır. Parça-odaklı yaklaşımda dilbilgisi yapıları ve sözcükler metinden bağımsız olarak doğrudan test edilmiştir. Sesletimde kullanılan küçük çift (minimal pair) ayırt etme soruları başlıca örneklerindedir. Türkçe için “kol-yol”, “sar-sor”, “toz-tuz” örnek verilebilir. Çoktan seçmeli sorular ve duyduğu cümlenin eş anlamlısını bulma soruları yaygın olarak kullanılır (Buck, 2002). Diğer bir soru türü ise dinlediği cümleyle aynı anlama gelen seçeneği bulma sorularıdır. Bu tür sorular başlangıç seviyesine uygun değildir, fakat orta ve ileri düzeyde öğrencilere duydukları metnin farklı ifadesini buldurma şeklinde sunulabilir. Bütünlükçü yaklaşımda duyduğunu sadece algılama değil “kullanabilme” becerisinde odaklanılmaktadır. Dinleme becerisi boşluk doldurma, özetlerdeki boşlukları doldurma, anadiline çevirme, dikte etme gibi yöntemlerle test edilmiştir. Boşluk doldurma soruları hazırlanırken metni dinlemeden sadece okuyarak yapılmasının önüne geçilir. Son olarak iletişimsel yaklaşımda ise dili “kullanabilme” becerisinin derinine inerek dili belli bir hedef, belli bir işlev doğrultusunda kullanmak ölçülmüştür. Sorulardaki metinler ve görevler gerçekçi tutulmuştur.

Bu araştırmada B1 düzeyi tanımlayıcıları ve sınav hazırlama yaklaşımları göz önünde bulundurularak, geçerliği ve güvenilirliği yüksek bir sınav hazırlamak için bir çerçeve öne sunulmuştur. Bu çerçeve sınav amacının belirlenmesi, soru türleri ve metinlerin seçilmesi,

sınavın test edilip düzenlenmesi olmak üzere 3 basamaktan oluşmaktadır. İlk aşamada sınavın türü sınavın hangi becerileri ölçeceği ve sonuçlarının nerede kullanılacağı tespit edilir (Hughes, 2003). Sınavlar kullanım amaçlarına göre eğitim öncesi, sırası ve sonrası değerlendirme şeklinde sınıflandırılmaktadır (Güler, 2019:14). Eğitim öncesinde yerleştirme sınavları, eğitim sırasında öğrenim sürecindeki eksikleri belirlemek için belli aralıklarda yapılan quizler, ara sınavlar veya testler, eğitim sonrasında ise hedef doğrultusunda ulaştığı noktayı belirlemek için yeterlik sınavları uygulanır. Amacın belirlenmesi hedef kitleden beklenen becerilerin de net bir şekilde çizilmesidir. Bu çalışmada kullanılan sınavın amacı katılımcıların B1 düzeyindeki yeterliklerini ölçmektir. Seviye tespit sınavı olarak kullanılabilmesi gibi yeterlik sınavı olarak da sunulabilir. Hedef kitle ise Türkçeyi yabancı dil olarak öğrenmekte olan herkesi kapsamaktadır.

Sınav hazırlamanın ikinci aşamasında dinleme metni ve sorular hedefler doğrultusunda seçilir. Dinleme metni seçilirken konuşmacı sayısı, metnin süresi ve hızı, konuşmada gereksiz, boşluk dolduran sözcüklerin bulunması, dilsel zorluk derecesinin seviyeye uygun olması, konusunun ilgi çekici olması ve konuşmada duraksama, tereddüt gibi doğal yapıların bulunmasına dikkat edilmelidir (Rost, 1990:171-172). Dinleme metinleri hazırlanırken veya seçilirken 3 yöntem öne çıkmaktadır. Field (2008) metinleri özgün, yarı-özgün ve yazılı olarak sınıflandırmaktadır. Özgün metinler orijinal kaynaklardan doğrudan alınan metinlerdir. Haber metinleri, hava alanındaki duyurular, radyo programları, internetteki yayımlar örnek gösterilebilir. Yazılı metinler ise dil öğrenimi için hazırlanmış, yazılı bir metinden seslendirilen kaynaklardır. Genellikle tereddüt ifadeleri, duraksamalar yer almaz, konuşma geçişleri doğal değildir. Bu nedenle yaygın olarak yarı-özgün metinlere başvurulur. Yarı-özgün metinlerde orijinal kaynak yeniden seslendirilebilir. Diğer bir seçenek ise konuşmacılara konu-soru vererek doğaçlama yaparak konuşturmadır. Metinlerde konuşmacılara ne kadar doğaçlama fırsatı verilirse metin o kadar özgün olacaktır. Soru türleri belirlenirken katılımcıların hedeflenen becerileri sergileyebilecekleri görevler seçilmelidir. AOÇM B1 düzeyi ve Field (2008) dinleme çerçevesindeki ileri düzey süreçler göz önünde bulundurulduğunda soruların ana fikri, neden-sonuç ilişkisini, metinden yapılacak çıkarımları sorması gerekmektedir. Sınavlarda seçmeli ve üretici olmak üzere iki farklı soru türüyle karşılaşılır. IELTS, TOEFL, JLPT, TOPIK gibi uluslararası geçerliği olan sınavlarda yaygın olarak seçmeli soru türünde yer alan çoktan seçmeli, sıraya koyma, doğru-yanlış, eşleştirme gibi sorular yer almaktadır. Çoktan seçmeli sorularda seçeneklerin uzunluklarının birbirine yakın olmasına, okuma oranının çok olmamasına dikkat edilmelidir. Bu tür sorularda okuma oranının aza indirilmek istenmesinin sebebi soruyu okuyup seçenekleri aklında tutma becerisinin dinlediğini anlama becerisinin önüne geçmesini engellemektir. Ayrıca “yukarıdakilerin hepsi” veya “a ve b” gibi seçeneklere yer verilmemelidir (Brown, 1996). Seçmeli sorular puanlamasının güvenilir olması sebebiyle tercih edilmektedir. JLPT sınavında diğerlerinden farklı olarak dinledikleri kısa metnin ana fikrini bulma, TOPIK sınavında ise konuşmacının söylediklerinden neden-sonuç ilişkisi kurma soruları yer almaktadır. Üretici sorularda ise boşluk doldurma, kısa cevap, sesli ve yazılı özetlerdeki boşlukları doldurma soruları yer almaktadır. Boşluklar içerik (content) sözcüklerinden seçilmesi ve boşluğa gelecek

sözcük sayısının belirtilmesi beceriyi ölçmede daha etkilidir. Puanlama sırasında farklılıklar oluşmasının önüne geçmek için detaylı cevap anahtarının hazırlanması önem taşımaktadır. Sınav öncesinde kabul edilebilecek eş anlamlı sözcükler ve yazım yanlışları belirtilmelidir. Son zamanlarda anlamı bozmadığı sürece yazım yanlışları göz ardı edilmektedir. Son aşamada ise sınavın pilot uygulaması gerçekleştirilir ve bu doğrultuda gerekli düzeltmeler yapılır. Sınavın işleyişi ile ilgili bölüm, metin, soru, metni dinleme sayısı ve süre gibi gerekli açıklamalar yazılı veya sözlü olarak sunulur, sınavı uygulamak için fiziksel şartlar sağlanır. Özellikle dinleme sınavlarında dinlemenin nasıl sunulacağı önem taşımaktadır. Öğrenciler kendi cihazlarından bireysel dinleyebileceği gibi tek bir cihazdan bir sınıf da dinleyebilir. Sınav öncesinde bu şartların göz önünde bulundurulması gerekir. Hazırlanan sınav uygulamaya geçilmeden önce öğretmenler ve uzmanların görüşleri almak veya anadil konuşucuları üzerinde uygulamak olası problemlerin erken fark edilmesini ve önüne geçilmesini sağlayacaktır.

Tablo 1.1: Uluslararası Kullanılmakta Olan Sınav Formatları					
Sınav Adı	Sınav süresi	Soru Sayısı	Metin Sayısı	Dinleme sayısı	Soru türleri
IELTS	30 dk	40	4	1	Çoktan seçmeli Eşleştirme Tablo/resim/özet vs. tamamlama Cümle tamamlama
TOEFL	41-57	28-39	5-7	1	Çoktan seçmeli (4 şıklı)
TestDaf	40 dk	25	3		Kısa Cevap Doğru Yanlış Açık uçlu
DELE	40 dk	30	5	2	Çoktan seçmeli (4 şıklı) Eşleştirme
JLPT	N2 - 50 dk N3 - 40 dk	N2 - 32 N3 - 28		1	Çoktan seçmeli (3-4 şıklı)
TOPIK	Yaklaşık 1 saat	TOPIK-II - 50		1	Çoktan seçmeli

Bu araştırmada kullanılan sınav hazırlanırken bahsedilen 3 aşama takip edilmiştir. Bu süreçte B1 düzeyi beceri çerçeveleri ve alan yazındaki çalışmalar haricinde uluslararası kullanılmakta olan geçerliği ve güvenilirliği ispatlanmış sınavlar da incelenmiştir. Yukarıdaki tabloda incelenen sınavların genel içerikleri sunulmuştur.

Tabloda yer alan sınavlardan IELTS, TOEFL ve TestDaf seviyelere ayrılmadan tek bir sınav olarak sunulmaktadır. Ancak genel yeterliği ölçtükleri için soruları B1 ve üzeri seviyelerdedir. DELE (İspanyolca Yeterlik Sınavı)'de A1, A2, B1, B2 ve C1 olmak üzere her seviye için farklı sınavlar kullanılmaktadır. JLPT (Japonca Yeterlik Sınavı)'de 5 farklı (N1-N5) sınav bulunmaktadır. N1 başlangıç seviyesine N5 ileri düzeye denk gelmektedir. AOÇM benzeven kendi yeterlik kriterleri bulunmaktadır. B1 seviyesi JLPT'de N3-N2 seviyelerine

denk gelmektedir. TOPIK (Korece Yeterlik Sınavı) 2 seviyeden oluşur. TOPIK I başlangıç seviyelerini (1-2), TOPIK II (3-4, 5-6) ileri seviyeleri ifade etmektedir. B1 seviyesinin becerileri göz önünde bulundurularak inceleme için TOPIK II sınavı tercih edilmiştir. Tablodan anlaşıldığı üzere bu dinleme sınavlarında çoğunlukla çoktan seçmeli sorular tercih edilmiştir. Sebep olarak sınavların geniş kitleler tarafından kullanılması gösterilebilir. TOEFL sınavında diğerlerinden farklı olarak bir metinle ilgili farklı türde sorular sorulmuştur. Güvenirliği sağlamak için sınavlarda en az 3 metin kullanıldığı görülmektedir. JLPT ve TOPIK sınavlarında metin başına 1-2 sorulmuştur. Bu nedenle metinler diğer sınavlarınkine göre kısadır. DELE haricindeki sınavlarda metinlerin bir kez dinletildiği görülmektedir. Sınav süresinin ise 30 dakika – 1 saat arasında değişiklik gösterdiği görülmektedir.

Sınav hazırlama aşamalarının son adımında Türkçede uygulanmakta olan sınavlar incelenmiştir. Yabancı dil olarak Türkçe öğretiminde genel geçer tek bir Türkçe Yeterlik sınavının bulunmadığı görülmektedir. Bu konuda Yunus Emre Enstitüsü çalışmalarını yürütmektedir. Hazırladığı sınav 2012 yılından itibaren yurt içi ve yurtdışında kullanılmaktadır. Ancak kullanım alanı genellikle kendi kurumlarıyla sınırlı kalmaktadır. 45 dakikalık, 6 metin ve 30 sorudan oluşan bir sınav olduğu görülmektedir. Diğer dinleme sınavlarının ise genellikle kurum içerisinde hazırlandığı gözlemlenmiştir. Seviyelere göre ayrılmış Türkçe dinleme metni kaynağı sıkıntısı olduğu için sınavda kullanılan metinlerin profesyonel olmayan ortamlarda kaydedildiği görülmektedir. Türkçede sınav geliştirmeye yönelik akademik çalışmaların var olan sınavların değerlendirilmesiyle sınırlı kaldığı görülmektedir. Işıkoğlu (2015) araştırmasında Mersin Üniversitesi İnternet Tabanlı Türkçe Yeterlik Sınavı ile Sakarya Üniversitesi Uzaktan Türkçe Sınavını madde bakımı açısından incelemiştir. Sınavlarda çoktan seçmeli, kısa cevap, doğru-yanlış soruları yer almaktadır. Soru köklerinde seçeneklerle binişlikliğe, zıt ifadelere, dilbilgisi ve yazım hatalarına rastlanmıştır. Seçeneklerde ise çeldiricilerin zayıf olduğu ve seçenekler arası uyumsuzluk görülmüştür. Dinleme sınavına odaklanan sadece bir çalışma bulunmaktadır. Tozlu (2017) çalışmasında A1-B2 seviyelerinde dinleme soruları hazırlamış, geçerliğini ve güvenilirliğini test etmiştir. Sınav sonrası katılımcılara sınavı değerlendirmeleri için sınav değerlendirme anketi sunulmuştur. Sınav ve anket sonuçları değerlendirildiğinde seçeneklerin metinle paralel olmadığı ve doğru-yanlış soruları gibi şans faktörü yüksek olan bazı soru türlerinin seviyeye uygun olmadığı anlaşılmıştır.

Bu nedenlerle de yabancı dil olarak Türkçe öğretiminde kullanılmak üzere geçerliği ve güvenilirliği yüksek bir sınav hazırlanmasının önemi ortaya çıkmaktadır. Bu çalışmada geçerliği ve güvenilirliği yüksek ve uluslararası düzeyde kullanılacak B1 düzeyinde bir dinleme sınavı taslağı oluşturulmak hedeflenmiştir. Sınav hazırlama sürecinde AOÇM'ne ve Field (2008) dinleme becerisi çerçevesine başvurulmuş, uluslararası geçerliği yüksek TOEFL, IELTS, TestDaf, DELE, JLPT ve TOPIK sınavları incelenmiştir. Hazırlama sürecinde ve sınav uygulandıktan sonra aşağıdaki sorulara cevaplar aranmıştır:

1. Sınavdaki metinler ve sorular B1 düzeyine uygun mu?
Metinlerin dili, içeriği, hızı seviyeye uygun mu?
Sorular ve seçenekler B1 düzeyi için anlaşılır özellikte mi?
2. Sınavın bağlam geçerliği sağlanmış mıdır?
Sınavdaki sorular ve beceriler B1 becerilerini ölçer nitelikte mi?
Metinlerin konuları ders konularını ve günlük hayattaki konuşmaları kapsıyor mu?
3. Sınavın puanlama geçerliği sağlanmış mı?
Puan dağılımı ve puanlandırma nasıl yapılmıştır?
4. Sınavın güvenilir midir?
Madde ayırt ediciliği ve madde zorluğu hedeflenen aralıkta mı?

2. Yöntem

Araştırmanın evrenini, İstanbul Üniversitesi Dil Merkezi ve Fatih Sultan Mehmet Vakıf Üniversitesi Türkçe Öğretim Merkezi oluşturmaktadır. Örneklem ise İstanbul Dil Merkezinde B1 düzeyindeki 21 öğrenci ile Fatih Sultan Mehmet Vakıf Üniversitesindeki 10 öğrencidir. Pandemi sebebiyle azalan Türkçe öğrenci sayısı sınavın uygulanacağı B1 öğrenci sayısını da etkilediği için var olan bu sınırlılıklar nedeniyle uygulama ulaşılabılır örneklem seçeneğine başvurulmuştur (Büyüköztürk, 2012; Yıldırım-Şimşek, 2018).

Belirtilen Türkçe öğretim merkezlerindeki öğrenciler akademik eğitimlerini tamamlama amacıyla Türkçe öğrenmektedirler. Hedef kitle B1 yeterliğini alıp B2'ye yeni başlamış öğrencilerden oluşmaktadır. Katılımcılar buldukları kurumlarda A1 düzeyinden başlayarak B2 düzeyine ulaşmışlardır. Sınavın geçerliğini sağlamak amacıyla sınav, B2 seviyesinin ilk haftasındaki öğrencilere uygulanmıştır. Sınavdaki boşluk doldurma bölümünü boş bırakan 5 öğrencinin sınavları incelemeye katılmamış, değerlendirme 26 katılımcı üzerinden yapılmıştır. Sınav hazırlanırken katılımcıların farklı dillerden ve kültürlerden geldiği göz önünde bulundurulmuştur. Katılımcıların hepsi B1 seviyesini başarıyla tamamlamış olmalarına rağmen bireysel farklılıkları, farklı artalan bilgileri ve farklı eğitim durumları sergilemektedir. Ancak bu çalışmada tüm katılımcıların Avrupa Ortak Dil Çerçevesinde belirtilen B1 özelliklerini taşıdığı varsayılmaktadır.

Field (2008) dinleme çerçevesi, AOÇM'de yer alan B1 dinleme becerileri ve incelenmiş olan diğer sınavlar göz önünde bulundurularak 3 uzun, 8 kısa metinden oluşan 32 soruluk bir sınav hazırlanmıştır. Sınav 3 bölümden oluşmaktadır. Metinlerde monolog ve diyaloglara yer verilmiştir. Yabancı dil olarak Türkçe öğretiminde seviyelere göre düzenlemiş metinler bulunmadığı için yazılı metinler hazırlanıp seslendirilmiştir. İlk metinde günlük bir telefon konuşması verilerek metne göre çoktan seçmeli 6 soru cevaplanması istenmiştir. İkinci metinde bir grup oyunuyla ilgili yönlendirme monoloğu dinletilmiş ve katılımcılardan çoktan seçmeli ve kısa cevaplı soruları cevaplanması istenmiştir. Çoktan seçmeli sorularda seçeneklerin ve soru metninin seviyeye uygun olmasına dikkat edilmiştir. Boşluk doldurma kısımlarında ise

cevaplarının net ve puanlamasının güvenilir olması için boşluklar rakamlardan seçilmiştir. Bu bölümdeki sorular oyunun fiyatlandırması ve süresi gibi bilgilerin takibini ölçmektedir. Üçüncü metinde ise izledikleri film ile ilgili karşılıklı yorum yapan iki konuşmacı yer almaktadır. Bu metinle ilgili konuşmacı-fikir eşleştirme soruları kullanılmıştır. Seçenekler metnin kendisinden alınmamış, duyduklarının farklı bir ifadesi olarak sunulmuştur. Birinci bölümdeki ilk 3 metin kullanılmakta olan sınavlar ve uzmanların görüşleri doğrultusunda 2 kez dinletilmiştir. İkinci bölümde ise JLPT ve TOPIK sorularında olduğu gibi kısa bir metinle ilgili tek bir çoktan seçmeli soru sorulmuştur. 8 metin kullanılmış ve metinler bir kez dinletilmiştir. Bu sorular neden-sonuç ilişkisi ve ana fikri buldurmaya yöneliktir. İkinci bölümdeki bir soru, soru kökündeki anlam bozukluğu sebebiyle sınavdan çıkarılmıştır. Sınavın süresi 35 dakika olarak belirlenmiştir. Ancak uygulama sırasında sürenin yetersiz kaldığı anlaşılmış ve 45 dakika olarak değiştirilmiştir ve tüm sınıflarda aynı şekilde uygulanmıştır. Sınav formatı alışık oldukları formattan farklı olduğu için sınav öncesi sınavla ilgili gerekli yönergeler dinletilmiş ve olası problemlere karşı araştırmacı da sınav gözetiminde bulunmuştur. Sınav çevrimiçi olarak Google form üzerinden uygulanmıştır. Katılımcılar ise sınav süresinde Zoom üzerinden takip edilmiştir. Sınav boyunca kameralarını açık tutmaları istenmiştir.

Sınav uygulandıktan sonra sınavı uygulayan ve Türkçe öğretimi deneyimi bulunan öğretmenlere ve sınava katılan öğrencilere sınavla ilgili sınav değerlendirme formu sunulmuştur. Sınav hazırlanırken Gülle'nin (2015) konuşma sınavı için hazırlanmış olduğu değerlendirme formu esas alınmıştır. Bu çalışmada soruların seviyeye uygunluğu, konuşmacıların hızı, yönergelerin anlaşılabilirliği, metinlerin konularının derslerdeki ve günlük hayattaki konulara yakınlığı ve soruların zorluk derecesini 1-4 arasında değerlendirmeleri istenmiştir. Öğrencilere sunulan ankette iki bölümü ayrı ayrı değerlendirmeleri istenmiştir. Öğretmen anketinde ise aynı sorular verilmiş ve sınavı genel olarak değerlendirmeleri istenmiştir. Öğrenci anketinden farklı olarak öğretmenlerden bu sınavda kullanılan farklı soru türlerini de değerlendirmeleri istenmiştir. Anketlerin sonunda sınavla ilgili görüşlerini belirtmeleri için yorum kısmı bırakılmıştır. Bu anketlere 29 öğrenci, 5 öğretmen katılmıştır.

3. Bulgular

Bu bölümde sınav ve anket sonuçları kullanılarak sınavın geçerlik ve güvenilirliği test edilmiştir. Araştırmanın ilk 3 sorusunu cevaplamak için Weir'in (2005) sosyal bilişsel çerçevesi kapsamında bağlam ve puanlama geçerliğine bakılmıştır. Bağlam geçerliği için Field (2008) dinleme becerisi çerçevesi ve AOÇM B1 düzeyi tanımlayıcıları da göz önünde bulundurulmuştur. Bu süreçte öğretmen ve öğrenci anketlerinin sonuçları da değerlendirmeye katılmıştır. İkinci bölümde ise sınavın güvenilirliği üzerinde durulmuştur. Soruların tutarlılığını incelemek için sırasıyla madde güçlük analizi ve madde ayırt edicilik analizi uygulanmıştır. Son bölümde ise öğretmen ve öğrenci anketlerinin sonuçları paylaşılmış ve genel olarak değerlendirilmeye katılmıştır.

3.1. Geçerlik

Sınavın bağlam geçerliği için Weir'ın (2005) çerçevesi doğrultusunda sırasıyla görevlerin yapısı, görevlerin içeriği ve görevlerin uygulanışı incelenmiştir. Görevlerin yapısı için sınavın amacına, cevap formatına, bilinen kritere, soruların ağırlığına, sıralanışına ve sınav süresine bakılmıştır. Hazırlanan sınav katılımcıların B1 düzeyi yeterliklerini ölçmeyi hedeflemektedir. Sonuçları B1 düzeyi tamamlama sınavında kullanılabilmesi gibi yerleştirme sınavı olarak da kullanılabilir. Burada hedeflenen kitle B1 düzeyinde Türkçe öğrenen genel öğrencidir. Puanlaması ve uygulanması geniş kitleye hitap edeceği düşünülerek oluşturulmuştur. Sorular alışık oldukları soru türlerinden başlayıp yeni soru türleri sonda gelecek şekilde sıralanmıştır. B1 düzeyindeki hedef kitleden günlük konuşmaları takip etmeleri, akademik ortamlarda alan bilgisi gerektirmeyecek hızlı olmayan konuşmaları takip etmeleri, radyo ve televizyon programlarını büyük oranda takip edebilmeleri beklenmektedir. Sınavdaki metinler belirlenirken bu ortamlar göz önünde bulundurulmuştur. Bu bölümde metinlerin ve soruların içeriği üzerinde durulmaktadır. Sınavın ilk metninde günlük hayatta sıkça karşılaşılan telefon şirketi müşteri hizmetleriyle müşteri arasında geçen telefon konuşması yer almaktadır. Bu bölümde kampanyayı seçme nedenleri, kampanyaların özellikleri gibi ana fikir, önemli detay ve neden-sonuç ilişkisi kuran sorular yer almıştır. Karşılaştıkları ilk bölümde aşına oldukları soru türü olan çoktan seçmeli kullanılmıştır. Seçenekler birbirine paralel ve kısa tutularak okuma oranı azaltılmıştır. İkinci metinde ise oyunun kurallarını anlatan bir monolog yer almaktadır. Bu bölümde 2 farklı soru türü kullanılmıştır. Metnin ilk bölümü 2 kez dinletilip tamamlandıktan sonra diğer bölüme başlamışlardır. Böylece soruları akılda tutma süresi kısaltılmıştır. İlk bölümde çoktan seçmeli ikinci bölümde boşluk doldurma soruları yer almıştır. Boşluk doldurma kısmında duyduklarını takip ederek oyunların fiyatlarını, sürelerini ve katılımcı sayılarını takip etmeleri istenmiştir. Buradaki sayıları takip etmek için dinlemek günlük hayatta da oldukça karşılaşılan bir durumdur. Bu nedenle sorularda verilen görevlerin günlük hayatı yansıttığı görülmektedir. Ancak bu soru türünün Field'ın (2013) çerçevesindeki düşük seviyelerde yer alan sözcük arama becerisini ölçtüğü için B1 seviyesine uygun olmadığı görülmektedir. Üçüncü metinde ise iki arkadaşın izledikleri film hakkındaki yorumları dinletilmiştir. Verilen eşleştirme sorusunda ise görüşleri kişiyle eşleştirmeleri istenmiştir. Soruda yer alan ifadeler dinlediklerinin farklı ifadesi şeklinde sunulmaktadır. Böylece A1-A2 düzeylerinde olduğu gibi duyduklarını doğrudan seçeneklerde görmeleri yerine anlamlandırma süzgecinden geçirerek B1 düzeyi bir beceri göstermeleri beklenmektedir. Sınavın son bölümünde 8 farklı kısa metin ve metinlerle ilgili tek bir soru yer almıştır. Metinlerdeki diyaloglar kütüphanede kitap işlemleri, televizyonda online eğitimi değerlendirme, otobüste karşılaşılan program, üniversitedeki ilk ders gibi günlük ve akademik ortamlardan seçilmiştir.

Sınavın tamamına bakıldığında konuların B1 çerçevesinde belirtilen konularla paralel olduğu görülmektedir. Öğretmenlerin ve öğrencilerin anket sonuçlarında da konuların derslerde ve günlük hayatta karşılaştıkları konulara paralel olduğu anlaşılmaktadır. Anketlerde sınavlardaki

metinlerle ilgili konuşmacının tonunun takibi zorlaştırdığı yorumu yer almaktadır. Sınav boyunca aynı 2 kişinin metni seslendirmesi, vurgu ve tonlamanın tekdüze olması dinlemeyi zorlaştırdığı belirtilmiştir. Ayrıca öğrencilerin bazıları ilk bölümdeki konuşmaların hızlı olduğunu ifade etmiştir. Metinlerle ilgili diğer bir yorum ise metin sayısının fazla olması, dolayısıyla –ilk bölümde 2 kere dinletildiği göz önünde bulundurulurak- sınav süresinin de fazla olmasıdır. Öğretmen anketlerinde eşleştirme sorusunun B1 seviyesine uygun olmadığı, daha çok A2'de kullanılabilceği yorumu yapılmıştır.

Bağlam geçerliğini ölçmek için son olarak sınavın uygulanışı üzerinde durulmuştur. Pandemi sebebiyle sınav çevrimiçi ortamda uygulanmıştır. Sınav boyunca öğrenciler kameralarını açık tutmuşlardır. Sınav linki öğrencilerle paylaşılmış ve belirtilen sürelerde tamamlamaları istenmiştir. Sınavın süresi 35 dakika olarak düşünülmüş ancak uygulama sırasında tüm sınıflar için 45 dakika şeklinde değiştirilmiştir. Herkes metni kendi cihazından açıp dinlemiştir. Böylece sesin yüksekliği herkese uygun bir şekilde sağlanabilmiştir. Sınav öncesinde sessiz bir ortamda bulunup bulunmadıkları kontrol edilmiştir. Herkesin aynı sürede tamamlayabilmesi için paylaşılan linkteki metni sadece bir kere açıp sınavı tamamlamaları istenmiştir. Ancak sınav sırasında anlayamadıkları yerleri tekrar dinleyen öğrenciler olmuş, süre sıkıntısı yaşanmıştır. Kaydın kendisinde gerekli yerlerde metinler 2 kez dinletilmiş, yönergeler ve soruları cevaplamaları için gereken süreler verilmiştir. Sınavın paralel yürütülmesini sağlamak için araştırmacı da sınıfın öğretmeniyle birlikte sınav gözetiminde bulunmuştur.

Araştırmanın üçüncü sorusunda puanlama geçerliği üzerinde durulmuştur. Sınav sonuçları doğrultusunda öğrencilerin sınavı “başarılı” olarak tamamlamaları farklı kriterlere tabidir. Katılımcıların notları puan sırasına göre dizilerek norm dayanaklı değerlendirme yapılabileceği gibi sınav öncesinde belli bir doğru sayısı hedef konularak ölçüt dayanaklı değerlendirme de yapılabilir (Güler, 2019). Bu çalışmada öğrencilerin başarısına değil, sınavın başarısına odaklanıldığı için sonuçlar herhangi bir değerlendirmeye alınmamıştır. Katılımcılar buldukları kurumlarda B1 yeterliğini almış öğrencilerdir. Sınav öncesinde de sonuçlarının taraflarına sadece doğru-yanlış cevap şekilde iletileceği, herhangi bir şekilde puanlandırılmayacağı konusunda bilgilendirilmişlerdir. Cevap anahtarı öncesinde detaylı olarak hazırlanmış ve puanlama (doğru-yanlış cevap belirleme) bilgisayar ortamında gerçekleştirilmiştir. Herhangi bir hata olmaması amacıyla boşluk doldurma soruları araştırmacı tarafından da kontrol edilmiştir. Sınavın bilgisayar ortamında puanlandırılması olası puanlayıcı farklılıklarının, etkilerinin önüne geçilmesini sağlamıştır. Sonuç olarak, sınavın puanlama geçerliği sağlanmıştır.

3.2. Güvenirlilik

Sınavın başka sınıflarda başka zamanda uygulandığında aynı sonuçlar verip vermediğini, tutarlılığı güvenirliliği ifade etmektedir (Bachman, 1996). Bu çalışmada güvenirliliği ölçmek için sınav aynı anda 2 farklı gruba uygulanmıştır, ancak seçilen sınıflar ulaşılabılır örneklerden seçildiği için geniş kitlelere ulaşma imkânı olmamıştır. Güvenirlilikteki diğer basamak

değerlendirici güvenilirliğidir. Değerlendirici ve değerlendiriciler arası güvenilirlik (Mackey & Gass, 2005) sınav çoktan seçmeli olduğu için ve boşluk doldurma sorularının cevap anahtarı detaylı olarak hazırlandığı için bir problem oluşturmamıştır. Sınavın uygulama sırasındaki güvenilirliğini sağlamak için gerekli tedbirler alınmıştır. Sınav çevrimiçi olarak yürütülmüştür. Herkesin eşit şartlarda sınava girmesini ve sınavın paralel bir şekilde uygulanmasını sağlamak, karşılaşılabilecek herhangi bir teknik aksaklığın önüne geçmek için sınıfın öğretmeniyle beraber araştırmacı da sınav gözetiminde bulunmuştur.

Sınavın güvenilirliğini sınamadaki diğer basamak ise sınavın kendi yapısından kaynaklanan hatalardan dolayı katılımcıları yanlış değerlendirip değerlendirilmediğini incelemektir. Bunun için test-tekrar test, paralel (eşdeğer) formlar ve eşdeğer yarılar yöntemi kullanılmaktadır (Mackey & Gass, 2005). Test-tekrar test, aynı sınavın farklı zamanlarda aynı katılımcılara uygulanmasıdır. Güvenirliği yüksek bir sınavda iki sınavda da benzer sonuçlar çıkması beklenir. Eşdeğer formlarda ise benzer iki test aynı sınıfa uygulanır. 2 testteki soruların birbirinden kolay veya zor olmamasına dikkat edilmelidir. Eşdeğer formlarda ise sınav sorulara ikiye bölünür ve sonuçları birbirleriyle karşılaştırılır. Bu çalışmada sınav sadece bir kez uygulandığı için tek bir sınav üzerinden güvenilirliği ölçülmüştür. Bunun için madde analizlerine başvurulmuştur. Sırasıyla soruların güçlük derecesine ve ayırt ediciliğine bakılmıştır.

Madde güçlük indeksi (P) soruların doğru cevaplanma oranını ifade etmektedir. P değeri 0-1 arasında yer almaktadır. Güçlük indeksi 0'a yakın ise madde zor, 1 ise kolay, 0,5 ise orta zorlukta değerlendirilmektedir (Tekin, 2000). Bir maddenin olması gereken ideal aralık 0,4-0,6 arasındadır, ancak uygulama sırasında çoğunlukla 0,3-0,7 aralığının kabul edildiği görülmektedir (Heaton, 1988:179; Henning, 1987 akt. Fulcher & Davidson, 2007).

Aşağıdaki tabloda birinci metinle ilgili çoktan seçme soruları S1-S6, ikinci metindeki sorular S7-S16 şeklinde ifade edilmiştir. İkinci metindeki boşluk doldurma soruları S12'den başlamaktadır. Üçüncü metinle ilgili eşleştirme soruları S17-S24, ikinci bölüm ise S25-S32 arasındadır. İlk metindeki 2., 3. ve 5. Sorunun hedeflenen P değerleri aralığında olmadığı görülmektedir. 3. soruya bakıldığında sadece 7 öğrencinin doğru cevapladığı, sorunun seviyenin üzerinde olduğu görülmektedir. Ayrıca bu soruda d seçeneğini hiçbir katılımcı işaretlememiştir. 2. soru P değeri yüksek olan seviyeye göre kolay bir sorudur. Benzer şekilde seçeneklerden birini hiçbir katılımcı işaretlememiştir. Bu durum seçeneklerin çeldiriciliğinin düşük olabileceğini göstermektedir. İkinci metindeki çoktan seçmeli soruların P değerleri 11. soru haricinde istenilen aralıktadır. Bu soruların P değerlerine bakıldığında bu seviye için kolay oldukları görülmektedir. Boşluk doldurma sorularında ise sadece 12. soru seviye için kolay olduğu görülmektedir. İkinci metindeki soruların P değeri diğer bölümlere kıyasla en yüksek olan bölümdür. Üçüncü metindeki eşleştirme sorularından sadece 2 soru istenilen aralığa girmiştir. Diğer soruların belirlenen seviye için kolay olduğu anlaşılmaktadır. İkinci bölümde soru kökündeki hatadan dolayı 26. soru çıkarıldığı için 7 soru yer almaktadır. 27 ve 31. soruların P değerinin düşük olduğu görülmektedir. Bu soruları doğru cevaplayan sayısı

sırasıyla 5 ve 7'dir. Bu bölümde kısa bir metin sunularak ana fikri bulma, neden-sonuç ilişkisi kurma görevleri verilmiştir. Metinler kısa olduğu için karışık ifadeler olmamasına dikkat edilmesi zorluk derecesini dengelemeye yardımcı olabilir. Ayrıca JLPT sınavlarında olduğu gibi bölümden hemen önce örnek bir soru çözümlenerek katılımcılar yeni soru türüne hazırlanabilir. Tablodan hareketle ikinci bölüm sorularının zorluk derecesinin ilk bölüme oranla seviyeye daha uygun olduğu görülmektedir. Sınavda P değeri hedeflenen aralığa girmeyen en çok soru üçüncü bölümden çıkmıştır. Field'ın (2008) çerçevesinde de belirttiği gibi eşleştirme sorusunun düşük seviye beceriler için daha uygun olduğu görülmektedir.

Bölüm1	Cevaplayan	Doğru cevaplayan	P değeri	D değeri
S1	26	17	0.653	0.428
S2	26	22	0.846	0.142
S3	26	7	0.269	0.285
S4	26	10	0.384	0.285
S5	26	19	0.730	0.142
S6	26	9	0.346	0.571
S7	26	18	0.692	0.714
S8	26	8	0.307	0.428
S9	26	10	0.384	0.714
S10	26	9	0.346	0
S11	26	23	0.884	0.142
S12	26	20	0.769	0.571
S13	26	13	0.5	0.857
S14	26	12	0.461	0.714
S15	26	14	0.538	0.571
S16	26	17	0.653	1
S17	26	14	0.538	0.285
S18	26	20	0.769	0.428
S19	26	14	0.538	0.571
S20	26	23	0.884	0.142
S21	26	22	0.846	0.285
S22	26	21	0.807	0.428
S23	26	24	0.923	0.142
S24	26	21	0.807	0.571
S25	26	18	0.692	0.428
S27	26	5	0.193	0.142
S28	26	15	0.576	1

S29	26	15	0.576	0.428
S30	26	9	0.346	0.571
S31	26	7	0.269	0.142
S32	26	11	0.423	0.285

İkinci aşamada soruların başarılı ve başarısız öğrencileri ayırt etme özelliği incelenmiştir. Bunun için öncelikle öğrenciler başarı sırasına göre dizilmiştir. Öğrencilerin doğru sayısı 9-27 arasında dağılmaktadır. Üst %27'lik grupta alt %27'lik grubun her bir soruyu doğru yapma oranlarına bakılmıştır. Üst ve alt grupta 7'şer kişi yer almaktadır. Üst grup ortalama 25,85 soruya doğru cevap verirken alt grubun ortalama 12 doğrusu vardır. Ortalamalara bakıldığında iki grup arasında anlamlı bir fark görülmektedir. Madde ayırt ediciliği (D) +1 ve -1 arasında dağılmaktadır. D'nin -1 olması soruyu başarılı öğrencilerin yapamadığı başarısız öğrencilerin yaptığı anlamına gelmektedir. Genelde madde ayırt ediciliğinin 0.3 üzerinde olması istenmektedir.

Birinci metindeki soruların D değeri 0,42-0,57 çıkmıştır. Bu durumda bu soruların ayırt edici olduğu anlaşılmaktadır. İkinci metinde ise 10 ve 11. sorular haricinde ayırt edicilik sağlanmıştır. 10. sorunun D değeri 0 çıkmıştır. Bu sonuç bu sorunun iki grup için de ya zor ya da kolay olduğunu ifade etmektedir. P değeri (0,346) ile birlikte incelendiğinde sorunun iki grup için de zor olduğunu, iki grubu birbirinden ayırt edemediği anlaşılmaktadır. 11. soruda ise P ve D değeri birlikte incelendiğinde sorunun iki grup için de kolay olduğu anlaşılmaktadır. Üçüncü metinde ise soruların yarısının (17, 20,21,23) D değeri hedefin altında kalmıştır. Soruların ayırt ediciliği düşüktür. İkinci bölümde 27, 31 ve 32. soruların D değeri 0,3'ün altındadır. P değerleriyle birlikte ele alındığında bu soruların iki grup için de zorlayıcı olduğu görülmektedir. Tüm sınav madde ayırt ediciliği yönünden ele alındığında birinci metinden 4, ikinci metinden 2, üçüncü metinden 4 ve ikinci bölümden 3 olmak üzere toplamda 13 sorunun başarılı ve başarısız öğrenciyi ayırt edemediği görülmektedir.

Sınavın madde ayırt ediciliği ve madde güçlüğü birlikte değerlendirildiğinde birinci metnin diğer bölümlere göre daha başarılı olduğu görülmektedir. İkinci metindeki 10 ve 11. soruların madde güçlüğü ve ayırt ediciliği değerlerinin düşük olduğu görülmektedir. Sorularda çeldiriciliği düşük seçenekler bulunduğu, soruların bu seviye için kolay olduğu, üst ve alt grubun kolayca cevaplayabilmesi sebebiyle ayırt ediciliği düşük olduğu yorumu yapılabilir. İkinci metindeki boşluk doldurma soruları ayırt edicilik ve güçlük açısından seviyeye uygundur. Üçüncü metinde ise 20, 21 ve 22.'nci soruların madde güçlüğü ve ayırt ediciliğinin düşük olduğu görülmektedir. Soruların B1 seviyesinin altında kaldığı görülmektedir. Sebep olarak iki seçeneikli, şans faktörünün yüksek olduğu bir soru türü olması gösterilebilir. Son bölümün madde güçlük ve ayırt edicilik indeksleri incelendiğinde 27. ve 31. sorunun değerlerinin düşük olduğu görülmektedir. Bu durumun sebebi öncesinde de değinildiği gibi metnin içeriği olabilir. Ayrıca metinler kısa olduğu için çeldiriciler 4 yerine 3'e indirilebilir. Böylece birbirine yakın seçeneklerle katılımcıların asıl becerilerini gösterememesinin önüne geçilebilir.

3.3. Sınav Değerlendirme Formları

Bu bölümde sırasıyla öğretmenlere ve öğrencilere uygulanan anketlerin sonuçları incelenmektedir. Anketler anonim olarak doldurulmuş, hiçbir katılımcıdan ismini vermesi istenmemiştir. Öğretmenlere uygulanan ankette bölümlere ayrılmadan tüm sınavı değerlendirmeleri istenmiştir. Formu 5 öğretmen doldürmüştür. Tablo 2.2.'ye bakıldığında soruların seviyesinin öğrencilere uygun olduğu ancak metinlerin seviyesinin uygun olmadığı görüşü ortaya çıkmaktadır. Farklı soru türlerinin kullanılması da yaygın olarak desteklenmiştir. Metinlerin konuları hakkındaki görüşlere bakıldığında %40 oranında katılımcının konuların derslerdeki ve sınavlardaki konulardan biraz farklı olduğunu düşündükleri anlaşılmaktadır. Bunlarla birlikte birinci bölümdeki metinlerin 2 kez dinletilmesinin onaylandığı (%100) görülmektedir. Tek metin – tek soru türünü kullanmaya %60 oranında öğretmen olumlu yaklaşmıştır.

Tablo 2.2: Sınav Değerlendirme Anketi Sonuçları (Öğretmenler)				
Anket soruları	1 (%)	2 (%)	3 (%)	4 (%)
1. Soruların seviyesi öğrencilere uygun.			80	20
2. Sınav süresi yeterli.			60	40
3. Sınavdaki metinler derslerde ve sınavlarda karşılaştıkları konularla paralel.		40	40	20
4. Sınavdaki dinleme metinlerinin seviyesi öğrencilere uygun.		40	40	20
5. Sınavda farklı soru türleri kullanılması faydalı.			60	40
6. Bölüm 1'de metinlerin 2 kere dinletilmesi uygun.				100
7. Bölüm 2'deki tek metin tek soru türünü sınavlarda görmek isterim.		40	40	20

Katılıyorum 1 – 2 – 3 – 4 Katılıyorum

Ankette sorular haricinde sınavla ilgili görüşlerini belirttikleri bölümde metinlerin seviyeye uygun olmamasının sebepleri üzerine görüş bildirilmiştir. Metinlerdeki konuşmacıların vurgu ve tonlamalarının zayıf olması ve tüm metinlerde aynı kişilerin bulunmasının metni takip etmeyi ve anlamayı güçleştirebileceği düşünülmüştür. İlk bölümde yer alan boşluk doldurma sorusunun B1 seviyesine uygun olmadığı, daha çok A2 seviyesinde kullanılması uygun bulunmuştur. Sınavda farklı türde sorular kullanılmasını, metinlerin uzunluğunu ve çeşitliliğini olumlu olarak değerlendirmişlerdir. Öğretmenlerin anket sonuçları ve sınavla ilgili yorumları değerlendirildiğinde sınavı formatıyla ilgili bir problemden ziyade metinlerin yapısı ve soru türleri üzerinde eksiklikler olduğu anlaşılmaktadır. Sınavın uzunluğu ile ilgili olan yorumlar konuşmacıların tekdüze konuşmasından kaynaklanabilir. Profesyonel ortamlarda profesyoneller tarafından hazırlanacak kayıtlarla bu sorunun önüne geçilebilir. Eşleştirme bölümünün çıkarılması ile sınavın uzunluğu ve soru türleri daha etkili hale getirilebilir.

Öğrenci anketlerinde her bölümü ayrı olarak değerlendirmeleri istenmiştir. Sınava katılan tüm öğrencilerden anketi doldurmaları istenmiştir. Değerlendirmede 29 öğrencinin anket sonucu yer almaktadır. Sonuçlara bakıldığında her iki bölümde de, konuların dersteki ve günlük hayattaki konulara benzediği görülmektedir. Birinci bölüm için öğrencilerin %13,8'i sınav süresinin yeterli olmadığını belirtmiştir. Bu bölümün zorluk derecesi normal-zor olarak değerlendirilmiştir. İkinci bölümün diğerine kıyasla yönlendirmelerinin yeterince anlaşılmadığı ifade edilmiştir. Konuşmaların hızı iki bölümde de benzer şekilde normal bulunmuştur. İkinci bölümde ise yönlendirmelerin anlaşılmadığı (%13,4) görülmektedir. Sınavın zorluk derecesine bakıldığında ise ilk bölümü normal-zor, ikinci bölümü çoğunlukla zor olarak değerlendirmişlerdir.

Tablo 2.3: Sınav Değerlendirme Anketi Sonuçları (Öğrenciler)									
Bölüm 1: Anket soruları	1 N %		2 N %		3 N %		4 N %		
Yönlendirmeler anlaşılır.	1	3,4	5	17,2	13	44,8	10	34,5	
Sınav süresi yeterli.	4	13,8	3	10,3	10	34,5	12	41,4	
Günlük yaşamdaki konulara benziyor.	2	6,9	3	10,3	10	34,5	14	48,3	
Dersteki konulara benziyor.	1	6,9	6	20,7	10	34,5	11	37,9	
Konuşmaların hızı uygun.	1	3,4	8	27,6	11	37,9	9	31	
Sınavın zorluk derecesi*	1	3,4	13	44,8	12	41,4	3	10,3	
Bölüm 2: Anket soruları	1 N %		2 N %		3 N %		4 N %		
Yönlendirmeler anlaşılır.	4	13,4	6	20,7	11	37,9	8	27,6	
Sınav süresi yeterli.	3	10,3	8	27,6	10	34,5	8	27,6	
Günlük yaşamdaki konulara benziyor.	1	3,4	5	17,2	14	48,3	9	31	
Dersteki konulara benziyor.	0	0	7	24,1	14	48,3	8	27,6	
Konuşmaların hızı uygun.	3	10,3	7	24,1	14	48,3	5	7,12	
Sınavın zorluk derecesi*	2	6,9	9	31	17	58,6	1	3,4	

Katılmıyorum 1 – 2 – 3 – 4 Katılıyorum

Anket sonunda görüşlerini belirttikleri notlarda genel olarak sınavı iyi, faydalı şekilde değerlendirmişlerdir. Öğretmenlerle benzer şekilde ikinci bölümdeki metinleri 1 kez dinlemenin yetersiz olduğunu belirtmişlerdir. Ayrıca sınavın genel olarak uzun olduğunu belirtmişlerdir.

4. Sonuç ve Tartışma

Bu çalışmada uluslararası geçerliği olan ve güvenilirliği ispatlanmış B1 dinleme sınavı hazırlanmıştır. Hazırlama aşamasında alan yazından ve uluslararası geçerliği olan sınavlardan faydalanılmıştır. Sınav hazırlama aşamasında Weir (2005) sosyal bilişsel çerçevesi, Field (2008) dinleme becerisi çerçevesi ve Avrupa Dilleri Ortak Çerçeve'deki B1 düzeyi esas alınmıştır. Ayrıca diğer dillerdeki sınavlar incelendiğinde en az 3 metin kullanıldığı, çoğunlukla çoktan seçmeli soruların tercih edildiği, soru sayısının 25 ve üzerinde olduğu, metinlerin genellikle

1 kez dinletildiği not edilmiştir. Alışılmış sınav formatlarından farklı olarak JLPT ve TOPIK sınavlarında kısa bir metinle ilgili tek bir soru sorulmuştur. Bu sorular genellikle ana fikri, neden-sonuç ilişkisini ve çıkarımları sormuştur. Bunlar B1 kapsamında gösterilmesi beklenen beceriler olduğu için bu soru türleri de sınavda kullanılmıştır. Tüm bunların sonucunda 2 bölümden oluşan 32 soruluk bir sınav hazırlanmıştır. İlk bölümde 3 metin sunulmuş ve 2 kez dinletilmiştir. Metinler belirlenirken konularının günlük hayatta ve akademik ortamlarda karşılaşılan diyaloglardan ve monologlardan olmasına dikkat edilmiştir. Metinlerin uzunluğu ve metin sayısı belirlenirken genel geçer sınavlara ve kurumlarda kullanılan sınavlara bakılmıştır. Metinlerle ilgili çoktan seçmeli, boşluk doldurma ve eşleştirme soruları sorulmuştur. İkinci bölümde ise 8 kısa metin üzerinden birer soru sorulmuştur. Bu bölümdeki sorular metnin ana fikri, nerede geçtiği, neden-sonuç ilişkisi gibi ileri düzey becerileri ölçmektedir. Sınav toplamda 45 dakika sürecek şekilde uygulanmış ve 31 kişiye uygulanmıştır. Ancak 5 kişi boşluk doldurma sorusunu tamamlamadığı için madde analizi değerlendirmelerinden çıkarılmışlardır. Son olarak, sınav sonrası katılımcı öğretmen ve öğrencilere sınav değerlendirme formu doldurtulmuştur.

Araştırmanın birinci ve ikinci sorusunda sınavın seviyeye uygunluğu ve bağlam geçerliği incelenmiştir. Dinleme metinlerinin konuları ders seçimi, röportaj, haber metni, bilgilendirici metin ve tartışma programı gibi günlük hayatta karşılaştıkları konulardan seçilmiştir. İlgili çerçevelere bakıldığında metinlerin B1 becerisini ölçmede başarılı olduğu görülmektedir. Kullanılan sorular da çerçevede belirtildiği üzere çıkarım yapma, detayları ve ana fikri takip etme, neden-sonuç ilişkilerini kurma şeklinde hazırlanmıştır. Ancak sınavın üçüncü metninde yer alan eşleştirme sorularının Field (2008) çerçevesi, AOÇM, madde analizleri ve öğretmen görüşleri doğrultusunda belirlenen seviyenin altında olduğu görülmüştür. Sınavı uygulayan öğretmenlerin ve katılımcıların değerlendirme formlarında konuların günlük hayatla paralel olduğunu ifade etmişlerdir. Sonuç olarak, sınavın bağlam geçerliğinin sağlandığı düşünülmektedir.

Araştırmanın 2 ve 3. sorularında ise puanlama geçerliği ve güvenilirlik üzerinde durulmuştur. Öğrenciler sınav sonunda sadece doğru-yanlış cevaplarını görebilecekleri, puan almayacakları konusunda bilgilendirilmişlerdir. Çoktan seçmeli ve boşluk doldurma sorularının sayılardan seçilmesi sebebiyle puanlayıcı etkisinin önüne geçilmiştir. Sınavın güvenilirliğini ve eşit bir biçimde uygulanmasını sağlamak amacıyla araştırmacı da sınıfın öğretmenleriyle beraber sınavı gözetlemiştir.

Soruların güvenilirliğini incelemek için madde analizine başvurulmuştur. Madde gücüne bakıldığında 3, 27 ve 31'inci soruların P değeri sırasıyla 0.269, 0.192 ve 0.269 çıkmıştır. Soruların seviyenin üzerinde olduğu anlaşılmıştır. Özellikle ikinci bölümde zorlayıcı sorular görülmektedir. Ankette öğrencilerin bu bölümdeki metinleri bir kez dinletilmesinin yetersiz olduğunu belirtmişlerdir. 27'nci soruda doğru cevabı 5 (19,2) kişi seçerken d seçeneğini 16 (61,5) kişi seçmiştir. Çeldiricilerin soruyu doğru cevaplamanın önüne geçtiği yorumu yapılabilir. 31'nci soruda ise doğru cevap olan d seçeneğini 8 (%30,8), çeldirici olan d seçeneğini ise 11 (%42,3) katılımcı işaretlemiştir. 2, 5, 11, 12, 18, 20, 21, 22, 23 ve 24'üncü soruların P değerleri

0.730-0.923 aralığındadır. Soruların seviyenin altında olduğu anlaşılmıştır. 20, 21, 22, 23 ve 24'üncü sorular birinci bölümdeki üçüncü metinle ilgili kişi-yorum eşleştirme sorularıdır. Bu metin ve sorular sınavdan çıkarılarak sınavın güvenilirliği artırılabilir. Anketlerde öğrenciler sınavın zorluk derecelerini ağırlıklı normal ve zor olarak belirtmiştir. İyi bir sınav için orta güçlük düzeyinde olması önemlidir.

Madde ayırt ediciliğine bakıldığında birinci bölümdeki ilk metnin başarılı ve başarısız öğrencileri ayırt etmede yetersiz kaldığı görülmektedir. İkinci metinde 10 sorudan 2'si 0.3 değerinin altındadır. Diğer bölümlerle kıyaslandığında madde ayırt ediciliği en yüksek bu metin çıkmıştır. Bu durum soruların B1 seviyesine uygun olduğunu ve öğrencileri ayırt etmede güvenilir olduğunu göstermektedir. Üçüncü metindeki eşleştirme sorularının yarısı hedeflenen değerin altındadır. İkinci bölümdeki soruların ayırt ediciliği yüksektir.

Tüm sorular madde güçlüğü ve madde ayırt ediciliği değerleriyle değerlendirildiğinde birinci bölüm ikinci metindeki boşluk doldurma soruları istenilen aralıktadır. Sonrasında birinci ve ikinci metinler gelmektedir. Bu bölümlerde çeldiriciliği düşük seçenekler yer almaktadır. Seçeneklerin çeldiriciliği artırılarak güvenilirliği ve geçerliliği sağlanabilir. P ve D değerinin en düşük üçüncü metnin soruları olduğu görülmektedir.

Sonuç olarak B1 düzeyinde kullanılacak farklı türde metin ve sorular içeren, Avrupa Dilleri Ortak Çerçevesine uygun uluslararası geçerliliği olan bir sınav hazırlanmıştır. Sınavda farklı soru türlerinin kullanılması genel olarak olumlu sonuç vermiştir. İkinci metindeki boşluk doldurma sorusu öğretmenlerin yorumlarına göre B1 seviyesinin altında görünse de madde analizi sonuçlarında seviyeye uygun zorlukta ve ayırt edici oldukları anlaşılmıştır. Sınavın konuları incelendiğinde B1 seviyesine ve sınavın amacına uygun olduğu görülmektedir. Ancak öğretmen anketleri incelendiğinde konularının dersteki ve sınavlardaki konulardan biraz farklı olduğu anlaşılmaktadır. Anketlerde metinlerle ilgili hızlı oldukları ve takiplerinin zor olduğu yorumu gelmiştir. Sebep sınav boyunca aynı konuşmacıların yer alması, konuşmadaki vurgu ve tonlamaların yetersiz kalması gösterilmiştir. Metinlerin profesyonel olmayan kişilerce hazırlanmasının sınavı olumsuz etkilediği anlaşılmıştır. Madde ayırt ediciliği ve güçlüğü sonuçlarına bakıldığında 3. Metindeki eşleştirme sorularının seviyenin altında olduğu anlaşılmıştır. Bu bölümdeki sorular sınavdan çıkarılarak veya değiştirilerek sınavın güvenilirliği artırılabilir. B1 becerisini ölçmede başarısız kalan diğer sorularda çeldiricilerin zayıf olduğu anlaşılmıştır. Seçenekler düzenlenerek sınavın geçerliliği artırılabilir.

5. Öneriler

Son olarak sınavın geliştirilmesi, geçerlik ve güvenilirliğini artırılması için bazı öneriler sunulmuştur. Sınav tek aşamalı olarak 2 farklı grup üzerinde uygulanmıştır. Daha geniş kitleye ulaşarak ve aynı gruplara benzer soruları farklı zamanlarda sunarak sınavın geçerliliği ve güvenilirliği ile ilgili daha detaylı sonuçlara ulaşılabilir.

Sınavda kullanılan metinler profesyoneller tarafından uygun bir ortamda hazırlanmadığı için sınavın takibi zor olmuştur. Anket sonuçlarında da metinlerde vurgu ve tonlamaların anlamayı zorlaştırdığı belirtilmiştir. Bu nedenlerle metinlerin uzmanlar tarafından hazırlanması ya da seviyeye uygun özgün kaynaklar kullanılması sınavın başarısını artıracaktır. Metinlerin hem takibinin zor olması hem de 2 kez dinletilmesi sebebiyle sınavın uzun geldiği görülmüştür. Sınavın güvenilirliğini zedelemeyecek şekilde bir metin çıkartılıp diğer metin ve sorular ona göre uzatılabilir. Anket sonuçlarında metinlerin 2 kez dinletilmesinin faydalı olduğu anlaşılmaktadır. Bu nedenle sınav uzunluğunu dengelemek için metinlerin 1 kez dinletilmesi sınavın verimini düşürecektir.

Üçüncü metindeki eşleştirme sorularının sınavın güvenilirliği ve düzeye uygunluğunu düşürdüğü görülmektedir. Öncelikle bu bölüm sınavdan çıkarılarak güvenilirlik artırılabilir. Diğer bir seçenek ise konuşmacı sayısını artırmak ve eşleştirecekleri yorumları konuşmaların özeti şeklinde sunmak olabilir.

İkinci bölümdeki tek metin – tek soru türünden istenilen sonuç elde edilememiştir. Soru metinleri seçilirken neden-sonuç ilişkisi kurmaya, ana fikri bulmaya olanak sağlamasına, açık ve anlaşılır olmalarına dikkat edilmelidir. Ayrıca kısa metinlerden 4 seçenekli çoktan seçmeli sorular sormak yerine seçenek sayısı azaltılarak sınavdaki okuma oranı azaltılabilir, dikkat metnin ana fikrine daha çok ayrılabilir. Ayrıca sınavda seçeneklerin birbirine yakın olması sebebiyle de ayırt ediciliğinin düşük olduğu görülmüştür. Seçenek sayısının azaltılması bu problemlerin de önüne geçebilir. Son olarak, bu bölümdeki soru tipi öğrenciler için yeni olduğu için ve ilk bölümden farklı olarak metinler bir kez dinletildiği için bölüm başında yönlendirme sonrası örnek soru çözümü sunmak faydalı olabilir.

Hakem Değerlendirmesi: Dış bağımsız.

Çıkar Çatışması: Yazarlar çıkar çatışması bildirmemiştir.

Finansal Destek: Yazarlar bu çalışma için finansal destek almadığını beyan etmiştir.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

Kaynakça/References

- Bachman, L. F., & Palmer, A.S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford University Press.
- Brown, D. J. (1996). *Testing in Language Programs*.
- Buck, G. (2002). *Assessing Listening* (J.Charles Alderson & Lyle F. Bachman, Ed.). Cambridge University Press.
- Büyüköztürk, Ş. (2012). *Örnekleme Yöntemleri* [PDF belgesi]. <http://w3.balikesir.edu.tr/~msackes/wp/wp-content/uploads/2012/03/BAY-Final-Konulari.pdf> adresinden edinilmiştir.

- Field, J. (2013). Cognitive Validity. Ardeshir G. & Lynda T., (Eds.), *Examining Listening* (pp. 77-151).
- Field, J. (2009). *Listening in the Language Classroom*. Cambridge University Press.
- Fulcher, G., Davidson, F. (2007). *Language Testing and Assessment: An Advanced Resource Book*. London & New York.
- Güler, N. (2019). *Eğitimde Ölçme ve Değerlendirme* (14. baskı.). Pegem Akademi.
- Gülle, T. (2015). *Development of a Speaking Test for Second Language Learners of Turkish (İkinci Dil olarak Türkçe Konuşma Testi Geliştirilmesi)* (Yayımlanmamış yüksek lisans tezi). Boğaziçi Üniversitesi.
- Heaton, J.B. (1990). *Writing English Language Tests*. Longman.
- Hughes, A. (2003). *Testing for Language Teachers* (2nd ed.). Cambridge University Press.
- Işıkoğlu, M. (2015). *Yabancı Dil Olarak Türkçe Öğretiminde Kullanılan Yeterlik Sınavlarının Madde Yazımı Bakımından İncelenmesi (Mersin ve Sakarya Üniversiteleri Örneği)* (Yayımlanmamış yüksek lisans tezi). Atatürk Üniversitesi.
- Larsen-Freeman, D. (2000). *Techniques and Principles in Language Teaching*. Oxford University Press.
- Mackey, A., Gass, S. M. (2005). *Second Language Research: Methodology and Design*. Lawrence Erlbaum Associates Publishers.
- Richard, Jack C. (1983). Listening Comprehension: Approach, design, procedure. *TESOL Quarterly*, 8 (2), 219-240. <https://doi.org/10.2307/3586651>
- Richards, J., Rodgers, T.S. (2002). *Approaches and Methods in Language Teaching*. Cambridge University Press.
- Rost, M. (1990). *Listening in Language Learning*.
- Tekin, H. (2000). *Eğitimde Ölçme ve Değerlendirme*. Yargı Yayınları.
- Tozlu, E. (2017). *The Development of a Listening Test for Learners of Turkish as a Foreign Language* (Yayımlanmamış yüksek lisans tezi), Boğaziçi Üniversitesi.
- Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*.
- Yıldırım, A. , Şimşek, H. (2018). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri* (6. baskı). Seçkin Yayıncılık.

Online kaynaklar/Online references

DELE: https://examenes.cervantes.es/sites/default/files/dele_b1_modelo0.pdf

IELTS: <https://www.ielts.org/about-the-test/sample-test-questions>

JLPT: <https://www.jlpt.jp/e/samples/forlearners.html>

TestDaf: <https://www.testdaf.de/de/teilnehmende/der-papierbasierte-testdaf/vorbereitung-auf-den-papierbasierten-testdaf/modelltests/#c2729>

TOPIK seviyeleri: <http://www.topik.go.kr/usr/cmm/subLocation.do?menuSeq=2210101#none>

TOPIK guide: <https://www.topikguide.com/topik-overview/>

TYS örneği: <https://tys.yee.org.tr/images/ornek-sinav/ornek-dinleme.pdf>

