

İkinci/Yabancı Dil Olarak Türkçe Öğretiminde A1-A2 Düzeylerinde Konuşma Puanlama Anahtarı Önerisi

Günsu Taşköprü*

MAKALE BİLGİSİ

Geliş: 30.07.2024
Düzeltilme: 24.12.2024
Kabul: 08.01.2025
DOI: 10.31464/jlere.1524713

Anahtar Sözcükler:

yabancı dil olarak Türkçe
konuşma
dereceli puanlama anahtarı
ölçme ve değerlendirme

ÖZET

Yabancı dil öğretimi sürecinin bir parçası olarak öğrenci yeterliklerini gözlemek için hazırlanan sınavlarda konuşma becerisinin değerlendirilmesi ve puanlaması dereceli puanlama anahtarları, kontrol listeleri gibi araçlarla yapılmaktadır. Uygun puanlama araçlarının kullanılması, öğretim sürecinin ve öğrenci yeterliğinin doğru biçimde saptanması açısından önemlidir. Bu çalışmada değerlendiriciler arasında tutarlılığı gözetmeyi hedefleyerek yabancı dil olarak Türkçe öğrenen A1 ve A2 düzeylerindeki öğrencilerin konuşma performanslarını değerlendirme amacıyla puanlama anahtarları geliştirilmiştir. Keşfedici ardışık desenli karma yöntem kullanılarak hazırlanan puanlama anahtarlarının 10 kişilik uzman grubu ile Lawshe tekniği kullanılarak kapsam geçerlik çalışmaları yapılmış, A1 ve A2 düzeyleri için yüksek kapsam geçerlik indeksine ulaşılmıştır. 20 kişilik öğretici grubu güvenilirlik testleri için altışar A1 ve A2 öğrencisinin sözlü performanslarını puanlama anahtarlarıyla değerlendirmiştir. Kendall-W testi sonucu değerlendirici arası uyum her iki puanlama anahtarı için yüksek çıkmıştır. Değerlendirici deneyim süresinin değerlendirme sonuçlarına etkisi için yapılan Mann-Whitney U testi sonucunda anlamlı veri elde edilememiştir. Yapılan testlerle A1 ve A2 düzeyi için hazırlanan puanlama anahtarlarının geçerli ve güvenilir olduğu sonucuna ulaşılmıştır.

Bilgilendirme

Bu çalışma Hacettepe Üniversitesi Türkiyat Araştırmaları Enstitüsü Yabancı Dil olarak Türkçe Öğretimi Doktora Programı bünyesinde Doç. Dr. Meltem Ekti danışmanlığında hazırlanmış "Yabancı Dil Olarak Türkçe Öğretiminde Konuşmanın Değerlendirilmesi İçin Puanlama Anahtarı Önerisi: A1-A2 Düzeyi" adlı doktora tezinden üretilmiştir.

Yayın Etiği Bilgilendirme

Bu çalışma bilimsel yayın etiğine uygun olarak yapılmıştır.
Kurulun Adı: Hacettepe Üniversitesi Sosyal ve Beşeri Bilimler Araştırma Etik Kurulu

Tarih: 11.04.2023

Karar numarası: E-35853172-000-00002804092

Araştırma tek yazar tarafından yürütülmüştür.

Bu çalışmada bir çıkar çatışması yoktur.

Yazarların Katkı Oranı

Çıkar Çatışması

Gönderim

Taşköprü, G. (2025). İkinci/yabancı dil olarak Türkçe öğretiminde A1-A2 düzeylerinde konuşma puanlama anahtarı önerisi. *Dil Eğitimi ve Araştırmaları Dergisi, 11(1)*, 188-206.

* Dr, ORCID ID: <https://orcid.org/0000-0002-9724-4698>, Dokuz Eylül Üniversitesi, Dil Eğitimi Uygulama ve Araştırma Merkezi, gunsu.taskopru@deu.edu.tr

Giriş

Yabancı dil öğretimi ve ikinci dil ediniminde sürecin önemli basamaklarından biri ölçme değerlendirmedir. Öğrencilerin öğrenme hedeflerine ulaşip ulaşmadıklarını sınamak, yeterliklerini ölçebilmek, sınıflandırmak ve ilerleyişlerini sınamak için ölçme ve değerlendirme araçları kullanılır. Dil öğretiminde genellikle sınavlar yardımıyla öğrencilerin yeterliği sınanmaktadır. Dört temel beceri olan dinleme, konuşma, yazma ve okuma birlikte veya beceri özelinde hazırlanmış sınavlar aracılığıyla ölçülüp değerlendirilebilmektedir. Her bir beceri için ölçme ve değerlendirme araçları oluşturulurken farklı türlerde yöntem ve teknikler gerekmektedir. Dinleme ve okuma gibi alımlama becerileri ile konuşma ve yazma gibi üretim becerileri ölçme ve değerlendirme sürecinde farklı ölçme ve değerlendirme araçları gerektirir. Becerilerin farklı özelliklerine dikkat edilmesi hem sınavı hazırlayan kişiler açısından hem de değerlendiren kişiler açısından önem teşkil etmektedir. Bu nedenle hazırlanan ölçme ve değerlendirme araçları, hedeflenen bir veya birden çok becerinin gerektirdiklerini yansıtmalıdır. Hazırlanan bu araçların belirli kurallar çerçevesinde oluşturulması ve uygulanması kişilerin dil öğrenim süreçlerinin daha verimli değerlendirilebilmesini sağlayabilmektedir.

Konuşma becerisinin değerlendirilme sürecini etkileyen birçok faktör bulunmaktadır. Değerlendiricilerin, değerlendirme sırasında konuşmanın hangi boyutlarına yoğunlaşacaklarına hafızalarına veya aldıkları notlara dayanarak karar vermeleri gerekmektedir. Diğer becerilerin aksine konuşma becerisinin değerlendirilmesi eşzamanlı gerçekleşmektedir. Değerlendiricilerin çoğunlukla değerlendirme yaparken geriye dönme şansı yoktur. Bu nedenle bazı durumlarda değerlendiriciler doğru ve adil değerlendirmeler yapmakta zorlanabilirler. Örneğin, sınava giren öğrencinin konuşma becerileri yerine dünya bilgisini değerlendirmek, değerlendiricinin sınava giren kişi hakkında sahip olduğu bilginin anlık değerlendirmesini etkilemesi gibi durumlar adil değerlendirmenin önüne geçebilmektedir (Metruk, 2018). Bu nedenle değerlendiricilerin öğrencilerin üretim performanslarını değerlendirdikleri yazma ve konuşma becerilerinde sistemli, kullanışlı ve açık dereceli puanlama anahtarları kullanması değerlendirme sürecini kolaylaştırmak açısından önemlidir.

Sınıf içi değerlendirmede ve sınavlarda dereceli puanlama anahtarları sıklıkla kullanılmaktadır. Fulcher ve Davidson (2012, s. 378) puanlama anahtarlarını öğrencilerin performansını ölçmede kullanılan, yeterlik tanımlayıcılarından oluşan hiyerarşik bir sistem olarak tanımlanmıştır. Berger'e göre (2015, s. 62), dereceli puanlama anahtarlarının yatay ve dikey boyutları vardır; yatay boyut öğrencilerin dili kullanırken neleri yapabildiğini görevler ve işlevler açısından tanımlarken, dikey boyut ise önceden belirlenmiş kriterlere göre öğrencinin konuşma performansının niteliksel farklarını genellikle sayısal biçimde yer alan referans cetveli ile göstermektedir. Öğrenciler her iki boyutta yer alan bilgilere göre değerlendirilirler. Dereceli puanlama anahtarı boyutlarında yer alan bilgiler, ölçülmesi ve değerlendirilmesi hedeflenen yeterliklere göre belirlenmektedir. Dereceli puanlama anahtarları yapılacak değerlendirmenin niteliğine göre bütüncül ve analitik olmak üzere iki türde oluşturulmaktadır. Bütüncül dereceli puanlama anahtarları kullanıldığında değerlendirilmek istenen becerinin tüm boyutları bir arada ele alınırken, analitik dereceli puanlama anahtarlarında her bir boyut için ölçütler belirlenir ve ayrı olarak değerlendirilir

(Brookhart, 2013). TELC, Cambridge, British Council gibi uluslararası sınav kuruluşlarının kullandığı konuşma puanlama anahtarları incelendiğinde genellikle bütüncül türde hazırlandığı görülmüştür. British Council sınavları için puanlama anahtarları hazırlanmasında görev alan O’Sullivan (2015) değerlendiricilerin kolay puanlaması ve hazırlanan sınavların doğası gereği bütüncül puanlama anahtarı oluşturulduğunu belirtmiştir. Bütüncül puanlama anahtarları daha kısa zamanda daha hızlı değerlendirme ve puanlama yapılabilmesini sağlamaktadır. Analitik dereceli puanlama anahtarları ise konuşma performanslarını daha ayrıntılı değerlendirme imkânı sağlamaktadır. Davis (2018, s. 2) analitik puanlama anahtarlarının dilbilgisi, akıcılık, organizasyon gibi dilin boyutlarının birbirinden ayrı olarak puanlanmasının, öğrencinin performansındaki tutarsızlıkları tespit etmeyi kolaylaştırdığını belirtmiştir. Alanyazında araştırmacıların analitik ve bütüncül puanlama anahtarlarını karşılaştırdıkları çalışmalarında, bütüncül puanlama anahtarları deneyimli değerlendiriciler tarafından kullanıldığında deneyimsiz değerlendiricilerin kullanımına göre daha güvenilir sonuçlar elde edildiğini; daha detaylı bilgiler içeren analitik puanlama anahtarları kullanıldığında ise deneyimli ve deneyimsiz değerlendiriciler arasında daha az fark görüldüğünü belirtmişlerdir (Barkaoui, 2010; Deygers vd., 2013; Fulcher vd., 2011; Harsch & Martin, 2013). Analitik puanlama anahtarları ile değerlendirme süresi bütüncül puanlama anahtarlarına göre daha uzun sürebilmektedir. Ancak analitik puanlama anahtarları öğretmenlere, öğrenciler hakkında daha detaylı bilgiye erişilebilmelerinden dolayı bütüncül puanlama anahtarlarına göre geribildirim açısından avantaj sağlamaktadır. Kişiler ve kurumlar ihtiyaçlarına ve amaçlarına göre genellikle analitik veya bütüncül türlerden birini seçerek puanlama anahtarlarını tasarlamaktadır.

Yabancı dil olarak Türkçe öğrenmek için genellikle üniversitelerin dil merkezlerini tercih eden öğrencilere merkezler kendi sınavlarını uygulamaktadır. Bu sınavları çoğunlukla derslere giren öğretmenler hazırlamakta, uygulamakta ve değerlendirmektedir. Kurumlar arasında ölçme araçlarını geliştirme ve uygulama açısından bazı farklar görülmektedir. Son yıllarda yabancı dil olarak Türkçe öğretiminde ölçme ve değerlendirmede öğretici yeterliği ile ilgili çalışmalara bakıldığında Boylu ve Başar (2016) yaptıkları çalışma sonucunda kurumlar arasında bir standart görülmediğini belirtmiştir. Çerçi (2018), yabancı dil olarak Türkçede konuşma becerisini değerlendirme ile ilgili yaptığı çalışmada konuşma becerisi için değerlendirme araçlarının eksik olduğunu, bu alanda öğretim elemanlarının eksikliği gibi açılardan problemler tespit edildiğini belirtmiştir. Konuşma becerisi ile ilgili çalışmalara bakıldığında Boylu (2019) yabancı dil olarak Türkçe alanında ölçme ve değerlendirme uygulamalarıyla ilgili yaptığı çalışmada öğretmenlerin ölçme ve değerlendirme yeterlik algılarını ve konuyla ilgili tutumlarını incelemiş, uygulama aşamasında kişiler arasında farklılıklar ve tutarsızlıklar tespit etmiş, bu durumun önüne geçilebilmesi için örnek ve önerilere yer vermiştir. Kocayanak (2021), öğretmenlerin konuşma becerisini ölçme değerlendirme süreci ve bu konudaki yeterlikleri hakkında yaptığı çalışmada öğretmenlerin konuşma becerisini değerlendirirken hangi alanlara dikkat ettiklerini ve kullandıkları teknikleri araştırmış, öğretmenlerin Diller İçin Avrupa Ortak Öneriler Çerçeve Metni kazanımlarına uygun sorular ve puanlama anahtarları kullanmayı tercih ettiklerini gözlemlemiştir. Konuşma becerisi için puanlama anahtarı geliştirme çalışmalarına bakıldığında Yorgancı (2021) doktora tezinde podcast kullanımının konuşma

ve yazma becerilerini geliştirip geliştirmediğini gözlemlediği çalışmasında B1 düzeyinde konuşma ve yazma sınavlarının yanı sıra bağımsız ve karşılıklı konuşma için iki analitik puanlama anahtarı geliştirmiştir. Şimşek (2022), yüksek lisans tezinde B2 ve C1 düzeylerinde dereceli puanlama anahtarları geliştirmiş ve geçerlik güvenirlik sonuçlarına dair bilgilere yer vermiştir. Ayverdi (2023) ise C1 düzeyinde konuşma için puanlama anahtarı geliştirdiği çalışmasında sözlü anlatım ve karşılıklı konuşma için iki ayrı puanlama anahtarı oluşturmuş ve geçerlik güvenirlik çalışmalarına yer vermiştir. Yabancı dil olarak Türkçe alanında konuşma becerisi için puanlama anahtarı geliştirme çalışmalarının, başlangıç düzeylerinden daha çok orta ve üst düzeylerdeki öğrencilerin konuşma becerilerinin ölçülmesi ve değerlendirilmesine odaklandığı görülmektedir.

Öğrencilerin konuşma becerisinin gelişiminin doğru takibi için temel düzeylerde çalışmaların yapılması gerekmektedir. Diller İçin Avrupa Ortak Öneriler Çerçeve Metni (DAOÇM) konuşma becerisinde öğrencilerin yeterli ve yetkin olması gereken bilişsel, dilsel ve sesbilgisel becerilerden bahsetmektedir. Genel olarak A1 ve A2 düzeylerinde öğrencinin kendisini ve çevresini anlatabileceğini, basit ve kalıplaşmış ifadeler kullanarak ortaya koyduğunu belirtir (Avrupa Konseyi, 2013). Öğrenciler A1 ve A2 düzeylerinde hedef dilin özelliklerini ve dil bilgisini yeni öğrendiğinden sınırlı bilgiye sahiptirler. Bu bilgileri öğrencinin doğru biçimde öğrenebildiğini ve kullanabildiğini görebilmek için ölçme ve değerlendirme araçlarına başvurulması gerekmektedir. DAOÇM yabancı dilde konuşma becerisinde “dil/söz varlığı”, “doğruluk”, “akıcılık”, “etkileşim”, “tutarlılık”, “sesbilim” alanlarında A1’den C2 düzeyine kadar yeterliklere yer vermiştir. Türkiye’de ve uluslararası yabancı dillerde yapılan ölçme değerlendirme çalışmalarında da benzer başlıkların kullanıldığı görülmektedir. Köse (2008) sözlü sınavların değerlendirilmesinde sözlü anlatım ve karşılıklı konuşma başlıklarına yer vermiş ve ikisi için ayrı boyutlar belirlemiştir. Sözlü anlatım başlığı altında “dilbilgisi”, “sözcük bilgisi”, “akıcılık”, “tutarlılık”, “sesletim” boyutları yer alırken, karşılıklı konuşma başlığı altında “dilbilgisi,” “sözcük bilgisi”, “akıcılık”, “dinleme becerisi”, “iletişim kurma” ve “diyalogu sürdürme” boyutları bulunmaktadır. Kocayanak ve Tecelli (2023) çalışmalarında öğrencilerin doğruluk, tutarlılık, akıcılık özgünlük ve etkileşim, anlatım ve zamanı kullanma ölçütlerine önem verdiklerini belirtmiştir. Alanda yapılmış puanlama anahtarları çalışmalarında kişi ve kurumların benzer ölçütler kullandığı bazı ölçütlerin ise değişiklik gösterdiği görülmüştür. Uluslararası yabancı dil sınavları yapan kurumlardan TELC’in konuşma sınavları için kullandığı puanlama anahtarında “sözlü etkileşim, sözlü anlatım, genel dilsel araçlar alanı, dilbilgisel doğruluk, akıcılık, tutarlılık ve bağdaşıklık, telaffuz ve tonlama hakimiyeti” başlıklarına yer verildiği görülmektedir. Cambridge’in (2011) A2 düzeyi konuşma sınavı için hazırladığı puanlama anahtarında “sözvarlığı, dilbilgisi, sesletim ve interaktif iletişim” başlıkları görülmektedir. British Council (O’Sullivan & Dunlea, 2019) tarafından hazırlanan bütüncül puanlama anahtarı ise “görev başarımları/konuya uygunluk, dilbilgisi doğruluğu, sözcük dağarcığı, sesletim, akıcılık” hakkında DAOÇM’yi temel alan maddelerden oluşmaktadır. Alanyazında görüldüğü üzere DAOÇM’yi temel alan kişi ve kurumların hazırladığı puanlama anahtarlarında bazı başlıkların ortak olduğu, başlık içerisindeki tanımlayıcıların ise değişkenlik gösterdiği görülmektedir. Bu durumda kişi ve kurumlar arasında puanlama anahtarları kriterlerinde belirli bir standart görülmekle beraber

farklılıklar da gözlemlendiği söylenebilir. Öğrenci, öğretici ve kurumların hedefleri açısından A1'den C1 düzeyine kadar standart, sistematik ve kullanışlı puanlama anahtarlarının kullanması gerekmektedir. Puanlama anahtarı geliştirme çalışmalarının A1 ve A2 düzeylerine de odaklanması, dil öğrenim sürecinin başından itibaren ölçme ve değerlendirme sürecinin doğru biçimde ilerlemesi için önemlidir. Bu amaçla temel düzeylerden itibaren uygun puanlama anahtarlarının hazırlanması ve kullanımına başlanmasıyla öğretici ve değerlendiriciler standart ölçme ve değerlendirme sistemleri oluşturabilir ve bu şekilde öğrencilerin bir üst düzeye geçiş yaparken yeterliklerini ve ihtiyaçlarını belirlemek için doğru ve yeterli veriler elde edebilecekleri düşünülmektedir.

Araştırmanın Amacı ve Araştırma Soruları

Bu çalışmada başlangıç düzeyleri olan A1 ve A2 düzeylerinde analitik dereceli puanlama anahtarları oluşturularak yabancı dil ve ikinci dil olarak Türkçe öğretimi alanında temel düzeyde yansız sonuçlar elde edilmesine olanak sağlamak amaçlanmaktadır. Değerlendiricilerin hızlı, kolay ve açık değerlendirmeler yapabilmelerine katkı sağlamak amacıyla her iki düzey için birer analitik dereceli puanlama anahtarları geliştirilmesi amaçlanmıştır. Bu amaçtan hareketle araştırmada aşağıdaki soruların cevapları aranacaktır:

1. A1 düzeyinde yabancı dil olarak Türkçe konuşma becerisi için geçerli ve güvenilir dereceli puanlama anahtarları oluşturulabilir mi?
2. A2 düzeyinde yabancı dil olarak Türkçe konuşma becerisi için geçerli ve güvenilir dereceli puanlama anahtarları oluşturulabilir mi?
3. A1 ve A2 düzeylerindeki konuşma performanslarının değerlendirilmesinde değerlendiricilerin deneyim sürelerinin bir etkisi bulunmakta mıdır?

Yöntem

Araştırma Tasarımı/Modeli

Araştırma kapsamında dereceli puanlama anahtarı geliştirmek için hem nitel hem nicel verilerin toplanması gerekmektedir. Bu nedenle çalışmada keşfedici ardışık desenli karma yöntem uygulanmıştır. Keşfedici ardışık desende, öncelikle araştırma problemini ortaya koymak için nitel veri toplanır ve bu verinin analizi yapılarak bir veri toplama aracı oluşturulur ve oluşturulan araçla nicel veri toplanarak, elde edilen verilerle sonuçlara ulaşılır (Creswell, 2021). Karma yöntem, dereceli puanlama anahtarları oluşturma adımları ile uyumlu olduğundan tercih edilmiştir.

Dereceli puanlama anahtarlarının oluşturulmasında Stevens ve Levi (2005) ve Gehlbach ve Brinkworth'un (2011) puanlama anahtarı geliştirme aşamalarından yararlanılarak, Creswell (2021)'in karma yöntem modeline uygun biçimde 11 adımlık bir sistem oluşturulmuştur. Çalışma karma yöntemin içeriğine uygun olarak nitel ve nicel aşamalardan meydana gelmektedir. Nitel aşamada katılımcı gruplarının oluşturulması, literatür tarama, kapsam geçerlik testlerinin yapılması adımları bulunmaktadır. Nicel aşamada ise pilot uygulamanın yapılması ve güvenilirlik çalışmaları adımları yer almaktadır.

Araştırma “nitel aşama”, “nicel aşama” ve “verilerin analizi” ana başlıklarından oluşmaktadır. Bu adımlar aşağıda açıklanmıştır:

a. Nitel Aşama

1. Dereceli puanlama anahtarının geliştirilme amacının belirlenmesi: Bu aşamada puanlama anahtarının oluşturulması için genel bir plan hazırlanmıştır. Hangi amaçla puanlama anahtarlarının oluşturulacağı belirlenmiştir.

2. Odak grubun belirlenmesi: Dereceli puanlama anahtarının oluşturulması için öncelikle hedef grup belirlenmiştir. Başlangıç düzeyleri olan A1 ve A2 düzeylerindeki öğrencilerin konuşma becerilerini değerlendirmek hedef alınarak çalışma yürütülmüştür. Çalışmaya tamamen gönüllülük esasına dayanarak yabancı dil olarak Türkçe öğretimi alanında çalışan uzmanlar, A1 ve A2 düzeylerinde öğrenim gören uluslararası öğrenciler ve dil merkezlerinde eğitim veren öğretim elemanları katılmıştır.

3. Puanlayıcı türünün belirlenmesi: Dereceli puanlama anahtarlarının kullanılacağı sınav türüne göre bir puanlama sistemi oluşturulmuştur. Hazırlanan puanlama anahtarları A1 ve A2 düzey sonu sınavlarında kullanılmak üzere hazırlandığından çoğunlukla tercih edilen 25’lik veya 100’lük sistemde kullanılabilecek şekilde geliştirilmiştir.

4. Literatür taramasının yapılması: Hedef grubun değerlendirilmesinde kullanılan yeterlik ve kazanımlar Türkiye’de ve uluslararası alandaki çalışmalar taranarak belirlenmiş, düzey ve hedefler gözetilerek düzenlenen boyutlar ve boyutları oluşturan maddeler literatürden elde edilen bilgilere göre hazırlanmıştır.

5. Boyutların oluşturulması ve betimlenmesi: Literatür taraması sonucunda elde edilen veriler ve alan uzmanlarının katkılarıyla boyutlar ve ölçüt maddeleri belirlenmiştir. Her bir boyut A1 ve A2 düzeyinin gerektirdiği yeterliklere göre tanımlanmıştır.

6. Dereceli puanlama anahtarlarının taslaklarının oluşturulması: Boyutlar ve ölçütler betimlendikten sonra analitik dereceli puanlama anahtarı biçimine getirilip A1 ve A2 düzeyleri için iki adet puanlama anahtarı oluşturulmuştur.

7. Geçerlik testleri: Taslak dereceli puanlama anahtarlarının saha uygulamasından önce alanında uzman kişilerden görüş alınarak, elde edilen verilerle Lawshe (1975) tekniği kullanılarak dereceli puanlama anahtarlarının kapsam geçerlikleri hesaplanmıştır. Yabancı dil olarak Türkçe öğretiminde en az 10 yıl deneyimli sekiz alan uzmanı tarafından maddeler üç dereceli bir ölçek ile değerlendirilmiştir. Bu ölçek “madde hedeflenen yapıyı ölçüyor”, “madde yapı ile ilişkili ancak gereksiz”, “madde hedeflenen yapıyı ölçmüyor” şeklinde basamaklandırılmıştır. Uzmanların görüşleri sonucunda uygun olmayan maddeler çıkarılarak puanlama anahtarları saha uygulamasına hazır hale getirilmiştir.

8. Değerlendirici kılavuzlarının oluşturulması: Değerlendiricilerin puanlama anahtarlarını doğru biçimde kullanmayı bilmeleri puanlama anahtarlarının güvenilirliği açısından önemlidir. Bu nedenle değerlendiricilerin puanlama anahtarı kullanımı için eğitim görmeleri doğru değerlendirme yapılmasını olumlu biçimde etkileyebilir. Jeong (2015), değerlendiricilerin puanlama anahtarlarını doğru kullanmayı bilmek için eğitim verilmesi gerektiğini, eğitim verilmesinin zor olduğu durumlar için ise kılavuzların oluşturulmasının yararlı olacağını belirtmiştir. Değerlendiricilerin boyut ve ölçütleri doğru biçimde yorumlayıp puanlama yapabilmelerine olanak sağlayan hem A1 hem de A2 düzeylerinde

hazırlanan puanlama anahtarları için kılavuz kitapçıklar oluşturulmuştur. Oluşturulan kılavuz kitapçıklar da puanlama anahtarları gibi alan uzmanları tarafından incelenmiştir.

b. Nicel Aşama

9.Saha uygulaması: Geçerlik testlerinden sonra oluşturulan dereceli puanlama anahtarları öğrencilerin konuşma performanslarını değerlendirmede kullanılmıştır. A1-A2 düzeyi öğrencilerin konuşma performansı kayıtları deneyimli ve deneyimsiz öğretici grubu tarafından taslak dereceli puanlama anahtarları ile değerlendirilmesi sonucunda güvenilirlik çalışmaları için nicel veri elde edilmiştir.

10. Güvenirlik testleri: Güvenirlik çalışmaları için değerlendiriciler arası uyuma bakılmıştır. Bu uyum Kendall-W testi ile hesaplanarak dereceli puanlama anahtarlarının güvenilirlikleri hesaplanmıştır. Kendall-W testinde uyum değeri W ile gösterilir ve bu değer 0 (uyumsuz) ile 1 (tamamen uyumlu) arasında değişmektedir (Gwet, 2020, s. 382).

c. Verilerin Analizi

11. Son kontrol: Bu aşamada geçerlik ve güvenilirlik testlerinden elde edilen veriler analiz edilmiş ve elde edilen verilerle birlikte puanlama anahtarları son kez kontrolden geçirilmiştir.

Yayın Etiği

Kurulun Adı: Hacettepe Üniversitesi Sosyal ve Beşeri Bilimler Araştırma Etik Kurulu

Tarih: 11.04.2023

Karar numarası: E-35853172-000-00002804092

Katılımcılar

Çalışmada geçerlik çalışmaları için alan uzmanlarından oluşan bir grup, konuşma performanslarının elde edildiği öğrenci grubu ve konuşma performanslarını değerlendirmek için öğretmenlerden oluşan değerlendirici grubu olmak üzere üç çalışma grubundan yararlanılmıştır. Çalışmaya katılan bütün katılımcılar gönüllülük esasıyla katılmıştır. Kapsam geçerlik testleri için oluşturulan alan uzmanları grubunun seçiminde ölçüt örneklemeden yararlanılmıştır. Yıldırım ve Şimşek (2013)'e göre önceden belirlenen ölçütlere göre örneklem seçilmesi olarak tanımlanmıştır. Deneyim süreleri 10 yıl ile 18 yıl arasında değişen, daha önce yabancı dil olarak Türkçe dersi vermiş, doktora yapmış veya doktora devam eden araştırmacılar alan uzmanı grubunu oluşturmaktadır.

Güvenirlik çalışmalarının yapılabilmesi için hazırlanan dereceli puanlama anahtarlarının öğrenciler üzerinde uygulanması gerekmektedir. Bu nedenle 2022-2023 eğitim öğretim yılında Dokuz Eylül Üniversitesi Dil Eğitim ve Araştırma Merkezi'nde (DEDAM) öğrenim gören gönüllülük esaslı çalışmaya katılan uluslararası öğrencilerin A1 ve A2 düzeylerinde konuşma performansları kayıt altına alınarak, değerlendirici grubunun performansları değerlendirmesi istenmiştir. Aykırı örnekleme yöntemiyle her bir düzey için iki zayıf, iki orta ve iki iyi olduğu düşünülen konuşma performansı belirlenmiştir. Aykırı örnekleme birbirinden farklı durumlarda olan örneklerin seçilmesiyle grupların oluşturulduğu bir yöntemdir (Yıldırım ve Şimşek, 2013). Öğrencilerin her birinin farklı

ülkelerden olması, farklı anadillere sahip olmaları da seçim için önemli bir diğer etken olmuştur.

Maksimum çeşitlilik örnekleme yöntemiyle seçilen son çalışma grubu, konuşma performanslarını değerlendiren katılımcılardan oluşmaktadır. Maksimum çeşitlilik örnekleme yöntemi farklılık gösteren gruplar arasındaki ortaklık ve farklılıklara odaklanır (Yıldırım ve Şimşek, 2013). Türkiye’deki üniversitelere bağlı dil merkezlerinde farklı düzeylerde eğitim veren tecrübeli-tecrübesiz öğreticiler arasından 20 değerlendirici karışık olarak seçilmiştir. Çalışmaya katılan değerlendirmecilerin deneyimleri 6 aydan 18 yıla kadar çeşitlilik göstermektedir. Katılımcılardan 1 kişi yüksek lisans mezunu, 9 kişi doktora devam etmekte, 10 kişi doktora mezunudur. Katılımcıların son mezun oldukları bölümler İngiliz Dil Eğitimi, Fransız Dili, Yabancı dil Olarak Türkçe Öğretimi, Dilbilim, Türk dili edebiyatı ve Türkçe Eğitimi olarak görülmektedir. 8 katılımcı Yabancı dil Olarak Türkçe Öğretimi doktorası yapmış veya yapmaktadır.

Veri Toplama ve Verilerin Çözümlemesi

Kapsam geçerlik çalışmaları için gerekli veriler uzman görüşü ile alınmıştır. Uzman görüşü için gerekli form “madde hedeflenen yapıyı ölçüyor”, “madde yapı ile ilişkili ancak gereksiz”, “madde hedeflenen yapıyı ölçmüyor” tanımlayıcıları içeren üçlü likert tipi bir ölçek ile alan uzmanlarından toplanmıştır. Ölçek ile toplanan veriler Lawshe (1975) yöntemi kullanılarak analiz edilmiştir. Lawshe yöntemine göre en az 5 en çok 40 uzman sayısına göre hesaplanan madde değerleri, belirlenen minimum değerle karşılaştırılarak kapsam geçerliği elde edilmektedir. Kapsam geçerliği hesaplanırken uzman sayılarına göre göz önüne alınması gereken minimum değer tablosu aşağıdadır (Lawshe, 1975, s. 568):

Çizelge 1. Lawshe Minimum Değer Tablosu

Uzman Sayısı	Minimum Değer
5	.99
6	.99
7	.99
8	.75
9	.78
10	.62
11	.59
12	.56
13	.54
14	.51
15	.49
20	.42
25	.37
30	.33
35	.31
40	.29

Bu çalışmada 8 uzmandan görüş elde edilmiştir. Tabloya göre 8 uzman için maddelerin kapsam geçerlik oranlarının hesaplanmasında minimum değer .75 olmalıdır. Bu

oranın altında kalan maddeler kapsam geçerliğini sağlamadığından puanlama anahtarından çıkarılmıştır. Daha sonrasında puanlama anahtarlarının tamamı için kapsam geçerlik indeksi hesaplanmıştır.

Güvenirlik çalışmaları için ise hazırlanan dereceli puanlama anahtarları kullanılmıştır. 20 değerlendiricinin 6 A1 ve 6 A2 düzeyi öğrencilerinin sözlü performansını değerlendirdikleri puanlama anahtarlarından elde edilen puanlar SPSS 29 ile analiz edilmiştir. Analiz için Kendall-W testi kullanılmıştır. Değerlendiriciler arası uyumu gösteren Kendall-W testi için elde edilen veriler aşağıdaki yorumlama tablosuyla karşılaştırılmıştır (Schmidt, 1997):

Çizelge 2. Kendall-W Uyum Katsayısı Yorumlama Tablosu

W	Yorum	Güven Sıralaması
.1	Çok zayıf uyum	Uyum yok
.3	Zayıf uyum	Düşük
.5	Orta düzeyde uyum	Makul
.7	Güçlü uyum	Yüksek
.9	Çok güçlü uyum	Çok yüksek

Tabloya göre elde edilen değer 0'a ne kadar yakınsa değerlendiriciler arası uyum az, 1'e ne kadar yakınsa uyumun yüksek olduğu şeklinde yorumlanmaktadır. Bu test sonucunda puanlama anahtarlarının güvenilirlik sonuçları elde edilmiştir.

Çalışmada deneyimli ve deneyimsiz değerlendiricilerin yaptıkları değerlendirme sonuçlarında fark olup olmadığını belirlemek için ise Mann-Whitney U testi uygulanmıştır. İki bağımsız grup arasında ilişki olup olmadığını görmek için uygulanan bir testtir (Büyüköztürk vd., 2020, s. 187). Bu test için puanlama anahtarlarından elde edilen veriler ile kişisel bilgi formundan elde edilen deneyim süreleri SPSS 29 aracılığıyla karşılaştırılmıştır. Kişisel bilgi formu ile değerlendirici grubu hakkında deneyim, öğrenim düzeyleri gibi bilgiler elde edilmiştir.

Bulgular

A1 ve A2 düzeylerinde konuşma becerisi için hazırlanan dereceli puanlama anahtarlarının geçerlik ve güvenilirlik testlerine dair bulgulara aşağıda yer verilmiştir.

Çizelge 3. A1 Düzeyi Konuşma Ölçeği Madde Kapsam Geçerlik Oranları ve Kapsam Geçerlik İndeksi

		n	Minimum Değer	KGO
1. Bölüm: Sözcük Bilgisi	M1	8	.75	1
	M2	8	.75	1
	M3	8	.75	1
	M4	8	.75	1
	M5	8	.75	1
2. Bölüm: Dilbilgisel Doğruluk	M6	8	.75	1
	M7	8	.75	1
	M8	8	.75	1
	M9	8	.75	1
	M10	8	.75	0

	M11	8	.75	1
3. Bölüm: Tutarlık ve Bağdaşıklık	M12	8	.75	1
	M13	8	.75	1
	M14	8	.75	1
	M15	8	.75	1
4. Bölüm: Akıcılık	M16	8	.75	1
	M17	8	.75	1
5. Bölüm: Sesletim	M18	8	.75	1
	M19	8	.75	0
	M20	8	.75	0
6. Bölüm: Karşılıklı Etkileşim	M21	8	.75	1
	M22	8	.75	.75
	M23	8	.75	1
	M24	8	.75	1
	M25	8	.75	1
	M26	8	.75	1
	M27	8	.75	1
	M28	8	.75	1
7. Bölüm: Görevi Başarma	M29	8	.75	-0.25
	KGI			0.99

Çalışmanın kapsam geçerliği kısmında A1 düzeyinde 29 maddeden minimum değer olan .75'i karşılamayan dilbilgisel doğruluk bölümünden M10, sesletim bölümünden M19, karşılıklı etkileşim bölümünden M20 ve görevi başarma bölümünden M29 çıkarılmıştır. Çıkarılan 4 madde dışında minimum değeri karşılayan karşılıklı etkileşim bölümünden M22 için .75, diğer maddeler için ise 1 sonucu elde edilmiştir. 8 uzman görüşü sonucunda 25 maddede kapsam geçerlik oranı için minimum değer olan .75 sağlandığı görülmüştür. A1 düzeyi için kapsam geçerlik indeksinin .99 olduğu görülmüştür.

Çizelge 4. A2 Düzeyi Konuşma Ölçeği Madde Kapsam Geçerlik Oranları ve Kapsam Geçerlik İndeksi

		n	Minimum Değer	KGO
1. Bölüm: Sözcük Bilgisi	M1	8	.75	1
	M2	8	.75	1
	M3	8	.75	1
	M4	8	.75	1
2. Bölüm: Dilbilgisel Doğruluk	M5	8	.75	1
	M6	8	.75	1
	M7	8	.75	1
	M8	8	.75	1
	M9	8	.75	1
	M10	8	.75	-.25
	M11	8	.75	1
3. Bölüm: Tutarlık ve Bağdaşıklık	M12	8	.75	1
	M13	8	.75	1
	M14	8	.75	1
4. Bölüm: Akıcılık	M15	8	.75	1
	M16	8	.75	1
	M17	8	.75	1
5. Bölüm: Sesletim	M18	8	.75	1
	M19	8	.75	.75
	M20	8	.75	1

6. Bölüm: Karşılıklı Etkileşim	M21	8	.75	.75
	M22	8	.75	1
	M23	8	.75	1
	M24	8	.75	1
7. Bölüm: Görevi Başarma	M25	8	.75	1
	M26	8	.75	1
	M27	8	.75	0
KGI				.98

A2 düzeyinde ise 27 madde içerisinde dilbilgisel doğruluk bölümünden M10 ve görevi başarma bölümünden M27 maddeleri minimum değer olan .75'i karşılamadığından çıkarılmış, sesletim bölümünden M19 ve karşılıklı etkileşim bölümünden M21 için .75, diğer maddeler için ise 1 geçerlik kapsam oranı bulunmuştur. A2 düzeyi için kapsam geçerlik indeksi .98 olarak hesaplanmıştır.

Çalışmanın güvenirlik kısmında ise değerlendiricilerin A1 ve A2 düzeylerinde uluslararası öğrencilerin konuşma performanslarını değerlendirmesi sonucu elde edilen veriler arasındaki uyum, Kendall-W testiyle genel açıdan ve boyutlar açısından karşılaştırılmış, aşağıdaki tablolarda görüldüğü biçimde özetlenmiştir.

Çizelge 5. A1 Düzeyi Boyutlara İlişkin Kendall-W Kat Sayıları

Boyutlar	W	p
Sözcük Bilgisi	.705	.00
Dilbilgisel Doğruluk	.643	
Tutarlık ve Bağdaşıklık	.710	
Akıcılık	.616	
Sesletim	.708	.00
Karşılıklı Etkileşim	.616	
Görevi Başarma	.605	
A1 Düzeyi Genel	.862	

A1 düzeyi dereceli puanlama anahtarının güvenirlik çalışmaları için 20 değerlendirici arasındaki uyuma bakılmıştır. Kendall-W ile A1 düzeyi puanlama anahtarında yer alan boyutlara ilişkin değerlendiriciler arası uyuma bakıldığında p değeri .00 olarak görülmüştür. Bu değer .05'ten küçük olduğundan sonuçlar istatistiksel olarak anlamlıdır. Sözcük bilgisi boyutu için .705 değeri ile yüksek derece uyum, dilbilgisel doğruluk boyutunda .643 değeri ile orta derecede uyum, tutarlık ve bağdaşıklık boyutunda .710 değeri ile yüksek derece uyum, akıcılık boyutunda .616 değeri ile orta derecede uyum, sesletim boyutunda .708 değeri ile yüksek derece uyum, karşılıklı etkileşim boyutunda .616 değeri ile orta derecede uyum, görevi başarma boyutunda ise .605 değeri ile orta derecede uyum görülmektedir.

A1 düzeyinin geneli için p değerinin .00 olduğu görülmüştür. Bu değer .05'ten küçük olduğundan sonuçların istatistiki açıdan anlamlı olduğunu göstermektedir. Dereceli puanlama anahtarının geneli için Kendall-W katsayısı .862 olarak bulunmuştur. Bu değer yorumlama tablosuna göre yüksek derecede uyum anlamına gelmektedir.

Çizelge 6. A2 Düzeyi Boyutlara İlişkin Kendall-W Kat Sayıları

Boyutlar	W	p
----------	---	---

Sözcük Bilgisi	.727	.00
Dilbilgisel Doğruluk	.802	
Tutarlık ve Bağdaşıklık	.740	
Akıcılık	.727	
Sesletim	.634	
Karşılıklı Etkileşim	.663	
Görevi Başarma	.724	
A2 Düzeyi Genel	.845	.00

A2 düzeyi dereceli puanlama anahtarının güvenirlik çalışmaları için de 20 değerlendirici arasında uyuma bakılmış; p değeri hem boyutlar özelinde hem de puanlama anahtarı genelinde .00 olarak görülmüş, değer .05'ten küçük olduğundan elde edilen sonuçların istatistiki açıdan anlamlı olduğunu göstermiştir.

Boyutlar özelinde değerlendiriciler arasındaki uyuma bakıldığında sözcük bilgisi için .727 değeri ile yüksek derece uyum, dilbilgisel doğruluk .802 değeri ile yüksek derece uyum, tutarlık ve bağdaşıklık .740 değeri ile yüksek derece uyum, akıcılık boyutu .727 değeri ile yüksek derece uyum, sesletim boyutunda .634 değeri ile orta derecede uyum, karşılıklı etkileşim .663 değeri ile orta derecede uyum, görevi başarma boyutunda ise .724 değeri ile yüksek derece uyum görülmektedir. Dereceli puanlama anahtarının geneli için ise .845 değeri ile yüksek derecede uyum gözlemlenmiştir.

Değerlendirmeyi yapan kişilerin deneyimleri 6 aydan 18 yıla kadar değişiklik göstermektedir. Değerlendiricilerin 13'ü 10 yıl ve üzeri deneyime sahipken 7'si 5 yıl ve daha az deneyim sahibidir. Bu nedenle değerlendiricilerin deneyim sürelerinin yaptıkları değerlendirmelerde bir etkisi olup olmadığını görebilmek için Mann-Whitney U testi uygulanmış ve bu test sonucundaki verilere aşağıdaki tabloda yer verilmiştir:

Çizelge 7. Deneyim Süresinin Değerlendirme Sonuçlarına Etkisi: A1 Düzeyi

Konuşma Performansları	U	P
A1.1	43.50	.877
A1.2	39.50	.643
A1.3	40.50	.699
A1.4	51.50	.643
A1.5	46.50	.938
A1.6	46.00	1.00

Değerlendirmeye katılan katılımcıların deneyim süresinin yaptıkları değerlendirme sonuçlarına etkisini görebilmek için performanslara verdikleri puanlar Mann-Whitney U testi ile incelenmiştir. Test sonucunda p değerleri .05'ten büyük olduğundan deneyim sürelerinin değerlendiricilerin yaptıkları değerlendirme sonuçlarını etkileyip etkilemediğine dair istatistiki olarak anlamlı bir sonuç elde edilememiştir.

Çizelge 8. Deneyim Süresinin Değerlendirme Sonuçlarına Etkisi: A2 Düzeyi

Konuşma Performansları	U	P
------------------------	---	---

A2.1	44.00	.938
A2.2	65.50	.115
A2.3	42.00	.817
A2.4	50.00	.757
A2.5	46.50	.938
A2.6	50.50	.699

A2 düzeyi değerlendirme sonuçlarına bakıldığında da değerlendiricilerin performans değerlendirmelerinin Mann-Whitney U testi sonucunda p değerleri .05'ten büyük olduğundan deneyim sürelerinin değerlendiricilerin yaptıkları değerlendirme sonuçlarını etkileyip etkilemediğine dair istatistiki olarak anlamlı bir sonuç elde edilemediği görülmüştür.

Sonuç ve Öneriler

Bu çalışmada yabancı dil olarak Türkçe öğretiminde temel düzeyler olan A1 ve A2 düzeylerinde öğrenim gören öğrencilerin kur sonu sınavlarında değerlendirilmelerinde kullanılabilecek genel içerikli iki dereceli puanlama anahtarı geliştirilmiştir. Bu amaçla dereceli puanlama anahtarları oluşturulurken nitel ve nicel birçok adım uygulanmış ve oluşturulan dereceli puanlama anahtarlarının sınanması için gerekli geçerlik ve güvenilirlik testlerine başvurulmuştur. Araştırmada A1 ve A2 düzeyi konuşma becerisini değerlendirmek için oluşturulan dereceli puanlama anahtarlarının geçerlik testleri en az 10 yıl deneyimli 8 uzman tarafından gerçekleştirilmiştir. Elde edilen verilerle hem maddeler düzeyinde hem de puanlama anahtarları genelinde geçerli olduğu bulunmuştur.

Güvenirlik testleri ise 20 değerlendiricinin, 6 A1 ve 6 A2 düzeyi öğrencinin konuşma performans kaydını izleyerek yaptığı değerlendirmeler sonucunda elde edilmiştir. Kendall-W testi ile kişiler arası uyum, her iki puanlama anahtarında .80 üzeri olmasından dolayı yüksek derecede olduğu belirlenmiştir. Test sonuçları hem A1 hem de A2 düzeylerindeki dereceli puanlama anahtarları için değerlendiricilerin notlandırmalarında yüksek uyum olduğunu göstermiştir. 20 değerlendirici arasında hem A1 hem de A2 düzeyinde görülen yüksek uyum, dereceli puanlama anahtarlarının güvenilir olduğunu göstermiştir. Bununla birlikte öğreticilerin deneyim sürelerinin değerlendirmelerinde bir etkisi olup olmadığının tespit edilebilmesi için Mann-Whitney U analizi yapılmış ve istatistiki olarak anlamlı bir sonuç elde edilememiştir. Değerlendiricilerin deneyim süreleri ve yaptıkları değerlendirme sonuçları arasındaki ilişkiye dair elde edilen sonuçlar Boylu'nun (2019) çalışmasındaki sonuçlar ile benzerlik göstermektedir. Çalışma sonucunda elde edilen verilerle, A1 ve A2 düzeyi konuşma becerisi dereceli puanlama anahtarlarının geçerli ve güvenilir oldukları söylenebilir.

Bu çalışmada A1 ve A2 düzeyinde kur sonunda kullanılabilecek iki adet dereceli puanlama anahtarı önerisi geliştirilmesi amaçlanmıştır. Ölçme ve değerlendirme süreci dil öğrenim sürecinin her aşamasında yer almaktadır. Bu nedenle tanılayıcı ve biçimlendirici ölçme değerlendirme araçlarıyla birlikte kullanılabilecek puanlama anahtarlarının geliştirilmesinin önemli olacağı düşünülmektedir. Genel puanlama anahtarları haricinde sınıf içi uygulamalar için görev ve tür özelinde de puanlama anahtarlarının oluşturulması öğrencilere hızlı, kolay ve ayrıntılı dönütlerin verilebilmesi açısından önem teşkil

etmektedir. Ayrıca temel düzeylerle uyumlu biçimde diğer düzeyler için puanlama anahtarlarının oluşturulması, konuşma sınavlarında standart bir sistemin uygulanması açısından yararlı olabileceği düşünülmektedir. Yabancı dil öğrenim ortamlarının sadece sınıf olmaması, günümüzde sanal ortamlarda da gerçekleşmesi nedeniyle ölçme ve değerlendirme çalışmaları dijital olarak da yapılabilir. Bu nedenle değerlendirme araçlarının dijital ortama entegre edilebilmesinin gerekli olduğu düşünülmektedir. Dereceli puanlama anahtarlarının yansız ve doğru kullanımının sağlanması, değerlendirmenin sağlıklı yapılabilmesi için önemlidir. Bu nedenle değerlendiriciler için eğitici materyaller hazırlanabilir ve yüz yüze veya çevrimiçi eğitimler gerçekleştirilebilir.

Kaynakça

- Avrupa Konseyi. (2013). *Diller için Avrupa ortak öneriler çerçevesi: Öğrenim, öğretim ve değerlendirme* (2. baskı). TELC GmbH.
- Ayverdi, T. (2023). *Yabancı dil olarak Türkçe öğretiminde C1 düzeyi konuşma becerisi analitik dereceli puanlama anahtarı geliştirilmesi*. [Yayınlanmamış yüksek lisans tezi]. Gazi Üniversitesi.
- Barkaoui, K. (2010). Explaining ESL essay holistic scores: A multilevel modeling approach. *Language Testing*, 27(4), 515–535. <https://doi.org/10.1177/0265532210368717>
- Berger, A. (2015). *Validating analytic rating scales*. Peter Lang. <https://doi.org/10.3726/978-3-653-06183-3>
- Boylu, E. (2019). *Yabancılar Türkçe öğretiminde ölçme değerlendirme uygulamaları ve standart oluşturma*. [Yayınlanmamış doktora tezi]. Çanakkale Onsekiz Mart Üniversitesi.
- Boylu, E., & Başar, U. (2016). Türkçe öğretim merkezlerinin güncel durumu ve standartlaştırılması üzerine. *The Journal of Academic Social Sciences*, 24(24), 309–309. <https://doi.org/10.16992/asos.1042>
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD.
- Büyüköztürk, Ş., Çokluk, Ö., & Köklü, N. (2020). *Sosyal bilimler için istatistik* (24. baskı). Ankara: Pegem Akademi.
- Cambridge. (2011). *Level-A2 assessing speaking performance*. <https://static.cambridge.org/content/id/urn:cambridge.org:id:article:S1366728920000747/resource/name/S1366728920000747sup001.pdf#page=1.00>
- Çerçi, A. (2018). Türkçenin yabancı dil olarak öğretiminde konuşma becerisini ölçme değerlendirmede karşılaşılan sorunlar üzerine yabancılar Türkçe öğretmenlerin görüşleri. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, 31, 3–16. <https://doi.org/10.19171/uefad.430138>
- Creswell, J. W. (2021). *Karma yöntem araştırmalarına giriş* (M. Sözbilir (ed.); 3. baskı). Pegem Akademi. <https://doi.org/10.14527/9786053184720>
- Davis, L. (2018). Analytic, holistic, and primary trait marking scales. *The TESOL Encyclopedia of English Language Teaching* içinde (ss. 1–6). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118784235.eelt0365>
- Deygers, B., Van Gorp, K., Luyten, L., & Joos, S. (2013). Rating scale design: A comparative study of two analytic rating scales in a task-based test. E. Galaczi & C. Weir (Ed.), *Exploring language frameworks. Proceedings from the ALTE Kraków conference* içinde (ss. 273–289). Cambridge University Press. https://www.alte.org/resources/Documents/SILT_36_FOR_PRINT.pdf

- Fulcher, G., & Davidson, F. (2012). The Routledge handbook of language testing. G. Fulcher & F. Davidson (Ed.), *The Routledge handbook of language testing* içinde. Taylor & Francis Routledge.
- Fulcher, G., Davidson, F., & Kemp, J. (2011). Effective rating scale development for speaking tests: Performance decision trees. *Language Testing*, 28(1), 5–29. <https://doi.org/10.1177/0265532209359514>
- Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology*, 15(4), 380–387. <https://doi.org/10.1037/a0025704>
- Gwet, K. L. (2020). *Her yönüyle puanlayıcılar arası güvenilirlik rehberi* (İ. Karakaya & H. Yürekli (ed.); F. Elkonca, G. Ceylan, İ. Karakaya, H. Yürekli, & E. C. Çorbacı (çev.)). Pegem Akademi.
- Harsch, C., & Martin, G. (2013). Comparing holistic and analytic scoring methods: Issues of validity and reliability. *Assessment in Education: Principles, Policy and Practice*, 20(3), 281–307. <https://doi.org/10.1080/0969594X.2012.742422>
- Jeong, H. (2015). Rubrics in the classroom: do teachers really follow them? *Language Testing in Asia*, 5(1). <https://doi.org/10.1186/s40468-015-0013-5>
- Kocayanak, D. (2021). *Yabancı dil olarak Türkçe öğretmenlerin konuşma becerisini ölçme sürecine yönelik uygulamaları*. [Yayınlanmamış doktora tezi]. Hacettepe Üniversitesi.
- Kocayanak, D., & Tecelli, N. T. (2023). Yabancı dil olarak türkçe öğretiminde konuşma becerisini değerlendirme ölçütleri. *Hacettepe Üniversitesi Türkiyat Araştırmaları Dergisi*, 38, 245–267. <https://dergipark.org.tr/en/pub/turkiyat/issue/77208/1267689>
- Köse, D. (2008). Türkçenin yabancı dil olarak öğretiminde sınavların hazırlanması ve uygulanması. *Dil Dergisi*, 36–47. https://doi.org/10.1501/dilder_00000000089
- Lawshe, C. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Metruk, R. (2018). Ethical issues in the assessment of speaking skills. *Communications - Scientific Letters of the University of Žilina*, 20(1PartA), 109–114. <https://doi.org/10.26552/com.c.2018.1a.101-114>
- O'Sullivan, B. (2015). *Technical report: Aptis test development approach*. http://www.britishcouncil.org/sites/britishcouncil.uk2/files/tech_001_barry_osullivan_aptis_test_-_v5.pdf
- O'Sullivan, B., & Dunlea, J. (2019). *Appendix H: Rating scales for speaking and writing speaking task 1*. https://aptisonline.com/wp-content/uploads/2019/05/Speaking-and-Writing-Rating-Scales-aptis_general_technical_manual_v-1.0.pdf
- Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques. *Decision Sciences*, 28(3), 763–774. <https://doi.org/10.1111/j.1540-5915.1997.tb01330.x>
- Şimşek, A. (2022). *Yabancı dil olarak Türkçe öğretiminde konuşma becerisine yönelik biçimlendirici değerlendirme temelli dereceli puanlama anahtarı geliştirme çalışması (B2 ve C1 düzeyi için)*. [Yayınlanmamış yüksek lisans tezi]. Kocaeli Üniversitesi.
- Stevens, D. D., & Levi, A. J. (2005). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback and promote student learning*. Stylus Publishing.
- Yıldırım, A., & Şimşek, H. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri* (9. baskı). Seçkin Yayıncılık.
- Yorgancı, O. K. (2021). *Yabancı dil olarak Türkçe öğretiminde podcast (sesli yayın) kullanımının üretici dil becerilerine etkisi*. [Yayınlanmamış doktora tezi]. Yıldız Teknik Üniversitesi.

Ekler

EK A. A1 Düzeyi Analitik Konuşma Puanlama Anahtarı

		MADDELER	Çok iyi (4)	İyi (3)	Geliştiri lmesi (2)	Yeter siz (1)
SÖZCÜK BİLGİSİ	1.	Kendini tanıtırken adı, yaşı, ailesi, ülkesi gibi konularda sözcük bilgisine sahip ve doğru kullanabiliyor.				
	2.	Tanışma, selamlaşma sözcük bilgisine sahip ve doğru kullanabiliyor.				
	3.	Alışveriş gibi günlük basit soru cevap gerektiren konularla ilgili sözcük bilgisine sahip ve doğru kullanabiliyor.				
	4.	Günlük yaşam (bir gününü anlat gibi) hakkında sözcük bilgisine sahip, sözcükleri doğru ve yerinde kullanabiliyor. (Sabah uyanıyorum, okula gidiyorum, ders çalışıyorum vb.)				
	5.	Konuşucunun değerlendirici tarafından verilen konu için sözcük seçimi genellikle doğru.				
DİLBİLGİSEL DOĞRULUK	6.	A1 düzeyi dilbilgisi yapılarını doğru biçimde kullanabiliyor. (çoğul eki, soru eki, durum ekleri, -yor, -DI zaman ekleri, iyelik vb.)				
	7.	Kendini tanıtırken dilbilgisi yapılarını ve kalıp ifadeleri doğru biçimde kullanabiliyor. (Benim adım..., ben... yaşıyorum, '-DAn geliyorum, benim annem, babam... vb.)				
	8.	Günlük yaşamı hakkındaki dilbilgisi yapılarını ve kalıp ifadeleri doğru biçimde kullanabiliyor. (Sabah uyanıyorum, okula gidiyorum, ders çalışıyorum, futbol oynamıyorum vb.)				
	9.	Tanışma, selamlaşma, alışveriş gibi karşılıklı soru cevap gerektiren konularda dilbilgisi yapılarını ve kalıp ifadeleri doğru kullanabiliyor. (Nasılsın? Nereden geliyorsun? Bu ne kadar? vb.)				
	10.	Konuşmada zaman, mekan, durum gibi zamana ilişkin kalıp ifadeler ve dilbilgisi yapılarını doğru kullanabiliyor.				
TUTARLIK VE BAĞDAŞIKLIK	11.	Konuyu anlatırken sıralamada ara sıra hatalar görülse de basit ve doğru bir sıralamayla anlatabiliyor.				
	12.	A1 düzeyine uygun bağlaçları kullanabiliyor ve cümleleri birbirine doğru şekilde bağlayabiliyor. (ve, ya da, veya, ama, önce, sonra vb.)				
	13.	Konuşurken başka konulara sapmadan beklenen yanıtı verebiliyor.				
AKICILIK	14.	A1 düzeyine uygun bir akıcılıkla konuşabiliyor. Tanıdık/bilgisinin olduğu konularda akıcılıkta problemi yok.				
	15.	Konuşma içerisinde düşünmek ve düzeltme yapmak için ara sıra duraklama yapsa da konuşmasını yönetebiliyor.				
SESLETİM	16.	A1 düzeyine uygun ifadeleri az ya da sıfır hatayla sesletebiliyor.				
	17.	A1 düzeyine göre bazı hatalar görülse de konuşma sırasında doğru tonlama, vurgu yapabiliyor.				
KARŞILIKLI ETKİLEŞİM	18.	A1 düzeyine uygun basit konuşmaları takip edebiliyor.				
	19.	Kısa soru cevap şeklindeki konuşmalarda konuşmayı takip edebiliyor.				
	20.	Kısa soru cevap şeklindeki konuşmalarda konuşmada söz alabiliyor.				
	21.	A1 düzeyinde aşına olduğu konularda etkileşime girebiliyor.				
	22.	A1 düzeyinde kendi ülkesi, ailesi, okulu gibi konularda basit bir karşılıklı konuşmaya katılabiliyor.				
	23.	Basit selamlaşma, izin alma, vedalaşma, teşekkür etme gibi kalıplaşmış ifadeleri karşılıklı konuşma içerisinde doğru biçimde kullanabiliyor.				
GÖREVİ BAŞARMA	24.	Değerlendirici tarafından verilen görevi doğru biçimde tamamlayabiliyor.				

	25.	Kendisine yöneltilen basit yönergeleri anlayabiliyor.				
--	-----	---	--	--	--	--

EK B. A2 Düzeyi Analitik Konuşma Puanlama Anahtarı

☐Yönergeleri değerlendiricinin yardımına rağmen anlamamış ve bağlama uygun olmayan içerikte konuşmuştur.

☐Konuşmasında temel dil bilgisi yapıları ve sözcüklerde çok fazla hata görüldüğünden konuşması anlaşılmamaktadır.

☐Konuşma uzunluğu yeterli olmadığından değerlendirme yapılamamıştır.

*Yukarıdaki 3 kutuyu da işaretleyen öğretici konuşmayı değerlendirmeye almaz ve “Başarısız” olarak belirtilir.

		MADDELER	Çok iyi (4)	İyi (3)	Geliştirelmesi (2)	Yetersiz (1)
SÖZCÜK BİLGİSİ	1.	Kendini, başka birisini, bir yeri, eşyayı tarif ederken bazen yanlış kullanımlar görülse de kullanabileceği sözcük bilgisine sahip.				
	2.	Yol tarifi gibi yer yön gerektiren konularla ilgili sözcük bilgisine sahip ve doğru kullanabiliyor.				
	3.	Arkadaşları, ülkesi, hobileri, gelecek planları gibi verilen konuyla ilgili sözcük bilgisine sahip ve doğru kullanabiliyor.				
	4.	Değerlendiricinin verdiği konu için sözcük seçimi çoğunlukla doğru.				
DİLBİLGİSEL DOĞRULUK	5.	A2 düzeyi dilbilgisi yapılarını doğru biçimde kullanabiliyor. (emir, istek kipleri, -DI, -AcAK, -mı, -(I/A)r, -(y)Abil vb.)				
	6.	Kendini veya başka birisini, bir yeri/eşyayı tarif ederken dilbilgisi yapılarını ve kalıp ifadeleri doğru biçimde kullanabiliyor. (Ben uzunum. O üzgün bir kız. Arkadaşım kısa ve şişman vb.)				
	7.	Yol tarifi gibi yer yön tarifi gerektiren konularla ilgili dilbilgisi yapılarını ve kalıp ifadeleri doğru biçimde kullanabiliyor (Sağa dön. Parkı geç. Dümdüz yürü. Pastane sağda, vb.)				
	8.	Arkadaşları, ülkesi, hobileri, gelecek planları gibi verilen konuyla ilgili konularda dilbilgisi yapılarını ve kalıp ifadeleri doğru kullanabiliyor.				
	9.	A2 düzeyinde kullanması beklenen dilbilgisi yapılarında ve kalıp ifadelerinde hata görülüyor.				
	10.	Kendini A2 düzeyine göre doğru ifade edebiliyor.				
TUTARLIK VE BAĞDAŞIKLIK	11.	Verilen konuyu basit ve doğru bir sıralamayla konudan fazla sapmadan anlatabiliyor.				
	12.	A2 düzeyine uygun bağlaçları kullanabiliyor ve cümleleri birbirine doğru şekilde bağlayabiliyor. (ile, çünkü, bu sebeple, bu nedenle, bunun için, hem hem de, ne ne de, ya ya vb.)				
	13.	“Gibi” veya “örneğin” vb. düzeyine uygun bağlayıcı sözcükleri kullanarak basit örnekler verebiliyor.				
AKICILIK	14.	A2 düzeyine uygun bir akıcılıkla konuşabiliyor. Aşına olduğu konularda akıcılıkta problemi yok.				
	15.	Düşünmek ve düzeltme yapmak için ara sıra duraklasa da konuşmasını yönetebiliyor.				
	16.	Kısa konuşmaları idare etmek için duraklama ve düzeltmeler görülse de aşına olduğu konularda öbekler ve tümceler kurabiliyor.				
SESLETİM	17.	A2 düzeyine uygun ifadeleri az ya da sıfır hatayla sesletebiliyor.				
	18.	A2 düzeyine göre bazı hatalar görülse de konuşma sırasında doğru tonlama, vurgu yapabiliyor.				
	19.	Genel olarak sesletimde hatalar görülse de anlaşılabilirliğini etkilemiyor.				
KARŞILIKLI ETKİLEŞİM	20.	Basit günlük karşılıklı konuşmalarda kendisine söyleneni anlayabiliyor.				
	21.	Karşı tarafı anladığını belirtebiliyor.				
	22.	Karşılıklı konuşmada anladığını belirtip gerekli yanıt verebiliyor.				
	23.	Karşılıklı konuşmayı başlatıp, devam ettirip, bitirebiliyor.				

GÖREVİ BAŞARMA	24.	Değerlendirici tarafından verilen görevi doğru biçimde tamamlayabiliyor.				
	25.	Yönergeleri doğru biçimde anlayabiliyor.				

EXTENDED ABSTRACT

From the very beginning, evaluation and assessment are crucial stages in the process of language teaching. In exams designed to assess the productive skill of speaking, learners demonstrate their oral performance. Graded scoring rubrics can be utilized to measure and facilitate the assessment of a learner's performance in such exams. Scoring rubrics are typically created in two forms: holistic and analytical. Holistic scoring rubrics evaluate the individual's performance as a whole, while analytical scoring rubrics allow for a more detailed assessment based on specific dimensions and criteria.

It is important to conduct foreign language learning processes with a certain standard in mind. The use of scoring rubrics that facilitate objective behavior by evaluators and establish a standard across institutions is crucial. In the field of Turkish as a foreign language, studies for beginner levels A1 and A2 are limited. Therefore, this study is a significant step towards preparing and using valid and reliable assessment tools for A1 and A2 levels.

This study aims to provide unbiased results at the basic level in the field of teaching Turkish as a foreign language. To this end, analytical rubrics will be created for the A1 and A2 levels, which correspond to the beginner levels. The objective is to develop analytical rubrics for both levels, thereby enhancing evaluators' capacity to conduct expedient, straightforward, and transparent assessments.

In order to develop rubrics within the scope of the research, both qualitative and quantitative data must be collected. Therefore, an exploratory sequentially patterned mixed method was applied in the study since it is compatible with the steps of creating rubrics. In order to create the rubric, the target group and the competencies and achievements needed for the levels aimed at being evaluated were first determined. Then, the main dimensions and criteria were determined according to the specified levels and competencies. A literature review was used to make these determinations. Expert opinions were obtained in order to create draft rubrics, and content validity studies were conducted using the Lawshe technique. Afterwards, the speaking performance records of the A1-A2 level learners were evaluated with the draft rubrics prepared by the experienced and inexperienced instructor groups, and quantitative data was collected for reliability. The interrater reliability was examined using the Kendall's W test in order to ascertain the extent of agreement between the raters.

Ensuring the impartial and correct use of rubrics is important for a healthy assessment. Therefore, evaluators should be trained about the rubrics they use. Guidelines should be created for the assessments to be healthy. In the study, the validity tests of the rubrics designed to evaluate A1 and A2 level speaking skills were carried out by eight experts with ten years of experience. The obtained data were found to be valid both at the

item level and in the overall rubric. Reliability tests were obtained as a result of the assessments made by 20 instructors by monitoring six A1 and six A2 speaking performance records. It was determined that interpersonal agreement was high in both rubrics with Kendall's W. However, in order to determine whether the instructor's experience duration had an effect on their assessments, a Mann-Whitney U analysis was performed, and no statistically significant result was obtained. In light of these results, it can be concluded that the A1 and A2 level speaking skill rubrics created are valid and reliable.