

ÇOKTAN SEÇMELİ TESTLERİN GÜVENİRLİK ANALİZİ: WEB UYGULAMASI GELİŞTİRME VE KULLANILABİLİRLİK DEĞERLENDİRMESİ

Mehmet ÇAKMAKTEPE¹

Gür Emre GÜRAKSIN²

Muharrem Kürşat YALIM³

Öz

Bu çalışma, çoktan seçmeli testlerin güvenilirlik analizini gerçekleştiren bir çevrimiçi platformun kullanılabilirliğini değerlendirmeyi amaçlamaktadır. Araştırma kapsamında, kendi geliştirdiğimiz istatistiksel parametreleri analiz eden bir web sitesi olan 'sinavanalizlerim.com' incelenmiş ve kullanıcı deneyimi, dört alt boyuttan oluşan Web Sitesi Kullanılabilirlik Ölçeği temel alınarak ölçülmüştür. Örneklem grubunu oluşturan 50 akademisyen, web sitesinin gezinme kolaylığı, tasarım tutarlılığı, erişim hızı ve kullanım kolaylığı olmak üzere dört alt boyutta değerlendirilmiştir. Elde edilen sonuçlar, web sitesinin tasarım tutarlılığı, menü ve içerik uyumu ile hızlı erişim özellikleri açısından kullanıcılar tarafından olumlu karşılandığını göstermektedir. Bununla birlikte, platformun etkinliğini artırmak için konu temelli soru eşleştirme, yapay zekâ destekli geri bildirim mekanizmaları ve notlandırma entegrasyonunun eklenmesi önerilmektedir. Bu tür iyileştirmelerin, ölçme ve değerlendirme süreçlerinin kalitesini artırarak eğitim alanında daha etkin bir kullanım sağlayabileceği öngörülmektedir. Araştırmanın metodolojisi çerçevesinde, veri toplama yöntemi olarak anket kullanılmış ve elde edilen veriler SPSS programı ile analiz edilmiştir. Web sitesi kullanılabilirliği açısından faktör analizleri gerçekleştirilmiş ve Cronbach's Alpha katsayısı 0,885 olarak hesaplanmıştır, bu da ölçeğin güvenilir olduğunu göstermektedir. Sonuç olarak, 'sinavanalizlerim.com', çoktan seçmeli testlerin değerlendirilmesi konusunda çok etkili bir dijital araç olarak öne çıkmaktadır. Ancak, sistemin daha kullanıcı dostu hale getirilmesi için yukarıda belirtilen teknolojik iyileştirmeler önerilmektedir. Bu çalışma, eğitimde dijital araçların etkin kullanımı konusunda önemli bir referans kaynağı olarak değerlendirilmektedir.

Anahtar Kelimeler

Çoktan Seçmeli Testler
Güvenirlik Analizi
Kullanılabilirlik
Eğitimde Ölçme ve Değerlendirme
Web Platformları

Makale Hakkında

Araştırma Makalesi

Gönderim Tarihi : 19.02.2025
Kabul Tarihi : 06.04.2025
E-Yayın Tarihi : 15.06.2025
DOI : 10.58702/teyd.1642933

¹ Yüksek Lisans Öğrencisi., Afyon Kocatepe Üniversitesi, Fen Bilimleri Enstitüsü, İnternet ve Bilişim Teknolojileri Yönetimi Bölümü, e-posta: cakmaktepemehmet@gmail.com, ORCID: 0009-0005-6264-1959.

² Doç. Dr., Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, e-posta: emreguraksin@aku.edu.tr, ORCID: 0000-0002-1935-2781.

³ Doktora Öğrencisi, İstanbul Topkapı Üniversitesi, Lisansüstü Eğitim Enstitüsü, Yönetim Bilişim Sistemleri Bölümü, e-posta: nebiseren@uludag.edu.tr@gmail.com, ORCID: 0000-0003-4080-4823.

RELIABILITY ANALYSIS OF MULTIPLE-CHOICE TESTS: DEVELOPMENT OF A WEB APPLICATION AND USABILITY EVALUATION

Abstract

This study aims to evaluate the usability of an online platform designed for conducting reliability analysis of multiple-choice tests. Within the scope of the research, a self-developed website, *sinavanalizlerim.com*’, which performs statistical analyses based on specific parameters, was examined. User experience was assessed using the Website Usability Scale, which comprises four sub-dimensions: ease of navigation, design consistency, access speed, and overall usability. A sample group of 50 academics evaluated the platform based on these dimensions. The findings indicate that the website was positively received, particularly in terms of design consistency, alignment between menus and content, and fast access features. However, to enhance the platform’s effectiveness, it is recommended to incorporate features such as topic-based question matching, AI-supported feedback mechanisms, and integrated grading systems. These improvements are expected to enhance the quality of assessment processes and promote more effective use in educational settings. Methodologically, data were collected through a survey and analyzed using SPSS software. Factor analysis was conducted in the context of website usability, and the Cronbach’s Alpha coefficient was calculated as 0.885, indicating a high level of reliability for the scale. In conclusion, *sinavanalizlerim.com*’ stands out as a highly effective digital tool for evaluating multiple-choice tests. Nonetheless, the incorporation of the aforementioned technological enhancements is suggested to further improve user experience. This study serves as a significant reference for the effective integration of digital tools in educational assessment.

Keywords

Multiple-Choice
Tests Reliability
Analysis Usability
Educational Assessment
Web Platforms

Article Info

Research Article

Received : 19.02.2025
Accepted : 06.04.2025
Online Published : 15.06.2025
DOI : 10.58702/teyd.1642933

Kaynakça Gösterimi: Çakmaktepe, M., Güraksın, G. E. ve Yalim, M. K. (2025). Çoktan seçmeli testlerin güvenilirlik analizi: web uygulaması geliştirme ve kullanılabilirlik değerlendirme. *Toplum, Ekonomi ve Yönetim Dergisi*, 6 (2), 252-277

Citation Information: Cakmaktepe, M., Guraksin, G. E. ve Yalim, M. K. (2025). Reliability analysis of multiple-choice tests: development of a web application and usability evaluation. *Journal of Society, Economics and Management*, 6 (2), 252-277.

GİRİŞ

Eğitim, bireyin bilinçli ve planlı bir süreçle istenen davranış değişikliklerini kazanmasını hedefler ve insanlık tarihinin en temel amaçlarından biridir. Günümüzde giderek karmaşılaşan bu olgu, bireylerin tutarlı davranışlar ve etkili problem çözme becerileriyle donatılmasını zorunlu kılmaktadır. Devlet yönetiminden sanayiye, doğal kaynakların işlenmesinden toplumsal sorunların çözümüne kadar birçok alandaki sorumluluk, bireylerin eğitim düzeyine dayanmaktadır. Bu nedenle, hedeflenen toplumsal yapıya ulaşmak ancak sistemli ve nitelikli bir eğitim süreci ile mümkün olabilir (Çakmak, 2008, s. 34). Bu çalışma çoktan seçmeli ölçme araçlarını değerlendiren bir sitenin kullanılabilirliğini değerlendirmeyi amaçlamaktadır.

Ölçme ve değerlendirme, öğrenmenin kalitesini ve kalıcılığını artırmada kritik bir role sahiptir. Ölçme, bir niteliğin gözlemlenip sayılar veya sembollerle ifade edilmesi; değerlendirme ise ölçme sonuçlarının belirli ölçütlere dayalı olarak bir değer yargısına ulaşılması sürecidir. Eğitim süreçlerinde bu uygulamalar, öğrencilerin kazanımlara ulaşma düzeylerini belirlemeye, süreçteki sorunları tespit ederek gerekli düzenlemeleri yapmaya ve eğitim programının etkinliğini artırmaya olanak tanır. Böylelikle eğitim süreci sürekli izlenebilir ve uygun dönütlerle iyileştirmeler sağlanarak kalite kontrolü mümkün hale gelir (Benzer ve Eldem, 2013, s. 650).

Çoktan seçmeli testler ise birden fazla seçenek arasından yalnızca bir doğru yanıtın belirlenmesine dayalı maddelerden oluşan yaygın bir ölçme aracıdır. Eğitim sisteminin çeşitli düzeylerinde, sınıf içi değerlendirmelerden ulusal ve uluslararası sınavlara kadar geniş bir uygulama alanına sahiptir. Bu testler, objektif puanlama imkânı sunmasıyla, öğrencilerin bilgi düzeylerini, yeteneklerini ve başarılarını etkili, güvenilir ve pratik bir şekilde değerlendirmeye olanak tanımaktadır (Yöntem, 2022, s. 121-122).

Çoktan seçmeli test maddeleri, sorunun ana bileşeni olan madde kökü ve cevap seçeneklerinden oluşur. Madde kökü, sorunun temel bilgilerini içeren ve genellikle en uzun bölüm olup bir soru cümlesi veya tamamlanması gereken bir ifade şeklinde yapılandırılabilir. Doğru cevabın yanı sıra yanlış cevapları temsil eden seçenekler, çeldirici olarak adlandırılır. Çeldiriciler, öğrenme hedefini gerçekleştiremeyen öğrenciler için makul görünürken, hedefe ulaşan öğrenciler tarafından mantıksız bulunarak elenmektedir (Yöntem, 2022, s. 121-122).

Bir çoktan seçmeli ölçme aracının kapsamlı bir şekilde analiz edilebilmesi için çoktan seçmeli ölçme aracının ortalaması, güvenilirlik katsayısı, ortalama güçlük düzeyi, standart sapması, varyansı, merkezi eğilim ve dağılım ölçüleri gibi temel istatistiksel değerlerin hesaplanması gereklidir. Benzer şekilde, maddelerin özelliklerini değerlendirmek amacıyla madde güçlük indeksi, madde ayrıricılık indeksi, madde varyansı, standart sapması, madde güvenilirlik katsayısı ve madde-kriter korelasyonu gibi ayrıntılı istatistiksel analizlerin gerçekleştirilmesi önem taşır. Bu analizler hem çoktan seçmeli ölçme aracının genel yapısının hem de maddelerin ölçme yeterliliğinin nesnel bir biçimde değerlendirilmesine olanak sağlar (Aydın, 2013, s. 2). Bu çalışma da söz konusu olan site ölçme kavramlarının (test güvenilirlik katsayıları, madde güçlük ve ayrıricılık indeksleri gibi) analizlerini desteklemektedir.

Bu makalenin amacı, tarafımızca tasarlanan ve belirli istatistiksel hesaplamaları gerçekleştirebilen bir web platformunun kullanılabilirliğini değerlendirmektir. Bu

değerlendirme, dört alt boyuttan oluşan ve 5'li Likert Ölçeği temel alınarak geliştirilen ve 'Web Sitesi Kullanılabilirlik Ölçeği' olarak adlandırılan bir anket aracı kullanılarak gerçekleştirilmiştir. Platformun sunduğu istatistiksel analizler, eğitim ve öğretim alanında, öğretmenlerin sınavlarının öğrencilere uygulanabilirliğini belirlemeleri açısından büyük bir önem taşımaktadır. Yapılan incelemeler sonucu bu hizmeti türkçe veren bir web sitesine rastlanılmamıştır. Günümüzde bu tür hesaplamaların elle yapılması veya yabancı kaynaklardan yararlanılmaya çalışılması süreci oldukça zorlaştırmaktadır. Web platformunun geliştirilmesindeki temel amaçlardan biri, bu zorlukları ortadan kaldırmaktır.

1. Araştırmanın Metodolojisi

1.1. Araştırmanın Konusu, Amacı ve Önemi

Bu çalışma, çoktan seçmeli testlerin güvenilirlik analizini gerçekleştirmek amacıyla tarafımızca geliştirilen 'sinavanalizlerim.com' adlı web sitesinin kullanılabilirlik düzeyini değerlendirmektedir. Bu web platformu, test güvenilirlik katsayıları, madde güçlük ve ayrımcılık indeksleri gibi istatistiksel analiz araçları sunarak eğitimde ölçme ve değerlendirme süreçlerini desteklemeyi amaçlamaktadır. Araştırmada, web sitesinin gezinme kolaylığı, erişim hızı, tasarım tutarlılığı ve kullanım kolaylığı gibi kullanılabilirlik unsurları incelenmiş ve kullanıcı dostu bir arayüz sunup sunmadığı analiz edilmiştir. Elde edilen bulgular, web sitesinin genel olarak çok olumlu değerlendirildiğini, konu temelli soru eşleştirme ve yapay zekâ destekli geri bildirim mekanizmalarının entegrasyonu ile daha verimli hale getirilebileceğini göstermektedir. Çalışma, eğitim teknolojilerinin geliştirilmesinde kullanılabilirlik faktörlerinin önemine dikkat çekmekte ve bu alandaki uygulamalara katkı sağlamayı hedeflemektedir.

1.2. Araştırmayla İlgili Literatür İncelemesi

Çoktan seçmeli testlerin güvenilirliğini etkileyen temel unsurlardan biri, testin içerik geçerliliğidir. Ercan ve Kan (2004) ile Seale (2004) gibi araştırmacılar, geçerlilik ve güvenilirlik düzeyi yüksek olan testlerin değerlendirme açısından daha uygun araçlar olduğunu vurgulamaktadır. Benzer şekilde, Yaman (2016) da çoktan seçmeli testlerin ölçme ve değerlendirme süreçlerinde en yaygın kullanılan yöntemlerden biri olduğunu ifade etmektedir (Boz vd., 2022, s. 150). Bununla birlikte, öğrenci grubunun büyüklüğü ile testte yer alan madde sayısının da puanlama güvenilirliği üzerinde belirleyici etkiler yarattığı belirtilmektedir. Nitekim, çoktan seçmeli testlerin geniş ölçekli öğrenci gruplarına uygulanması ve daha fazla sayıda madde içermesi, testin hem geçerlilik düzeyini hem de değerlendirme kalitesini artırmaktadır (Birgin ve Özcan, 2022, s. 322).

Rehman ve arkadaşlarının (2018, s. 291-293) yapmış oldukları çalışma, diş hekimliği eğitimi bağlamında yalnızca çoktan seçmeli soruların kalite analizine odaklanmıştır. Çoktan seçmeli sorular, tıp ve diş hekimliği öğrencilerinin değerlendirilmesinde yaygın olarak kullanılan bir yöntemdir. Ancak, düşük kaliteli soruların kullanımı, objektif ve doğru bir değerlendirme yapılmasını engelleyebilmektedir. Çalışmada, çoktan seçmeli testlerin kalitesini belirlemek için kullanılan üç temel ölçüt detaylı olarak incelenmiştir: zorluk indeksi, ayırım indeksi ve yanıtlayıcı etkinliği. Burada bahsi geçen kavramların tümü yapılan web sitesinde bulunmaktadır. Çalışma sonuçları, kullanılan soruların kalite açısından iyileştirilmesi gerektiğini ve gelecekteki değerlendirme süreçlerinin daha özenli bir şekilde

düzenlenmesinin önemini vurgulamaktadır. Eğitim ve değerlendirme süreçlerindeki bu tür iyileştirmeler, genel olarak sağlık eğitimi alanında daha yüksek bir standartın yakalanmasına katkı sağlayacaktır. Bu çalışmada incelenen web sitesi yalnızca çoktan seçmeli soruları analiz edebilmektedir.

Tavakol ve Dennick'in (2011 , s. 447-458) çalışması, tıp eğitimi değerlendirmelerinde objektif test sonuçlarının analizine odaklanmaktadır. Çalışmada, değerlendirme süreçlerinde meydana gelebilecek hataların en aza indirilmesi ve öğrencinin 'gerçek' performansını en doğru şekilde yansıtacak sonuçlara ulaşılabilmesi için güvenilirlik ve geçerliliği artırma yöntemleri ele alınmıştır. Objektif testlerin analizi, öğrenme hedefleri, test maddelerinin kalitesi ve genel değerlendirme döngüsü ile doğrudan ilişkilidir. Araştırma bulgularına göre, düşük zorluk düzeyine sahip veya ayırt edici nitelik taşımayan sorular, değerlendirme süreçlerinin genel kalitesini olumsuz yönde etkilemekte ve öğrencilerin gerçek bilgi ve beceri düzeylerini yeterli derecede ölçmede başarısız kalmaktadır. İncelemesini yaptığımız web sitesinin eğitimciler tarafından kullanılması, eğitimciler tarafından çoktan seçmeli ölçme araçları kullanarak yaptıkları ölçmeleri analiz ederek daha sağlıklı ölçme araçları geliştirmelerine yardımcı olabilir. Bu bağlamda, öğretim üyelerinin değerlendirme tasarımı ve soru hazırlama süreçleri konusunda yeterli eğitim almaları büyük önem taşımaktadır. Ayrıca, sınav sonrası analizlerin yalnızca sınav sonuçlarını değerlendirmeye katkı sağlamakla kalmayıp, aynı zamanda öğrenme hedefleri ve öğretim yöntemleri gibi eğitim süreçlerinin genel olarak iyileştirilmesine olanak sunduğu vurgulanmıştır.

1.2.1. Ölçme ve Değerlendirme

Bu çalışmanın konusu çoktan seçmeli ölçme araçlarının analizidir ve ölçme, bir özelliği belirleme ve değerlendirme süreci olarak, varlıkların veya olayların niteliklerini ortaya koymayı ve 'Ne kadar?' sorusuna yanıt aramayı amaçlayan sistematik bir süreçtir. Uzunluk, sıcaklık, derinlik, nem, yağış miktarı, nabız, depremin şiddeti, zekâ ve başarı gibi kavramlar, bu nitelikleri tanımlayan unsurlardır. Ölçme işlemi, insanın doğasında bulunan merak duygusunun bir yansıması olmasının yanı sıra, yaşamı daha güvenli ve sağlıklı bir şekilde sürdürebilmek için de kritik bir rol oynar. Varlıkların veya olayların özelliklerini ve bu özelliklere sahip olma derecelerini tespit süreci, genel anlamda 'gözlem' olarak adlandırılabilir. Ancak, bir gözlemin 'ölçme' olarak değerlendirilebilmesi için elde edilen verilerin mutlaka 'sayı' veya 'semboller' ile ifade edilmesi gerekmektedir (Yılmaz, 1996, s. 12).

Ölçme, bir nesnenin ya da nesne grubunun belirli bir özelliğe sahip olup olmadığını ve sahip olduğu durumda bu özelliğin derecesini gözlemleyerek tespit etme, ardından bu gözlem sonuçlarını sembollerle, özellikle sayısal ifadelerle temsil etme süreci olarak tanımlanır. Örneğin, bir bireyin cinsiyetini (kadın-erkek), bir kişinin medeni durumunu (evli, bekâr, dul, boşanmış) belirlemek; bir günün sıcaklığını derece cinsinden ifade etmek ya da bir çocuğun ağırlığını kilogram cinsinden göstermek, ölçme sürecine somut örneklerdir (Tekin, 1979, s. 20). Web sayfasında ölçme sonuçları sayısal ifadelerle literatürde ve eğitim sisteminde geçerli formüllerle ölçülerek kullanıcıya soruların zorluk düzeyini belirleyebildikleri geçerli bir çıktı vermektedir.

Ölçme, doğrudan, dolaylı ve türetilmiş ölçme olmak üzere üç temel kategoride sınıflandırılabilir. Doğrudan ölçme, ölçülen özellik ile ölçüm aracı arasında doğrudan bir

uyumun bulunduğu durumları ifade eder. Örneğin, ağırlığın başka bir ağırlıkla veya uzunluğun başka bir uzunlukla ölçülmesi bu tür ölçmeye örnek teşkil eder. Dolaylı ölçme ise ölçülen özellik ile ölçüm aracı arasında doğrudan bir ilişki bulunmadığında kullanılır ve geçerli bir varsayıma dayanır. Örneğin, civanın yüksekliği sıcaklığı, yayın gerilimi ağırlığı veya başarı testi bireyin başarısını ölçer. Dolaylı ölçmenin doğruluğu, kullanılan varsayımların geçerliliğine bağlıdır ve bu nedenle ölçümün güvenilirliği açısından kritik bir rol oynar (Yılmaz, 1996, s. 42). Türetilmiş ölçme ise, bir değişkenin iki veya daha fazla değişken ve bunlar arasındaki ilişkiler üzerinden tanımlandığı durumlarda gerçekleştirilir. Bu tür ölçümlerde, tanımlamada yer alan değişkenler ve bu değişkenler arasındaki bağıntıyı ifade eden aritmetik işlemler sonucu elde edilen değerler kullanılır. Bu nedenle, türetilmiş ölçmeler karmaşık değişkenlerin değerlendirilmesinde yaygın bir yöntemdir (Atılğan ve ark., 2014, s. 3). Oluşturulan web sayfasının ölçme yöntemi dolaylı ölçmeye dayanmaktadır. İncelenen sorular geçerli bir varsayıma dayanan ölçme istatistik formülleri ile hesaplanmaktadır.

Değerlendirme, karar verme sürecine dayanan bir işlem olup, bu yönüyle ölçmeden ayrılır. Ölçme, mevcut durumu gerçeğe en yakın şekilde betimlemeyi amaçlarken, değerlendirme, ölçme sonuçlarının belirlenmiş bir ölçütle karşılaştırılmasını içerir. Bu süreçte, ölçme sonuçlarının ilgili ölçütle tanımlanan şartları karşılayıp karşılamadığı analiz edilir. Sonuç olarak, değerlendirme, ölçülen özelliğin belirli bir amaca uygunluğunu veya işlevselliğini belirlemek üzere bir karara ulaşılmasını sağlar (Özcelik, 1998, s. 252). Oluşturulan web sayfasında sınavın öğrencilere uygun olup olmadığını belirlemek için hesaplamalar yapılmaktadır ve bu konuyla alakalı olarak bir değerlendirme çıktısı vermektedir.

Ölçüt, ölçme sonuçlarının değerlendirilmesinde temel alınan ve karar verme süreci boyunca sabit kalan belirli standartlar veya referans değerlerdir (Gömlüksiz ve ark., 2008, s. 20). Oluşturulan web sayfasında kullanılan formüller bahsedilen ölçüt kavramına dayanmaktadır. Formüller sayesinde ölçme sonuçlarını değerlendirebilmekteyiz.

Karar değerlendirme sürecinin temel sonucunu ifade eder. Değerlendirme sürecinde doğru kararlar almak ve sürecin güvenilirliğini sağlamak büyük bir önem taşır. Hatalı bir karar hem sistemin işleyişi hem de bireyler hakkında yanlış çıkarımlara yol açarak süreci olumsuz etkileyebilir. Örneğin, bir programın hedeflenen davranışları kazandırıp kazandırmadığını belirlemeyi amaçlayan bir değerlendirme sürecinde, değerlendirme alt sisteminin hatasız olduğu varsayılırsa, gerçekte programın iyileştirilmesi gerekirken öğretimin yetersiz olduğu sonucuna ulaşılabilir. Bu durum, sistemin gereksiz bir şekilde eleyici bir yapıya dönüşmesine neden olabilir. Benzer şekilde, öğrencilerin düzeylerinin altında bir eğitim programında yüksek başarı göstermesi, sistem kaynaklarının israf edilmesine ve zaman, emek ile maliyetin boşa harcanmasına yol açabilir. Dolayısıyla, değerlendirme süreçlerinde dikkatli ve doğru kararlar vermek hem sistemin etkinliğini artırmak hem de bireylerin gerçek potansiyellerini ortaya koymak açısından hayati bir öneme sahiptir (Baykul ve ark., 1993, s. 110). Web sayfasında bulunan ölçümler eğitmene sorduğu soruların kolay ya da zor olduğu ve öğrencinin konuyu anlayıp anlayamama düzeyini tespit eden bir karar çıktısı vermektedir.

1.2.2 Çoktan Seçmeli Testler

Çoktan seçmeli testler, cevaplayıcının verilen seçenekler arasından sorunun doğru cevabını seçmesini gerektiren bir test türüdür. Seçenek sayısının yalnızca iki ile sınırlı olduğu doğru-yanlış testleri ise, genellikle çoktan seçmeli testlerin bir alt kategorisi olarak değerlendirilir. Bu test türleri, farklı ölçme gereksinimlerini karşılamak için kullanılır ve yapısal özelliklerine bağlı olarak çeşitli avantajlar ve sınırlamalar içerir (Aslan, 2018, s. 112). Yapmış olduğumuz web sayfasında ise sadece çoktan seçmeli test ölçümleri ele alınmaktadır.

Çoktan seçmeli testlerin geçerlilik, güvenilirlik ve nesnellik gibi avantajları, ancak iyi yapılandırıldıkları takdirde değerlendirilebilir. Ancak, sıklıkla değerlendirme amacıyla kullanılan kötü yapılandırılmış çoktan seçmeli testler, Bloom'un taksonomisinin gerektirdiği bilişsel seviyeleri ölçememekte, öğrenciler için ya çok kolay ya da çok zor olmakta, belirsizlik içermekte, kolayca 'tahmin edilebilmekte' ya da yüksek başarı düzeyine sahip öğrenciler ile düşük başarı düzeyine sahip öğrenciler arasında ayırım yapamamak gibi problemlere neden olabilir. Bu nedenle, çoktan seçmeli testlerin kalitesinin düzenli olarak değerlendirilmesi büyük önem taşımaktadır. Oluşturduğumuz web sayfasında bu değerlendirmeye yönelik her ölçülen sınavda sınav zorluğu hesaplanmaktadır. Kalite değerlendirmesi bulgularına dayanarak iyileştirmeler yapılabilir ve bu maddeler daha sonraki testlerde kullanılabilir (Shete ve ark., 2015, s. 108).

Çoktan seçmeli testlerin belirgin bir avantajı, kısa bir süre içinde çok sayıda adayı geniş bir kavramsal kapsamda değerlendirme yeteneğidir. Bu nedenle çoktan seçmeli ölçme araçları eğitimciler tarafından sıklıkla kullanılsa da bu ölçme araçları beklediğimiz geçerli ve güvenilir ölçmeyi her zaman sağlamayabilir. Bilgisayar kullanılarak yapılan puanlama nesnel, güvenilir ve hızlıdır. Çoktan seçmeli testler genellikle bilgiye dayalı hatırlama becerisini ölçmek için kullanılsa da uygulama, analiz, sentez ve bilgi değerlendirmesi gibi üst düzey düşünme becerilerini de ölçebilir. Ancak, çoktan seçmeli testlerin kullanımıyla ilgili birkaç sorun arasında öğrencilerin tahmin yapma veya ipuçlarından faydalanma etkisi ve geri bildirim eksikliği yer alır. Hatalı sorular, test puanlarının doğru ve anlamlı bir şekilde yorumlanmasını engeller ve değerlendirme süreci üzerinde olumsuz bir etki yaratır (Singh ve ark., 2021, s. 92). Tarafımızca geliştirilen ve bu çalışma ile incelenen web sitemiz öğretmenlerin daha kaliteli ölçme araçları geliştirebilmeleri için çoktan seçmeli ölçme araçlarını analiz edebilmelerini sağlamaktadır.

1.2.3. Test ve Madde Analizi

Oluşturulan web sitesinde, test analizi sürecinde bu bölümde açıklanan formüller kullanılarak analiz sonuçları elde edilmekte ve kullanıcıya sunulmaktadır. Web sitesinde yer alan sonuç çıktıları Şekil 1'de örnek olarak gösterilmiştir.

Şekil 1.Web Sitesi Test Analizi Sonuç Çıktıları

Test Analizi		
En Yüksek Puan	35	
En Düşük Puan	16	
Ortalama	21.893	
Ranj	19	
Tepe Değer (Mod)	20, 19,	
Ortanca (Medyan)	21.000	
Testin Ortalama Güçlüğü	0.626	
KR20	0.720	
Standart Sapma	4.601	
Varyans	21.167	
Bağıl Değişim Katsayısı	21.015	
Çarpıklık Katsayısı	1.583	

Kaynak: Çakmaktepe, 2024.

Test uygulamasından elde edilen sonuçların puanlanmasının ardından, eğitim-öğretim süreçlerinin etkinliğini artırmak ve geliştirmek amacıyla bu sonuçların detaylı bir şekilde analiz edilmesi ve değerlendirilmesi büyük önem taşımaktadır. Bu süreç, ölçme sonuçlarından anlamlı çıkarımlar yaparak, süreçlerin güçlendirilmesine yönelik kararların alınmasını sağlar (Yılmaz, 1996, s. 209).

1.2.3.1. Tepe Değer

Tepe değer, bir veri setindeki en sık gözlemlenen değeri tanımlayan önemli bir istatistiksel ölçüttür. (Başol, 2016, s. 189).

1.2.3.2. Aritmetik Ortalama

Merkezi eğilim ölçütleri arasında en yaygın kullanılanı aritmetik ortalamadır. Bunun temel sebebi, aritmetik ortalamanın tüm veri setini dikkate alarak hesaplanması ve ileri düzey istatistiksel analizlerde en uygun ve işlevsel ölçüt olarak kabul edilmesidir. (Baykul, 1997, s. 231).

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

(SEQDenklemlerARABIC 1)

1.2.3.3. Medyan

Sıralanmış bir veri setinin tam ortasında bulunan değer, medyan olarak tanımlanır. Medyan, bir veri setini %50'lik iki eşit gruba ayıran merkezi değer olmasıyla dikkat çeker. (Güler, 2011, s. 139).

$$x = \frac{n+1}{2}$$

(2)

1.2.3.4. Ranj

Bir veri kümesindeki en yüksek değer ile en düşük değer arasındaki fark, ranj olarak ifade edilir ve bu ölçüt, veri setinin yayılımını tanımlamak için kullanılan temel istatistiksel göstergelerden biridir (Baykul ve Güzeller, 2013, s. 269).

1.2.3.5. Varyans

Varyans, istatistiksel analizlerde bir veri setindeki değerlerin dağılımını ve ortalamadan sapma düzeylerini ölçmek amacıyla kullanılan temel bir istatistiksel ölçüttür. Bu ölçüt, değerlerin ortalamadan olan sapmalarının karelerinin toplamının evren için N'ye, örneklem için ise (n-1)'e bölünmesiyle hesaplanır. (Büyüköztürk ve ark., 2013, s. 41).

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

(3)

Varyans, verilerin ortalamadan olan sapmalarının karelerinin ortalamasını ifade eder. Ancak, varyans biriminin karesel olması nedeniyle, uygulamada genellikle karekökü alınarak kullanılır. (Alpar, 2016, s. 98).

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

(4)

1.2.3.6. Bağlı Değişim Katsayısı

Standart sapma ve varyans, aritmetik ortalamaya dayalı olmaları nedeniyle gruplar arası karşılaştırmalarda yanıltıcı neden olabilir. Bu tür hataları azaltmak için standart sapma

ile ortalamayı birlikte değerlendiren bağıl değişim katsayısı kullanılır. Bu katsayı, farklı ortalamalara sahip grupların daha objektif karşılaştırılmasını sağlar (Doğan, 2008, s. 268).

$$V = \frac{S_x}{\bar{X}} \cdot 100$$

(5)

1.2.3.7. Çarpıklık Katsayısı

Bir dağılımın simetri özellikleri hakkında bilgi edinmek, dağılımın sağa veya sola çarpıklık gösterip göstermediğini analiz etmekle mümkündür (Kasap, 2012, s. 92). Çarpıklık katsayısı, dağılımın simetri durumunu ve şekil özelliklerini değerlendiren bir ölçüttür ve pozitif, negatif veya sıfır değer alabilir. (Gömlüksiz ve ark., 2008, s. 266).

$$çk = \frac{3 \cdot (\bar{x} - m_{\text{ort}})}{s_x}$$

(6)

1.2.3.8. Kuder Richardson -20 (KR20)

KR20, bir çoktan seçmeli ölçme aracının iç tutarlılığını değerlendirmeye yönelik olarak kullanılan bir istatistiksel yöntemdir. 1-0 puanlama sistemine dayalı yapısı sayesinde, özellikle çoktan seçmeli ve doğru-yanlış türündeki testlerde yaygın bir şekilde uygulanmaktadır. (Kasap, 2012, s. 81).

$$K 20 = \frac{K}{K-1} \left(1 - \frac{\sum p \cdot q}{S_x^2} \right)$$

(7)

1.2.3.9. Kuder Richardson -21 (KR21)

Test maddelerine farklı ağırlıklar verildiğinde veya şans başarısı düzeltilmesi yapıldığında, KR20 formülü uygun değildir çünkü maddeler arası kovaryansların eşitliğini varsayar. Bu tür durumlarda, maddelerin eşit güçlükte olduğu varsayılıyorsa KR21 formülü tercih edilir. KR21, KR20'nin özel bir durumu olarak homojen testlerde yaygın şekilde kullanılır (Özen ve ark., 2006, s. 85).

$$K 21 = \frac{K}{K-1} \left(1 - \frac{K \bar{X} - \bar{X}^2}{k \cdot s_x^2} \right)$$

(8)

1.2.3.10. Ortalama Güçlük İndeksi

Bir çoktan seçmeli ölçme aracının ortalama güçlüğü, testte yer alan maddelerin güçlük indekslerinin aritmetik ortalaması şeklinde tanımlanır ve çoktan seçmeli ölçme aracının genel zorluk düzeyini değerlendirmek için kullanılan bir ölçüttür (Tekin, 1977, s. 220).

$$\bar{P} = \frac{\bar{X}}{K}$$

(9)

1.2.3.11. Madde Analizi

Madde analizi, çoktan seçmeli ölçme aracındaki maddelerin belirli niteliklere göre sistematik olarak değerlendirilmesini amaçlar. Bu sürecin sağlıklı ilerleyebilmesi için, analiz öncesinde bir ön uygulama yapılmalı ve elde edilen veriler dikkatle incelenmelidir (Tomak, 2013, s. 11).

1.2.3.12. Güçlük İndeksi

Güçlük indeksi, bir test maddesini doğru yanıtlayanların tüm gruba oranını gösterir ve maddenin zorluk düzeyini belirlemek için kullanılır. Değer 1.00'a yaklaştıkça madde kolay, 0.00'a yaklaştıkça zor kabul edilir (Dündar, 2019, s. 74).

$$p_j = \frac{n(D)}{N}$$

(10)

1.2.3.13. Maddenin Ayırt Ediciliği

Bir maddenin ayırt ediciliği, hedef davranışı gösteren ve göstermeyen öğrencileri birbirinden ayırma derecesini ifade eder ve genellikle korelasyon istatistikleriyle belirlenir. Bu çalışmada, madde ve test puanlarının sürekli değişkenler olması nedeniyle ayırt edicilik indeksleri Pearson momentler çarpımı korelasyon katsayısı ile hesaplanmıştır (Acar, 2018, s. 505).

$$r_{jx} = \frac{D\ddot{u} - D\ddot{a}}{N}$$

(11)

1.2.3.14. Madde Varyans ve Standart Sapması

Madde varyansı ve standart sapması, bir maddeye verilen yanıtların (0-1 puanlarının) dağılımındaki değişim düzeyini ifade eden istatistiksel ölçütlerdir. Bu değerler, maddeye ait puanların değişkenliğini nicel olarak değerlendirmeye olanak sağlar (Güler, 2011, s. 120).

$$s_j^2 = p_j(1 - p_j)$$

(12)

$$s_j = \sqrt{p_j(1 - p_j)}$$

(13)

1.2.3.15. Madde Güvenirlik Katsayısı

Madde güvenilirliğini temsil eden madde güvenirlilik katsayısı, bir maddenin ayırt edicilik gücü ile standart sapmasının çarpımıyla elde edilen bir değerdir (Baykul ve Güzeller, 2013, s. 588).

1.2.3.16. Korelasyon Katsayısı

Korelasyon katsayısı, istatistikte iki değişken arasındaki ilişkinin yönünü ve şiddetini nicel olarak ifade eden bir ölçüttür. Bu katsayı, değişkenler arasındaki ilişki düzeyine bağlı

olarak pozitif (+) ya da negatif (-) işaret alabilir ve ilişkinin gücüne göre yüksek veya düşük bir değer gösterebilir (Çağlar, 1970, s. 192).

1.2.3.17. Madde Ayırt Edicilik İndeksi

Madde ayırt edicilik indeksinin hesaplanmasında yaygın olarak kullanılan yöntemlerden biri, nokta çift serili korelasyon katsayısının belirlenmesidir. Bu katsayı, bir test maddesinin (doğru veya yanlış olarak puanlanan) bireyin toplam test puanı ile olan ilişkisini ölçer. (Büyüköztürk ve ark., 2013, s. 99).

$$r_{pb} = \left(\frac{\bar{Y}_p - \bar{Y}_q}{s_y} \right) \cdot \sqrt{pq} \quad (14)$$

1.2.3.18. Çift Serili Korelasyon Katsayısı

Çift serili korelasyon katsayısı, sürekli bir değişkenin iki kategorili hale getirildiği durumlarda kullanılan ve özellikle eğitim ile psikolojide test geliştirme sürecinde yaygın olarak tercih edilen bir istatistiksel yöntemidir. Bu süreçte, test maddeleri doğru yanıtlar için 1; yanlış, boş veya çoklu yanıtlar için 0 puanla değerlendirilir (Baykul ve Güzeller, 2013, s. 584).

$$r_b = \left(\frac{\bar{Y}_p - \bar{Y}_q}{s_y} \right) \cdot \frac{pq}{y} \quad (15)$$

1.2.4. Kullanılabilirlik

Kullanılabilirlik, ürünün kalitesi, verimliliği, kullanıcı memnuniyeti ve tatmini açısından genel bir kavram haline gelmiş ve genellikle 'kullanıcı dostu' ve 'kullanım kolaylığı' terimleriyle ilişkilendirilmiştir. Kullanılabilirlik, kullanıcı ihtiyaçlarının niteliği ve ürün kullanımında bireysel değerlendirmelerle ilgilidir. Bu kavram, uluslararası standartlarda genellikle üç ana bileşen üzerinden tanımlanmaktadır: 'verimlilik', 'etkililik' ve 'tatmin'.

Etkililik, bir ürünün veya sistemin belirlenen hedeflere ulaşmak için gerekli olan fonksiyonları ne derece yerine getirebildiğini ifade eder ve bu süreçte kullanılabilirlik kavramıyla büyük ölçüde örtüşür.

Verimlilik, kullanıcıların taleplerine uygun şekilde hareket edilmesini, ancak aynı zamanda kullanıcıların bilişsel ve fiziksel yeterlilikleri doğrultusunda bu hedefleri gerçekleştirebilmelerinin sağlanmasını ifade eder. Bu hem süreçlerin hem de kaynakların etkin bir şekilde yönetilmesini gerektirir.

Tatmin, kullanıcıların ürün veya sistemle etkileşimde bulunduklarında subjektif değerlendirmelerini ifade eder ve bu etkileşimi onların beklentilerini karşılayacak şekilde kabul edilebilir kılmayı amaçlar. Bu, kullanıcı deneyiminin olumlu algılanmasını sağlamayı hedefler (Telek, 2013, s. 6-7).

Bahsi geçen katsayılar, değerler ve tanımların anlatılmasındaki amaç makalemizin konusu olan web sayfasının kullanılabilirliğini test etmek açısından sınavın değerlendirmesini

ve hesapladığı değerler hakkında bilgi sahibi olup bunların doğruluğunu teyit etmektir. Verilen her açıklama sitede hesaplanan değerlerle bağlantılıdır. Bu metin, ölçme ve değerlendirme kavramlarının teorik temellerini detaylı bir şekilde ele alarak, bu süreçlerin farklı uygulama alanlarındaki önemini ve işlevini açıklamaktadır. Ölçme, olayların ya da varlıkların belirli özelliklerini sayısal veya sembolik ifadelerle tanımlamayı amaçlayan sistematik bir süreç olarak ele alınmakta; değerlendirme ise ölçme sonuçlarının önceden belirlenmiş ölçütlerle karşılaştırılarak karar verme sürecine katkıda bulunmasını içermektedir. Doğrudan, dolaylı ve türetilmiş ölçme türlerini tanımlayarak bu süreçlerin farklı bağlamlardaki uygulamalarını detaylandırmaktadır. Ayrıca, güvenilirlik, kullanılabilirlik ve test analizi gibi ölçme araçlarının temel özellikleri ve bu araçların sistematik karar alma süreçlerine etkisi üzerinde durulmaktadır. Bu bağlamda, ölçme ve değerlendirme süreçlerinin güvenilirlik, geçerlilik ve işlevsellik açısından taşıdığı kritik rol, ilgili kavramların teorik ve uygulamalı çerçevede anlaşılmasını sağlamak amacıyla kapsamlı bir biçimde analiz edilmiştir.

1.3. Geliştirilen Ölçme ve Değerlendirme Web Sayfası

Web sayfasının hazırlanması sürecinde aşağıda belirtilen tasarım ve programlama araçlarından faydalanılmıştır.

- HTML 5
- CSS 3
- BOOTSTRAP 5 (Css Framework)
- JavaScript
- Laravel 10 (PHP Framework)
- jQuery (JS Framework)
- MySQL

Modern web geliştirme süreçlerinde HTML5, CSS3, Bootstrap 5 ve JavaScript temel teknolojiler arasında yer almakta olup, bu teknolojilerin entegrasyonu ile daha işlevsel, estetik ve duyarlı web uygulamaları geliştirmek mümkün hale gelmektedir.

HTML5, web sitelerinin oluşturulmasında kullanılan metin işaretleme dili olan HTML'in (Hypertext Markup Language) beşinci ve en güncel sürümüdür. CSS3 ise, HTML ile yapılandırılan içeriklerin görsel açıdan daha zengin ve estetik bir biçimde sunulmasını sağlayan stil tanımlama dilidir. Bootstrap, HTML, CSS ve JavaScript teknolojilerini bir araya getiren, açık kaynak kodlu bir çerçeve (framework) olup, duyarlı (responsive) ve mobil uyumlu web tasarımlarının hızlı ve verimli bir şekilde geliştirilmesine olanak tanımaktadır. JavaScript ise, geliştiricilerin kullanıcı etkileşimini artıran dinamik ve etkileşimli web sayfaları oluşturmasını sağlayan bir programlama dilidir.

Bu projede, kullanıcıya sunulan tüm arayüzler Laravel View bileşeni aracılığıyla ve söz konusu teknolojiler kullanılarak oluşturulmuştur. Arka planda gerçekleşen veri yönetimi, kullanıcı işlemleri ve istatistiksel hesaplamalar Laravel Controller tarafından yürütülmektedir. Kullanıcı bilgileri, sınav içerikleri ve sınav ayarlarına ilişkin veriler ise MySQL veri tabanında saklanmaktadır.

Şekil 1. Web Sitesi Test Analizi Sayfası

Sınav Analizi

Sınav İşlemleri > Sınav Analizi

Sonuçları PDF olarak indir

Test Sonuç

Aday Kimlik	Kitapçık	Doğru	Yanlış	Boş	Detay
214207092	A	35	0	0	Detay
204207054	B	35	0	0	Detay
224207086	B	26	7	2	Detay
214208001	B	26	7	2	Detay
214207022	B	25	8	2	Detay
194208028	B	25	9	1	Detay
214209004	B	24	9	2	Detay
214207016	B	24	9	2	Detay
224207083	A	23	10	2	Detay
204209001	B	23	10	2	Detay
214207057	B	22	13	0	Detay
194208039	A	22	13	0	Detay

Test Analizi

En Yüksek Puan	35	
En Düşük Puan	16	
Ortalama	21.893	
Ranj	19	
Tepe Değer (Mod)	20, 19	
Ortanca (Medyan)	21.000	
Testin Ortalama Güçlüğü	0.626	
KR20	0.720	
Standart Sapma	4.601	
Varyans	21.167	
Bağıl Değişim Katsayısı	21.015	
Çarpıklık Katsayısı	1.583	

Test Frekans Tablosu

Puanlar	Frekanslar
35	2
26	2
25	2
24	2
23	2
22	2
21	3
20	4
19	4
18	2
16	3

Kaynak: Çakmaktepe, 2024.

Sınav analizlerini içeren bu sayfada, 'Test Sonuçları', 'Test Analizi,' ve 'Test Frekans Tablosu' olmak üzere üç ana bölüm bulunmaktadır (Şekil 1). Test sonuçları bölümünde, sınava katılan adayların kitapçık türlerine ek olarak, verdikleri doğru, yanlış ve boş yanıtlar listelenmektedir. Ayrıca, 'Detay' butonuna tıklanmasıyla birlikte, cevap anahtarıyla karşılaştırma yapılarak detaylı koşullu biçimlendirme sağlanmaktadır (Şekil 2). Bu biçimlendirme kapsamında, doğru yanıtlar yeşil renkte '+' sembolüyle, yanlış yanıtlar ise kırmızı renkte '-' sembolüyle belirtilmektedir.

Şekil 2. Öğrenci Cevapları Detaylar Bölümü

Detaylar				✕	
Soru Numarası	Aday Cevap	Doğru Cevap	Durum		
1		B	-		
2		N	-		
3	C	C	+		
4	B	B	+		
5	A	A	+		
6	B	D	-		
7	C	C	+		
8	B	B	+		
9	A	A	+		
10	B	B	+		

Kaynak: Çakmaktepe, 2024.

Test analizi bölümünde ise, ölçme ve değerlendirme literatüründen yararlanılarak gerçekleştirilen analiz sonucunda, çoktan seçmeli ölçme aracının aşağıdaki istatistiksel değerleri hesaplanmaktadır: en yüksek puan, en düşük puan, ortalama, ranj, tepe değer (mod), ortanca (medyan), çoktan seçmeli ölçme aracının ortalama gücü, KR-20 güvenirlik katsayısı, standart sapma, varyans, bağıl değişim katsayısı ve çarpıklık katsayısı. Her bir istatistiksel değer yanında, bu kavramın tanımı ve analiz sonuçlarına ilişkin yorumlara ulaşmak için 'i' simgeli bilgi butonları yer almaktadır. Şekil 3, bu açıklamalara ilişkin bir örnek sunmaktadır.

Şekil 3. Test Analizleri Değerlerin Detayları

Detaylar



Hesaplanan: Testin Ortalama Güçlüğü,

Tanım: Test puanlarının ortalama gücü ise test ortalamasının testten alınması mümkün olan en yüksek puana bölünmesiyle bulunur,

Değer: 0.626

Yorum:

- Eğitim öğretim etkinlikleri yeterli düzeydedir
- Öğrenciler, konuları öğrenmiştir
- Test, güçlük düzeyi düşük, kolay sorulardan oluşmaktadır

Kaynak: Çakmaktepe, 2024.

Çoktan seçmeli ölçme aracının frekans tablosunda ise, öğrencilerin aldıkları puanların büyükten küçüğe doğru sıralı bir dağılımı ve her puan aralığında yer alan öğrenci sayıları gösterilmektedir. Bu tablo, sınav sonuçlarının genel dağılımını görselleştirerek analiz sürecini kolaylaştırmaktadır.

Şekil 4. Web Sitesi Madde Analizi Sayfası

Madde Seçimi

A

7

Getir

Madde İstatistik Bilgileri

	Tümü	Üst Grup	Alt Grup
Doğru Sayısı	2	1	0
Yanlış Sayısı	12	2	3
TOPLAM	14	3	3

Madde Cevap ve Çeldiriciler

Madde Şık	Cevap Toplamı
C	12
D	2

Madde İstatistik Hesaplama Sonuçları

İstatistik	Değer
Madde Güçlük İndeksi (p_i)	0.143
Madde Ayırt Edicilik İndeksi (r_{px})	0.333
Madde Varyans (s_y^2)	0.122
Madde Standart Sapma (s_y)	0.350
Madde Güvenirlik İndeksi	0.117
Nokta Çift Serili Korelasyon (r_{pb})	0.590
Çift Serili Korelasyon (r_b)	0.000

Kaynak: Çakmaktepe, 2024.

'Madde analizi' seçeneğine tıklandığında, öğrencilerin yanıtladığı her bir soruya ilişkin ayrıntılı istatistiksel verilerin sunulduğu bir sayfa açılmaktadır. Bu sayfa, her bir sorunun bireysel olarak analiz edilmesine olanak tanıyarak, soruların istatistiksel özelliklerinin değerlendirilmesini sağlamaktadır. Şekil 4'te gösterildiği üzere, 'Madde Analizi' sayfası beş ana bölümden oluşmaktadır. İlk bölüm olan Madde Seçimi, sınava ait soruların kitapçık türüne göre seçilmesine ve soru maddelerinin belirlenmesine olanak tanımaktadır. Madde İstatistik Bilgileri bölümünde ise, seçilen soruya doğru ve yanlış yanıt veren kişi sayıları ile üst ve alt grup analiz değerleri sunulmaktadır.

Bunun hemen altında yer alan Madde Cevap ve Çeldiriciler bölümü, seçilen soruya verilen yanıtların dağılımını detaylı şekilde göstermektedir. Bu bölüm, doğru cevaba yönelen öğrenci sayısının yanı sıra diğer seçeneklere yönelen öğrenci sayılarını karşılaştırma imkânı tanımakta ve böylece doğru cevabı seçemeyen öğrencilerin hangi çeldirici seçeneklere daha çok yöneldiğini belirlemeye olanak sağlamaktadır. Madde İstatistik Hesaplama Sonuçları bölümünde ise ölçme ve değerlendirme literatürüne dayalı olarak madde güçlük indeksi, madde ayırt edicilik indeksi, madde varyansı, madde standart sapması, madde güvenilirlik indeksi, nokta çift serili korelasyon ve çift serili korelasyon gibi istatistiksel değerler hesaplanmaktadır.

Son olarak, Detaylı İstatistik Bilgileri bölümü, bu istatistiksel değerlerin tanımlarını ve analiz sonuçlarına yönelik yorumları içermektedir. Her bir değer yanındaki açıklama butonu kullanılarak, söz konusu değer anlamı ve soruya ilişkin sonuçları hakkında bilgi edinilebilmektedir. Bu yapı, madde bazlı analizlerin kapsamlı ve ayrıntılı bir şekilde gerçekleştirilmesini sağlayarak, sınav sorularının ölçme ve değerlendirme açısından etkinliğini ortaya koymaktadır.

2. Yöntem

2.1. Araştırmanın Modeli ve Hipotezleri

Bu çalışmanın temel problemi, 'Çoktan seçmeli testlerin güvenilirlik analizini hesaplama amacıyla geliştirilen 'sinavanalizlerim.com' adlı web sitesinin, bu amaca ne ölçüde katkı sağladığı ve bilişim terminolojisi bağlamında 'kullanılabilir' olup olmadığı' sorusuna yanıt aramaktır.

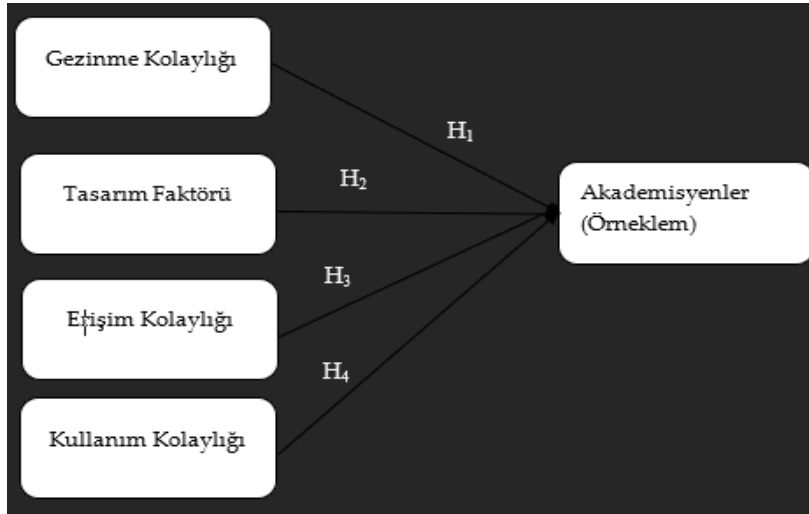
H₁: Web sitemizin kullanılabilirliği gezinme kolaylığı faktörü açısından yeterlidir.

H₂: Web sitemizin kullanılabilirliği tasarım kolaylığı faktörü açısından yeterlidir.

H₃: Web sitemizin kullanılabilirliği erişim kolaylığı faktörü açısından yeterlidir.

H₄: Web sitemizin kullanılabilirliği kullanım kolaylığı faktörü açısından yeterlidir.

Şekil 5. Araştırmanın Kavramsal Modeli



2.2. Araştırmanın Evreni/Örnekleme ve Kısıtları

Bu araştırmanın evrenini, Afyon Kocatepe Üniversitesi Eğitim Fakültesi'nde görev yapan akademik personel oluşturmaktadır. Ancak zaman, maliyet, erişim güçlükleri ve denetim sınırlılıkları gibi etkenler nedeniyle evrenin tamamına ulaşmanın pratikte mümkün olmadığı kabul edilmiştir. Bu doğrultuda, Formül 16'ya dayalı olarak gerçekleştirilen örneklem büyüklüğü hesaplamasında, evrenin 57 kişiden oluştuğu durumda %95 güven düzeyi ve $\pm 5\%$ hata payı ile 50 kişilik bir örneklem grubunun istatistiksel açıdan yeterli olduğu sonucuna ulaşılmıştır.

$$n = \frac{N \cdot Z^2 \cdot p \cdot (1 - p)}{(N - 1) \cdot e^2 + Z^2 \cdot p \cdot (1 - p)} \quad (16)$$

Tablo 1. Afyon Kocatepe Üniversitesi'nde Bulunan Akademisyenlerin Demografik Özelliklerine Göre Dağılımı

Demografik Özellikler	Sıklık	Yüzde	Toplam Yüzde
Cinsiyet			
Kadın	16	32,0	32,0
Erkek	34	68,0	100,0
Yaş			
22 - 29 yaş	9	18,0	18,0
30 - 37 yaş	10	20,0	38,0
38 - 45 yaş	18	36,0	74,0
46 - 53 yaş	9	18,0	92,0
54 yaş ve üzeri	4	68,0	100,0

Unvan			
Araştırma Görevlisi	13	26,0	26,0
Öğretim Görevlisi	6	12,0	38,0
Dr. Öğretim Üyesi	12	24,0	62,0
Doçent	15	30,0	92,0
Profesör	4	8,0	100,0
Meslekte Çalışma Süreniz			
1-5 yıl	11	22,0	22,0
6-10 yıl	0	0	22,0
11-15 yıl	14	28,0	50,0
16-20 yıl	12	24,0	74,0
21 yıl ve üzeri	13	26,0	100,0

Tablo 1'e göre %68', erkeklerden oluşan katılımcıların %36'sı 38 - 45 yaş aralığındadır. Katılımcıların çoğunluğu doçent unvanına sahiptir ve katılımcıların çoğunluğu 11 - 15 yıl süresinde mesleklerini sürdürmektedirler.

2.3. Araştırmanın Veri Toplama Yöntemi

Bu çalışmada veri toplama yöntemi olarak anket tekniği tercih edilmiştir. Anketler, katılımcıların gönüllülük esasına dayalı olarak yer aldığı yüz yüze görüşmeler yoluyla üniversite ortamında uygulanmıştır. Anket formu, akademisyenlerin sosyo-demografik özelliklerini (cinsiyet, yaş, unvan, çalışma süresi) belirlemeye yönelik sorular ile web sitesi kullanılabilirliğini değerlendirmeye yönelik bir ölçeği içermektedir. Web sitesinin kullanılabilirlik düzeyini ölçmek amacıyla, geçerlilik ve güvenilirlik çalışmaları Çakmak ve arkadaşları (2011, s. 35) tarafından gerçekleştirilen "Web Sitesi Kullanılabilirlik Ölçeği" bu çalışmada herhangi bir değişiklik yapılmaksızın kullanılmıştır. Ölçek; 1-9 numaralı maddelerle gezinme kolaylığı, 10-16 numaralı maddelerle tasarım kolaylığı, 17-20 numaralı maddelerle erişim kolaylığı ve 21-29 numaralı maddelerle kullanım kolaylığı alt boyutlarını kapsamaktadır.

2.4. Araştırmanın Veri Analiz Yöntemleri ve Güvenilirliği

Araştırma verileri, SPSS istatistik programı kullanılarak analiz edilmiştir. Çalışmada, katılımcıların demografik ve bireysel özelliklerine ilişkin dağılımlar, frekans ve yüzde olarak sunulmuştur. Ölçeklerin daha önceki çalışmalarda geçerlik ve güvenilirlik testlerinden geçmiş olmasına rağmen, bu çalışmada ilgili ölçekler için iç tutarlılık analizi uygulanarak, ölçeklerin orijinal yapılarıyla uyumlu olup olmadığı değerlendirilmiştir. İç tutarlılık analizi sonuçları, çalışmanın ilerleyen bölümlerinde detaylı şekilde ele alınmıştır. Bunun yanı sıra, kullanılan ölçeklerin güvenilirlik düzeylerini belirlemek amacıyla Cronbach's Alpha katsayısından yararlanılmıştır. Son olarak, akademisyenlerin web sitesi kullanılabilirliğine ilişkin değerlendirmeleri, frekans ve yüzde dağılımlarının yanı sıra aritmetik ortalama ve standart sapma değerleri hesaplanarak betimleyici istatistiklerle ortaya konulmuştur.

2.5. Araştırmanın Bulguları

2.5.1. Normallik Analizleri

Tablo 2. Araştırmada Kullanılan Ölçeğin Güvenilirlik Analizi

Ölçekler	Güvenilirlik Analizi
	Cronbach's Alpha (α)
Web Sitesi Kullanılabilirlik Ölçeği (Kılıç ve ark., 2011, s. 36).	0,885

Bu çalışmada kullanılan ölçeğin Türkçe geçerlilik ve güvenilirlik analizleri daha önce yapılmış olduğundan, ölçeğin geçerliliği yeniden değerlendirilmemiştir. Bunun yerine, ölçeğin güvenilirliğini belirlemek amacıyla Cronbach Alpha katsayısı hesaplanmıştır. Tablo 2'de belirtildiği üzere, elde edilen Cronbach Alpha değeri 0,885 olarak bulunmuştur. Çakmak vd. (2011, s. 36) yapmış oldukları çalışmada ölçeğin Cronbach Alpha değerini 0,930 olarak bulmuşlardır bu iki sonuç incelendiğinde ölçeğin güvenilir olduğu tespit edilmiştir. Cronbach Alpha katsayısının 0,70'in üzerinde olması, ölçeğin çalışmada güvenilir bir şekilde kullanılabileceğini göstermektedir (Gürbüz ve Şahin, 2018, s. 134).

Tablo 3. Katılımcıların 'Gezinme Kolaylığı' Faktörünü Oluştan Maddele İlişkin Algılarının Aritmetik Ortalaması

	N	X
1. Madde	50	4,36
1. Madde	50	4,42
1. Madde	50	4,42
1. Madde	50	4,50
1. Madde	50	4,52
1. Madde	50	4,48
1. Madde	50	4,32
1. Madde	50	4,30
1. Madde	50	4,36

Tablo 3 incelendiğinde, 'gezinme kolaylığı' faktörü altında yer alan maddelerden en yüksek puana sahip olanın 5. madde olduğu görülmektedir. 'Menülerdeki başlıklar ile sayfa içerikleri tutarlı' ifadesiyle tanımlanan bu madde, kullanıcıların web sitesine yönelik algılarının olumlu olduğunu göstermektedir. Söz konusu özellik, kullanıcılar tarafından 4,52 puanla 'çok iyi' olarak değerlendirilmiş ve web sitesinin sayfa içeriği ile menü uyumu açısından güçlü bir yönü olduğu ortaya konmuştur.

Tablo 4. Katılımcıların Tasarım Faktörünü Oluştan Maddele İlişkin Algılarının Aritmetik Ortalaması

	N	X
1. Madde	50	4,48
1. Madde	50	4,56
1. Madde	50	4,28
1. Madde	50	4,32
1. Madde	50	4,34
1. Madde	50	4,42
1. Madde	50	4,34

Tablo 4'te yer alan verilere göre, web sitesinin tasarım faktörüne ilişkin değerlendirmede bulunan maddelerin puanları genel olarak 'çok iyi' seviyesinde kaydedilmiştir. Bu durum, web sitesinin tasarım özelliklerinin güçlü bir yönü olduğunu göstermektedir. Maddeler arasında en yüksek puan, 'Kullanılan yazı tipleri tutarlı' ifadesine ait olan 11. maddeye aittir ve bu madde 4,56 puan almıştır. Bu bulgu, web sitesinde kullanılan yazı tiplerinin kullanıcılar açısından önemli bir tasarım gücü olduğunu ortaya koymaktadır.

Tablo 5. Katılımcıların Erişim Kolaylığı Faktörünü Oluştan Maddelere İlişkin Algılarının Aritmetik Ortalaması

		N	X
1.	Madde	50	4,54
1.	Madde	50	4,44
1.	Madde	50	4,14
1.	Madde	50	4,20

Tablo 5'te yer alan veriler incelendiğinde, web sitesinin erişim kolaylığına ilişkin değerlendirmelerin genel olarak 'çok iyi' seviyesinde olduğu görülmüştür. Bu durum, web sitesinin erişim kolaylığı açısından önemli bir güçlü yönü olduğunu ortaya koymaktadır. Maddeler arasında en yüksek puanı, 4,54 ile 'Site hızlı yükleniyor' ifadesini içeren 17. madde almıştır. Bu bulgu, sitenin yüklenme hızının kullanıcı deneyimi açısından dikkate değer bir erişim avantajı sunduğunu göstermektedir.

Tablo 6. Katılımcıların Kullanım Kolaylığı Faktörünü Oluştan Maddelere İlişkin Algılarının Aritmetik Ortalaması

		N	X
1.	Madde	50	3,88
1.	Madde	50	4,06
1.	Madde	50	3,30
1.	Madde	50	3,76
1.	Madde	50	4,34
1.	Madde	50	4,48
1.	Madde	50	4,48
1.	Madde	50	4,36
1.	Madde	50	4,54

Tablo 6'daki veriler incelendiğinde, 23. madde olan 'İçeriği henüz tamamlanmamış sayfa veya sayfalar var' ifadesinin 3,30 puan ile en düşük puanı aldığı ve 'orta' olarak değerlendirildiği görülmektedir. Buna karşın, 29. madde olan 'Madde analizi kısmı, soruları detaylı bir şekilde incelemeye yardımcı oluyor' ifadesi 4,54 puan ile en yüksek puanı almıştır. Bu bulgular, web sitesinin kullanım kolaylığı açısından genel olarak 'iyi' bir düzeyde olduğunu, ancak sonuç çıktılarının anlaşılması ve yorumlanması noktasında 'çok iyi' bir performans sergilediğini ortaya koymaktadır.

Tablo 7. Katılımcıların sinavanalizlerim.com'un Kullanılabilirliğine Yönelik Algıları

	N	X
Genel Kullanılabilirlik	50	113,70

Gezinme kolaylığı, tasarım kolaylığı, erişim kolaylığı ve kullanım kolaylığı faktörlerini ele alan ölçeğin genel kullanılabilirlik düzeylerine ilişkin puan aralıklarına göre, 93-123 arası

puanlar 'yüksek' düzeyde kullanılabilir olarak kabul edilmektedir. Tablo 7'de yer alan veriler doğrultusunda, 'sinavanalizlerim.com' web sitesi, 113,70 puan ile katılımcılar tarafından yüksek düzeyde kullanılabilir bulunmuştur. Bu sonuç, web sitesinin çoktan seçmeli testlerin güvenilirlik analizine katkı sağlama amacına uygun bir şekilde tasarlandığını göstermektedir. Bu bulgulara dayanarak, akademisyenlerin çoktan seçmeli test uygulamalarında 'sinavanalizlerim.com' web sitesini kullanmasının, ölçme ve değerlendirme faaliyetlerine önemli bir katkı sağlayabileceği düşünülmektedir.

SONUÇ

Araştırmada, gezinme kolaylığı, tasarım kolaylığı, erişim kolaylığı ve kullanım kolaylığı gibi unsurlar değerlendirildiğinde, 'sinavanalizlerim.com' adresinin genel kullanılabilirlik düzeyinin yüksek olduğu ve çoktan seçmeli testlerin güvenilirliği açısından etkili bir performans sergilediği görülmüştür.

Gezinme kolaylığı faktörü kapsamında yapılan değerlendirmelerde, en yüksek puanın 'Menülerdeki başlıklar ile sayfa içeriklerinin tutarlılığı' (4. madde) ifadesine verildiği belirlenmiştir. Benzer şekilde, Göker'in (2019, s. 56) Yabancı Dil Olarak Türkçenin Öğretiminde Uzaktan Eğitim Web Sitelerinin Kullanılabilirlik Açısından İncelenmesi başlıklı çalışmasında aynı ölçeği kullanmış ve en yüksek puanın 1. ve 4. maddelere verildiği görülmüştür. Bu bulgular incelendiğinde, mevcut çalışma sonuçlarının literatürdeki önceki araştırmalarla uyumlu olduğu tespit edilmiştir.

Tasarım kolaylığı faktörü özelinde yapılan analizlerde, en yüksek puanın 'Kullanılan yazı tiplerinin tutarlılığı' kriterine verildiği tespit edilmiştir. Bu bulgu, Saputra ve arkadaşları (2022, s. 2588) tarafından çevrimiçi otel rezervasyon web siteleri üzerine gerçekleştirilen ampirik bir çalışmada elde edilen sonuçlarla örtüşmektedir. Söz konusu çalışmada, kullanıcı memnuniyetini etkileyen temel kullanılabilirlik faktörleri arasında arama kutularının düzeni, yazı tipi seçimleri ve menü tasarımlarının bulunduğu, ayrıca yazı tipi tercihinin kullanıcı memnuniyeti üzerinde olumlu bir etkisi olduğu vurgulanmıştır. Literatürle karşılaştırıldığında, bu çalışmanın sonuçlarının benzerlik taşıdığı görülmektedir.

Web sitesinin geliştirilmesi gereken yönleri arasında öncelikle, konu temelli soru eşleştirme sistemi öne çıkmaktadır. Bu sistem, soruların belirli konu başlıkları ile ilişkilendirilmesini sağlayarak öğrencilerin hangi konularda zorluk yaşadığını tespit etmeyi mümkün kılabilir. Böylece, öğrencilerin eksikliklerine odaklanılarak öğrenme sürecinin daha ayrıntılı bir şekilde analiz edilmesi sağlanabilir. Ayrıca, kazanım odaklı öğrenme sistemi ile öğretim elemanlarının öğrencilerde geliştirmek istedikleri öğrenme hedeflerini tanımlayabilecekleri bir yapı oluşturulabilir. Bu hedeflere uygun sorular aracılığıyla öğrencilerin kazanımlara ulaşp ulaşmadığı değerlendirilebilir. Bir diğer önemli geliştirme önerisi, yapay zekâ destekli geri bildirim mekanizmasıdır. Bu mekanizma, öğretmenlerin karşılaştıkları zorluklara yönelik anlık destek sunarak hem pedagojik rehberlik hem de içerik yönetimi açısından önemli bir katkı sağlayabilir. Son olarak, notlandırma ve üniversite sistemleri ile entegrasyon sağlanarak sınav sonuçlarının otomatik olarak akademik sistemlere aktarılması mümkün kılınabilir. Bu entegrasyon, öğretim elemanlarının değerlendirme süreçlerini kolaylaştırırken, öğrencilerin akademik performanslarının düzenli bir şekilde takip edilmesini destekleyecektir.

Araştırma ve Yayın Etiği Beyanı

Bu araştırma içerisinde sunmuş olduğumuz verileri, bilgileri ve dokümanları akademik ve etik ahlak kuralları çerçevesinde elde ettiğimizi, tüm bilgileri, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumuzu, bu çalışmada yararlandığımız eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimizi, kullanılan verilerde herhangi bir değişiklik yapmadığımızı, çalışmanın özgün olduğunu bildiririz. Aksi bir durumda aleyhimize doğacak tüm hak ve kayıplarını kabul ettiğimizi beyan ederiz.

Yazarların Makaleye Katkı Oranları

Bu makalede yazarlar eşit oranda katkıda bulunmuştur.

Etik Kurul İzni

Bu araştırma için etik kurul izni Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Bilimsel Araştırma ve Yayın Etiği Kurulu'nun 18.11.2024 tarih ve 316818 sayılı kararı ile alınmıştır.

Çıkar Beyanı

Bu çalışmanın hazırlanmasında çıkar çatışması yaşanacak herhangi bir durum bulunmamaktadır.

KAYNAKÇA

- Acar, T. (2018). Yazılı yoklamalarda puanlama yönteminin test ve madde istatistiklerine etkisi. *Elementary Education Online*, 17(2), 500-509. <https://doi.org/10.17051/ilkonline.2018.418890>
- Alpar, R. (2016). *Spor, sağlık ve eğitim bilimlerinden örneklerle uygulamalı istatistik ve geçerlik-güvenirlilik* (4. Baskı). Detay Yayıncılık.
- Arıkan, S., Çelen, Ü., Demirtaşlı, R. N., Gülleroğlu, H. D., Gültekin, S., Kilmen, S. ve Demir, S. B. (2017). *Eğitimde ölçme ve değerlendirme* (3. Baskı). Anı Yayıncılık.
- Aslan, K. (2018). *Eğitim bilimleri ölçme ve değerlendirme* (3. Baskı). Kamupark Yayıncılık.
- Atılğan, H., Kan, A. ve Doğan, N. (2014). *Eğitimde ölçme ve değerlendirme* (6. Baskı). Anı Yayıncılık.
- Aydın, A. (2013). *Çoktan seçmeli ölçme sonuçlarının bilgisayar yardımıyla analizi* [Yayımlanmamış yüksek lisans tezi]. Afyon Kocatepe Üniversitesi.
- Başol, G. (2016). *Eğitimde ölçme ve değerlendirme* (4. Baskı). Pegem Akademi.
- Baykul, Y., Kelecioğlu, H. ve Gebal. (1993). *Ders geçme ve kredi düzeni araştırması* (2. Baskı). Milli Eğitim Basımevi.
- Baykul, Y. (1997). *İstatistik metotlar ve uygulamalar* (2. Baskı). Anı Yayıncılık.
- Baykul, Y. ve Güzeller, C. O. (2013). *Sosyal bilimler için istatistik* (1. Baskı). Pegem Akademi.
- Benzer, A. ve Eldem, E. (2013). Türkçe ve edebiyat öğretmenlerinin ölçme ve değerlendirme araçları hakkında bilgi düzeyleri. *Kastamonu Eğitim Dergisi*, 21(2), 649-664.

- Birgin, A., & Özcan, H. (2022). 8. sınıf öğrencilerinin mevsimlerin oluşumu ile ilgili bilgilerini ölçmeye yönelik bir başarı testinin geliştirilmesi. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 24(1), 305-326. <https://doi.org/10.26468/trakyasobed.817792>
- Boz, C., Bostan Sarioğlu, A., & Özcan, H. (2022). Kimyasal bağlar konusunda başarı testi geliştirme çalışması. *Manisa Celal Bayar Üniversitesi Eğitim Fakültesi Dergisi*, 10(2), 149-166. <https://doi.org/10.52826/mcbuefd.1065432>
- Büyüköztürk, Ş., Çokluk, Ö. ve Köklü, N. (2013). Sosyal bilimler için istatistik (12. Baskı). Pegem Akademi.
- Çağlar, D. (1970). *Başarının ölçülmesi ve istatistik metotlarla değerlendirme* (3. Baskı). Çağdaş Eğitim Kitapları.
- Çakmak, Ö. (2008). Eğitimin ekonomiye ve kalkınmaya etkisi. *D.Ü. Ziya Gökalp Eğitim Fakültesi Dergisi*, (11), 33-41.
- Çakmak, E. K., Güneş, E., Çiftçi, S. ve Üstündağ, M. T. (2011). Web sitesi kullanılabilirlik ölçeğinin geliştirilmesi: Geçerlik, güvenirlik analizi ve uygulama sonuçları. *Pegem Eğitim ve Öğretim Dergisi*, 1(2), 31-40. <https://doi.org/10.14527/C1S2M4>
- Çakmaktepe, M. (2024, Kasım 20). sinavanalizlerim.com. <https://app.sinavanalizlerim.com> adresinden 21 Kasım 2024 tarihinde alınmıştır.
- Doğan, N. (2008). *Ölçme ve değerlendirme konu anlatımlı soru bankası* (1. Baskı). Uzman Kariyer Yayınları.
- Dündar, M. (2019). *Geliştirilen genel yetenek çoktan seçmeli ölçme aracının psikometrik özelliklerinin klasik test kuramı ve madde tepki kuramına göre incelenmesi* [Yayımlanmamış yüksek lisans tezi]. Hasan Kalyoncu Üniversitesi.
- Göker, M. (2019). *Yabancı dil olarak Türkçenin öğretiminde uzaktan eğitim web sitelerinin kullanılabilirlik açısından incelenmesi (3 dakikada Türkçe örneği)*. [Yayımlanmamış yüksek lisans tezi]. Sakarya Üniversitesi.
- Gömlüksiz, M., Bıçak, B., Çetin, B., Karaca, E., Yurdabakan, İ. ve Nartgün, Z. (2008). *Eğitimde ölçme ve değerlendirme* (2. Baskı). Nobel Akademik Yayıncılık.
- Güler, N. (2011). *Eğitimde ölçme ve değerlendirme* (2. Baskı). Pegem Akademi.
- Gürbüz, S. ve Şahin, F. (2018). *Sosyal bilimlerde araştırma yöntemleri* (4. Baskı). Seçkin.
- Karadağ, N. (2014). *Açık ve uzaktan eğitimde ölçme ve değerlendirme: Mega üniversitelerdeki uygulamalar* [Yayımlanmamış yüksek lisans tezi]. Anadolu Üniversitesi.
- Kasap, Y. (2012). *Ölçme ve değerlendirme* (1. Baskı). Beyaz Kalem Yayıncılık.
- Özçelik, D. A. (1998). *Ölçme ve değerlendirme* (2. Baskı). ÖSYM Yayınları.
- Özen, Y., Gülaçtı, F. ve Kandemir, M. (2006). Eğitim bilimleri araştırmalarında geçerlik ve güvenirlik sorunsalı. *Erzincan Eğitim Fakültesi Dergisi*, 8(1), 69-88.
- Rehman, A., Aslam, A. ve Hassan, S. H. (2018). Item analysis of multiple-choice questions. *Pakistan Oral and Dental Journal*, 38(2), 291-293.

- Saputra, D., Syah, E. A. ve Darnis, F. (2022). Usability çoktan seçmeli ölçme aracının on the simponik website using the system usability scale (SUS). *Sinkron*, 7(4), 2584-2592.
- Shete, A. N., Kausar, A., Lakhkar, K. ve Khan, S. T. (2015). Item analysis: An evaluation of multiple-choice questions in Physiology examination. *Journal of Contemporary Medical Education*, 3(3), 106-109. <https://doi.org/10.5455/jcme.20151011041414>
- Singh, T., Gupta, P. ve Singh, D. (2021). *Principles of medical education*. Jaypee Brothers Medical Publishers.
- Tavakol, M. ve Dennick, R. (2011). Post-examination analysis of objective tests. *Medical Teacher*, 33(6), 447-458. <https://doi.org/10.3109/0142159X.2011.564682>
- Tekin, H. (1977). *Eğitimde ölçme ve değerlendirme* (4. Baskı). Mars Matbaası.
- Tekin, H. (1979). *Eğitimde ölçme ve değerlendirme* (2. Baskı). Yargı Kitap ve Yayınevi.
- Telek, C. (2013). *Kullanılabilirlik kavramı, tasarım süreci içindeki yeri ve benzer tasarım yaklaşımları ile ilişkisi* [Yayımlanmamış yüksek lisans tezi]. Mimar Sinan Güzel Sanatlar Üniversitesi.
- Tomak, L. (2013). *Madde analizi ve tıp fakültesi sınavlarının değerlendirilmesi* [Yayımlanmamış doktora tezi]. Ondokuz Mayıs Üniversitesi.
- Uğan, Ç. ve Akbolat, M. (2023). Şiddete maruz kalmanın iş performansına etkisinde şiddet görme korkusunun aracı rolü. *Journal of Organizational Behavior Studies*, 3(1), 25-37. <https://doi.org/10.5281/zenodo.8098890>
- Uysal, M. (2019). *Eğitimde ölçme ve değerlendirme* (3. Baskı). Nobel Akademik Yayıncılık.
- Yılmaz, H. (1996). *Eğitimde ölçme ve değerlendirme* (9. Baskı). Öz Eğitim Yayınları.
- Yöntem, M. K. (2022). *Uygulama örnekleri ile etkili üniversite eğitimi* (1. Baskı). Eğitim Yayınevi.

EXTENDED ABSTRACT

This study focuses on evaluating the usability of an online platform designed for the reliability analysis of multiple-choice tests. The platform, *sinavanalizlerim.com'*, developed for statistical analysis of test parameters, was assessed based on user experience using the Website Usability Scale. The research sample comprised 50 academics, who provided evaluations regarding ease of navigation, design consistency, accessibility, and usability. The study aimed to determine the platform's effectiveness in supporting assessment and evaluation processes in education.

The findings indicate that the platform was positively received by users, particularly in terms of design consistency, menu and content alignment, and rapid access features. These aspects were identified as strong contributors to a positive user experience. However, while the current system was considered effective, several recommendations were made for further improvement. Specifically, topic-based question matching, AI-supported feedback mechanisms, and grading integration were suggested as potential enhancements. These improvements would enable the platform to offer more advanced assessment capabilities, increasing its effectiveness in educational settings.

From a methodological perspective, the study employed a survey-based approach, with data collected from 50 faculty members. The survey results were analyzed using SPSS software, and factor analyses were performed to examine website usability. The Cronbach's Alpha coefficient was calculated as 0.885, indicating that the usability scale used in the study is highly reliable. The statistical analysis validated the consistency and reliability of the collected data, reinforcing the credibility of the study's findings.

Multiple-choice tests are widely used in education due to their efficiency in evaluating students' knowledge and competencies in a structured manner. These tests provide objective scoring, reduced grading bias, and large-scale applicability, making them a preferred assessment tool in academic and standardized examinations. However, despite their advantages, ensuring the reliability of multiple-choice tests remains a critical challenge. The accuracy of these assessments largely depends on the statistical properties of the test items, including item difficulty, discrimination indices, and reliability coefficients. The platform evaluated in this study aims to address these challenges by providing automated reliability analysis tools that assist educators in refining their assessments.

One of the critical aspects of the study was the usability evaluation of the platform. Usability is a key determinant of the effectiveness of educational technologies, as a system that is difficult to navigate or access can significantly hinder its adoption. The study assessed usability through four main factors: ease of navigation, design consistency, accessibility, and overall usability. The results revealed that the platform's design consistency, structured menus, and responsive interface were highly appreciated by users. These factors contributed to a seamless user experience, ensuring that educators could easily access and interpret test analysis results. However, certain limitations were also identified, particularly in terms of question matching, personalized feedback, and system integration with grading mechanisms.

One of the key findings was the necessity of integrating AI-driven feedback mechanisms to enhance the interpretability of test results. AI-powered systems could automatically analyze student performance patterns, generate tailored feedback, and provide insights on potential improvements. This would allow educators to gain a deeper understanding of student learning trends and adjust their teaching strategies accordingly. Additionally, topic-based question matching was identified as a crucial feature that could improve the alignment of test questions with learning objectives. By categorizing test items based on their content, the platform could help educators ensure that their assessments are comprehensive, balanced, and aligned with curriculum standards.

Another significant recommendation was the integration of grading systems within the platform. Currently, assessment tools often operate in isolation, requiring educators to manually transfer test results into grading software or academic management systems. Automated grading integration would streamline this process, reducing administrative workload and enhancing efficiency. Moreover, such an integration would allow for real-time tracking of student progress, providing valuable data for both instructors and educational administrators.

Despite these areas for improvement, the study highlights the potential of 'sinavanalizlerim.com' as a valuable tool for academic assessment. The platform's ability to

perform detailed item analyses, compute reliability coefficients, and visualize statistical parameters makes it a powerful asset for educators seeking to enhance their assessment strategies. The study also reinforces the importance of usability in educational technology, emphasizing that the success of digital tools is largely dependent on user experience, interface design, and system accessibility.

In conclusion, the findings suggest that 'sinavanalizlerim.com' serves as an effective digital platform for the reliability analysis of multiple-choice tests. While its current features provide significant benefits, further enhancements, particularly in AI-driven feedback, question matching, and grading integration, would further increase its value and usability. This study contributes to the broader discourse on digital assessment tools in education, offering insights into how usability and technological advancements can shape the future of academic evaluation. By implementing the recommended improvements, the platform could become a leading solution in the field of educational measurement and assessment, supporting instructors in developing more reliable, valid, and data-driven Methodologies of multiple-choice assessment instruments.