



**ÇOKTAN SEÇMELİ TESTLERİN
PSİKOMETRİK ANALİZİ İÇİN WEB
TABANLI BİR ARAÇ GELİŞTİRME VE
KULLANILABİLİRLİK DEĞERLENDİRMESİ**

YÜKSEK LİSANS TEZİ

Mehmet ÇAKMAKTEPE

Danışman

Doç. Dr. Gür Emre GÜRAKSIN

**İNTERNET VE BİLİŞİM TEKNOLOJİLERİ YÖNETİMİ
ANABİLİM DALI**

Mayıs 2025

AFYON KOCATEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

ÇOKTAN SEÇMELİ TESTLERİN
PSİKOMETRİK ANALİZİ İÇİN WEB TABANLI BİR ARAÇ
GELİŞTİRME VE KULLANILABİLİRLİK DEĞERLENDİRMESİ

Mehmet ÇAKMAKTEPE

Danışman

Doç. Dr. Gür Emre GÜRAKSIN

İNTERNET VE BİLİŞİM TEKNOLOJİLERİ YÖNETİMİ
ANABİLİM DALI

Mayıs 2025

TEZ ONAY SAYFASI

Mehmet ÇAKMAKTEPE tarafından hazırlanan “Çoktan Seçmeli Testlerin Psikometrik Analizi İçin Web Tabanlı Bir Araç Geliştirme Ve Kullanılabilirlik Değerlendirmesi” adlı tez çalışması lisansüstü eğitim ve öğretim yönetmeliğinin ilgili maddeleri uyarınca 26 / 05 / 2025 tarihinde aşağıdaki jüri tarafından **oy çokluğu** ile Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü **Anabilim Dalı Adı Anabilim Dalı’nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman : Doç Dr. Gür Emre GÜRAKSIN

Başkan : Doç. Dr. Semiral ÖNCÜ
Balıkesir Üniversitesi, Necatibey Eğitim Fakültesi **İmza**

Üye : Dr. Öğr. Üyesi Ertuğrul ERGÜN
Afyon Kocatepe Üniversitesi, Uzaktan Eğitim MYO **İmza**

Afyon Kocatepe Üniversitesi
Fen Bilimleri Enstitüsü Yönetim Kurulu’nun
..... /..... /..... tarih ve
..... sayılı kararıyla onaylanmıştır.

.....

Prof. Dr. Bekir YALÇIN

Enstitü Müdürü

BİLİMSEL ETİK BİLDİRİM SAYFASI

Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- Atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

26 /05 /2025

İmza

Mehmet ÇAKMAKTEPE

ÖZET

Yüksek Lisans Tezi

ÇOKTAN SEÇMELİ TESTLERİN PSİKOMETRİK ANALİZİ İÇİN WEB TABANLI BİR ARAÇ GELİŞTİRME VE KULLANILABİLİRLİK DEĞERLENDİRMESİ

Mehmet ÇAKMAKTEPE

Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü

İnternet ve Bilişim Teknolojileri Yönetimi Anabilim Dalı

Danışman: Doç. Dr. Gür Emre GÜRAKSIN

Bu tez çalışmasında, çoktan seçmeli testlerin psikometrik analizini çevrimiçi ortamda gerçekleştirebilecek bir web tabanlı sistem geliştirilmiş ve bu sistemin kullanılabilirliği değerlendirilmiştir. Laravel altyapısı ile geliştirilen yazılım; HTML, CSS, JavaScript, jQuery ve MySQL teknolojileriyle desteklenmiş, kullanıcı dostu bir arayüz aracılığıyla madde güçlüğü, ayırt edicilik, ortalama başarı ve KR-20/KR-21 gibi güvenilirlik katsayılarını otomatik olarak hesaplayabilir hâle getirilmiştir. Görselleştirilmiş analiz çıktıları ve çıktı alma seçenekleri ile kullanıcıların ölçme-değerlendirme süreçlerini daha hızlı ve etkili şekilde yürütmeleri hedeflenmiştir. Sistemin kullanılabilirliği, beş akademisyenin katıldığı görev temelli testler ve Likert ölçekli anketlerle ölçülmüştür. Bulgular, kullanıcıların büyük oranda görevleri başarıyla tamamladığını, ancak sınav yükleme ve analiz sonuçlarını yorumlama aşamalarında bazı zorluklarla karşılaştıklarını ortaya koymuştur. Pasta grafiklerle sunulan görev bazlı zorluk verileri, sistemin geliştirilmesi gereken yönlerini belirlemede yol gösterici olmuştur. Buna ek olarak, sistemin genel kullanıcı deneyimi düzeyini belirlemek amacıyla 50 akademisyenin katılımıyla Web Sitesi Kullanılabilirlik Ölçeği uygulanmıştır. Elde edilen verilere göre sistem, toplam 135 üzerinden ortalama 113,70 puan olarak yüksek düzeyde kullanılabilir olarak değerlendirilmiştir. Katılımcılar, özellikle sistemin görsel yapısını, menü başlıklarının işlevselliğini ve analiz ekranlarının açıklığını olumlu bulmuştur. En yüksek ortalama puan, “madde analizi ekranı soruları detaylı incelemeye olanak tanıyor” ifadesinden elde edilmiştir. Bu bulgular, sistemin sadece teknik işlevsellik açısından

değil, aynı zamanda kullanıcı deneyimi bakımından da güçlü ve tatmin edici bir düzeye ulaştığını göstermektedir. Sonuç olarak geliştirilen sistem, eğitimde dijital ölçme-değerlendirme uygulamalarının etkinliğini artırmakta; öğretmenlerin iş yükünü azaltmakta ve ölçme araçlarının daha güvenilir biçimde değerlendirilmesine katkı sağlamaktadır. Tez, eğitim teknolojileri alanında hem yazılım geliştirme hem de kullanıcı odaklı değerlendirme açısından bütüncül bir katkı sunmaktadır.

2025, xii + 101 sayfa

Anahtar Kelimeler: Web Tabanlı, PHP, Ölçme ve Değerlendirme, Güvenirlik, Eğitim Sistemi, Kullanılabilirlik.

ABSTRACT

M.Sc. Thesis

RELIABILITY ANALYSIS OF MULTIPLE-CHOICE TESTS IN A WEB-BASED ENVIRONMENT

Mehmet ÇAKMAKTEPE

Afyon Kocatepe University

Graduate School of Natural and Applied Sciences

Department of Internet and Information Technology Management

Supervisor: Assoc. Prof. Gür Emre GÜRAKSIN

In this master's thesis, a web-based system was developed to perform the psychometric analysis of multiple-choice tests in an online environment, and the usability of the system was evaluated. The software was built on the Laravel framework and supported by technologies such as HTML, CSS, JavaScript, jQuery, and MySQL. Through a user-friendly interface, the system is capable of automatically calculating item difficulty, item discrimination, average success rate, and reliability coefficients such as KR-20 and KR-21. The system also provides visualized analysis outputs and export options, aiming to facilitate and accelerate assessment processes for users. The usability of the system was evaluated through task-based tests and Likert-scale questionnaires with the participation of five academic users. The findings revealed that users successfully completed most of the tasks; however, certain difficulties were encountered in tasks related to uploading exams and interpreting analysis results. Task-specific difficulty levels were presented in a pie chart, serving as a guide for identifying areas of the system in need of improvement. In addition, a Web Site Usability Scale was administered to a larger sample of 50 academic users to assess overall user experience. According to the data obtained, the system achieved an average score of 113.70 out of 135, indicating a high level of usability. Participants particularly highlighted the effectiveness of the visual structure, the clarity and functionality of menu labels, and the transparency of the analysis screens. The highest-rated statement was “The item analysis screen enables detailed examination of questions.” These findings indicate that the system is not only functionally effective but

also provides a satisfactory and robust user experience. As a result, the developed system enhances the efficiency of digital assessment practices in education, reduces the workload of educators, and contributes to more reliable evaluation of test instruments. This thesis offers a comprehensive contribution to the field of educational technologies in terms of both software development and user-centered evaluation.

2025, xii + 101 pages

Keywords: Web-Based, PHP, Assessment and Evaluation, Reliability, Educational System, Usability.



TEŐEKKÖR

Bu arařtırmanın konusu, deneysel alıřmaların y nlendirilmesi, sonuların deęerlendirilmesi ve yazımı ařamasında yapmıř olduęu b y k katkılarından dolayı tez danıřmanım Sayın Do. Dr. G r Emre G RAKSIN'a, arařtırma ve yazım s resince yardımlarını esirgemeyen Sayın Muharrem K rřat YALIM'a her konuda  neri ve eleřtirileriyle yardımlarını g rd ę m hocalarıma ve bu s rete desteęini esirgemeyen Sayın Sedanur ORC N'e teőekk r ederim.

Bu arařtırma boyunca maddi ve manevi desteklerinden dolayı aileme teőekk r ederim.

Mehmet AKMAKTEPE

Afyonkarahisar 2025

İÇİNDEKİLER DİZİNİ

	Sayfa
ÖZET	i
ABSTRACT	iii
TEŞEKKÜR	v
İÇİNDEKİLER DİZİNİ.....	vi
SİMGELER ve KISALTMALAR DİZİNİ	viii
ŞEKİLLER DİZİNİ	ix
ÇİZELGELER DİZİNİ.....	x
RESİMLER DİZİNİ	xi
1. GİRİŞ.....	1
1.1 Literatürde Mevcut Sistemler	5
1.1.1 TestAn (Ali Aydın)	5
1.1.2 TAP – Test Analysis Program (Gordon P. Brooks, Ohio University)	5
1.1.3 Iteman (Assessment Systems Corporation)	6
2. LİTERATÜR TARAMA.....	7
2.1 Ölçme ve Değerlendirme	9
2.1.1 Ölçme	9
2.1.2 Değerlendirme.....	11
2.1.3 Ölçme Araçlarının Sahip Olması Gereken Temel Nitelikler	12
2.2 Test ve Madde Analizi	15
2.2.1 Temel Kavramlar	15
2.2.2 Test Analizi	16
2.2.3 Madde Analizi.....	23
2.3 Çoktan Seçmeli Testler	30
2.4 Kullanılabilirlik Kavramı ve Tespiti.....	32
2.4.1 Kullanılabilirliğin Bileşenleri.....	33
2.4.2 Kullanılabilirlik Ölçümleri.....	33
2.5 Konu İle İlgili Daha Önce Yapılmış Çalışmalar	34
3. MATERYAL ve METOT	42
3.1 Araştırmanın Genel Uygulama Süreci	42
3.2 Materyaller.....	42

3.2.1 HTML	43
3.2.2 CSS.....	43
3.2.3 BOOTSTRAP	44
3.2.4 JavaScript	44
3.2.5 Laravel.....	45
3.2.6 jQuery.....	45
3.2.7 MySQL.....	46
3.3 Test Ve Madde Analizi WEB Sitesi	46
3.3.1 Web Sitesi Kullanılabilirlik Ölçeği: Yapı, Uygulama ve Değerlendirme Yöntemi	46
3.3.2 Görev Temelli Kullanılabilirlik Testi: Yöntem ve Uygulama	47
3.3.3 Geliştirme Sürecine Temel Oluşturan Model	48
3.3.4 Sitenin Ara Yüzü ve Kullanımı.....	48
4. BULGULAR	65
4.1 Araştırmanın Amacı ve Önemi	65
4.2 Araştırmanın Hipotezleri	65
4.3 Araştırmanın Evreni/Örnekleme ve Kısıtları.....	66
4.4 Araştırmanın Veri Toplama Yöntemi	67
4.5 Araştırmanın Veri Analiz Yöntemleri ve Güvenilirliği	67
4.6 Web Sitesi Kullanılabilirlik Ölçeği'nden Elde Edilen Bulgular.....	68
4.7 Görev Temelli Kullanılabilirlik Değerlendirme Formunun Sonuçları	72
5. TARTIŞMA VE SONUÇ	74
5.1 Uygulamaya Dayalı Öneriler	75
5.2 Web Sayfasının Eksik Veya Daha Sonra Geliştirilebilecek Yönleri	76
5.3 Sınırlılıklar ve Gelecek Araştırmalar	77
6. KAYNAKLAR.....	78
ÖZGEÇMİŞ.....	93
EKLER	94

SİMGELELER ve KISALTMALAR DİZİNİ

Kısaltmalar

API	Application Programming Interface
ASC	Assassment Systems Corporation
CSS	Cascading Style Sheets
DOM	Document Object Model
HTML	HyperText Markup Language
KR-20	Kuder Richardson 20
KR-21	Kuder Richardson 21
MVC	Model View Controller
OLAP	On Line Analytical Processing
OLTP	On Line Transaction Processing
ÖDGP	Ölçme ve Değerlendirme Gelişim Programı
PDF	Portable Document Format
PHP	Hypertext Preprocessor
SQL	Structured Query Language
WHATWG	Web HyperText Application Working Group
XHTML	Extensible HyperText Markup Language
XML	Extensible Markup Language

ŞEKİLLER DİZİNİ

	Sayfa
Şekil 1.1 Çoktan seçmeli maddenin yapısı.....	3
Şekil 2.1 Eğitim sistemi bileşenleri arasındaki ilişki	7
Şekil 4.1 Görev bazlı ortalama zorluk dağılımı.	73



ÇİZELGELER DİZİNİ

	Sayfa
Çizelge 2.1 20 öğrencinin 40 maddelik bir sınavdan aldığı notlar.	17
Çizelge 2.2 Test Maddelerinin Ayırıcılık Gücü İndeksine Dayalı Değerlendirilmesi. 26	
Çizelge 2.3 Nokta Çift Serili Korelasyon Örnek Veriler	29
Çizelge 4.1 Afyon Kocatepe Üniversitesi’nde bulunan akademisyenlerin demografik özelliklerine göre dağılımı.	66
Çizelge 4.2 Katılımcıların ‘Gezinme Kolaylığı’ Faktörünü Oluştan Maddelere İlişkin Algılarının Aritmetik Ortalaması.....	68
Çizelge 4.3 Katılımcıların tasarım faktörünü oluştan maddelere ilişkin algılarının aritmetik ortalaması.	70
Çizelge 4.4 Katılımcıların erişim kolaylığı faktörünü oluştan maddelere ilişkin algılarının aritmetik ortalaması.	70
Çizelge 4.5 Katılımcıların kullanım kolaylığı faktörünü oluştan maddelere ilişkin algılarının aritmetik ortalaması.	71
Çizelge 4.6 Katılımcıların sinavanalizlerim.com’un kullanılabilirliğine yönelik algıları.	72

RESİMLER DİZİNİ

	Sayfa
Resim 3.1 Kullanıcı giriş sayfası.	49
Resim 3.2 Ana sayfa.	50
Resim 3.3 Gezinme çubuğu.	51
Resim 3.4 Kategori işlemleri.....	51
Resim 3.5 Kategori ekle.....	52
Resim 3.6 Kategoriler.	52
Resim 3.7 Tüm kategoriler.....	52
Resim 3.8 Sınav işlemleri.	53
Resim 3.9 Sınav ekle.....	53
Resim 3.10 Sınavlar.	53
Resim 3.11 Sınav düzenle.....	54
Resim 3.12 Sınav sil.....	54
Resim 3.13 Sınav ayarlar.	54
Resim 3.14 Sınav ayarlar sayfası ilk kısım.	55
Resim 3.15 Sınav ayarlar sayfası ikinci kısım.	56
Resim 3.16 Sınavı analiz et.....	56
Resim 3.17 Sınav analiz sayfası.....	57
Resim 3.18 Öğrenci cevapları detaylar.	57
Resim 3.19 Değerlerin detaylar sekmesi.....	58
Resim 3.20 Madde analizi.....	59
Resim 3.21 Madde analiz sayfası.....	59
Resim 3.22 Madde cevap ve çeldiriciler.	60
Resim 3.23 Paylaşımlar.....	61
Resim 3.24 PDF indirme.....	61
Resim 3.25 Madde analizi raporu.	62
Resim 3.26 Test analizi ilk sayfa.	62
Resim 3.27 Test analizi ikinci sayfa.	63
Resim 3.28 Profil aracı.....	63
Resim 3.29 Profil sayfası.	64
Resim 3.30 Şifre değiştir sekmesi.....	64

Resim 4.1 Başlıklar ve sayfa gezinme kolaylığı.	69
Resim 4.2 Yazı tipleri.	70
Resim 4.3 Madde analizi sonuçları.	71



1. GİRİŞ

Eğitim, bireyin deneyimlerinden yola çıkarak, bilinçli ve amaçlı bir biçimde istenen davranış değişikliklerini kazandırmayı hedefleyen planlı bir süreçtir. İnsanı belirli davranışlarla donatmak, diğer bir ifadeyle eğitmek, insanlık tarihinin en temel amaçlarından biri olmuştur. Günümüzde eğitim olgusu, giderek daha karmaşık bir yapıya bürünmekte ve etkisi daha yoğun şekilde hissedilmektedir. İnsan, en basit düzeydeki somut ilişkilerden en karmaşık düzeydeki soyut ilişkilere kadar, her türlü sürecin odağında yer alır. Devlet yönetimi, sanayi kuruluşlarının planlanması ve işletilmesi, ürünlerin dağıtımı ve tüketimi, doğal kaynakların işlenmesi, altyapı projelerinin hayata geçirilmesi ve toplumsal sorunların çözümü gibi geniş bir yelpazedeki sorumluluklar, insanın yetkinliğine bağlıdır. Bu bağlamda, bireyleri tutarlı davranışlarla ve etkili problem çözme bilgi ve becerileriyle donatmak bir zorunluluktur. Söz konusu hedefe ulaşmak ise yalnızca sistemli ve nitelikli bir eğitim süreci ile mümkün olabilir (Çakmak, 2008).

Eğitim süreci ne kadar önemliyse, bu süreçte gerçekleştirilen değerlendirmeler de aynı derecede önemlidir. Öğrencilerin, aldıkları eğitimin sonucunda uygun değerlendirme yöntemleriyle analiz edilmesi ve bu değerlendirmeler doğrultusunda gelişimlerinin izlenmesi gerekmektedir (Isoré 2009). Bu yaklaşım, hem öğrencilerin bireysel ilerlemelerini desteklemek hem de eğitim süreçlerinin etkinliğini artırmak açısından kritik bir role sahiptir.

Uzun yıllar boyunca eğitim sistemimizde, öğretim etkinlikleri genellikle öğretmen odaklı bir yaklaşıma dayalı olarak tasarlanmıştır. Bu süreçte, dersler büyük ölçüde anlatım yöntemiyle işlenmiş, öğrenciler ise bilgiyi pasif bir şekilde ezberleyerek kaydeden bireyler olarak görülmüştür. Değerlendirme süreçleri ise, sürece değil sonuca odaklanmış, yalnızca "geçti" ya da "kaldı" gibi yargılarla sınırlı kalmıştır. Bu anlayış çerçevesinde öğrenciler, bilgiyle doldurulması gereken boş kaplar olarak değerlendirilmiş ve öğretmenlerin bu kapları ne kadar çok bilgiyle doldurursa, öğrencilerin o kadar başarılı olacağı varsayılmıştır (Kutlu vd., 2024).

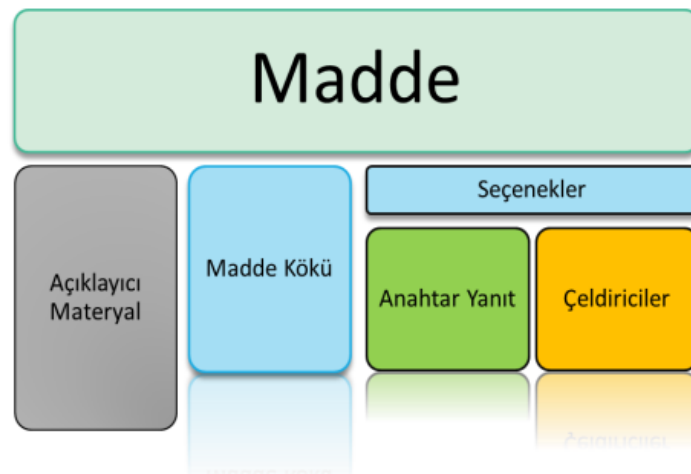
Eğitim sürecinin vazgeçilmez unsurlarından biri, bilişsel, alanlara ilişkin özelliklerin ölçülmesidir. Bilişsel alan, bilgi ve bu bilgiden türeyen zihinsel süreçleri kapsamaktadır (Bloom vd. 1956). Eğitim uygulamalarında genellikle bilişsel davranışların kazandırılmasına ve bu davranışların ne ölçüde gerçekleştirildiğinin belirlenmesine öncelik verilmektedir (Bao ve Redish 2002). Bu amaçla kullanılan ölçme araçları, genellikle bilişsel alan taksonomilerine dayalı olarak hazırlanmaktadır ve bu alandaki en yaygın taksonomi, Bloom Taksonomisi'dir. Bloom Taksonomisi, bilişsel alanı basitten karmaşığa doğru sıralanmış ve hiyerarşik bir yapı gösteren altı basamakta ele almaktadır: bilgi, kavrama, uygulama, analiz, sentez ve değerlendirme. Bu basamaklardan bilgi, kavrama ve uygulama alt düzey bilişsel süreçler; analiz, sentez ve değerlendirme ise üst düzey bilişsel süreçler olarak sınıflandırılmaktadır ve her bir basamak, bir sonraki basamağın ön koşulunu oluşturmaktadır (Krathwohl 2002). Güncellenmiş Bloom Taksonomisi'nde bu basamaklar, hatırlama, anlama, uygulama, analiz, değerlendirme ve yaratma şeklinde yeniden adlandırılarak, bilişsel süreçlerin daha geniş bir bağlamda ve modern yaklaşımlara uygun şekilde ele alınması hedeflenmiştir (Baran 2020).

Ölçme ve değerlendirmenin doğru bir şekilde uygulanmasının, öğrenmenin kalitesini ve kalıcılığını artırdığı bilinmektedir (Crooks 1988). Bu bağlamda, ölçme ve değerlendirme kavramlarının açıklanması önem arz etmektedir. Ölçme, bir niteliğin gözlemlenmesi ve bu gözlem sonucunun sayılarla veya sembollerle ifade edilmesi işlemidir; değerlendirme ise ölçme sonuçlarının belirli bir ölçüte dayandırılarak bir değer yargısına ulaşılması sürecidir (Sivakumar ve Thirumoorthy 2019). Eğitim ve öğretim süreçlerinde ölçme ve değerlendirme, bu süreçlerin sürekli izlenmesine olanak sağlayarak her aşamada ortaya çıkan sorunların tespit edilmesi ve düzenlenmesi için fırsat sunmaktadır. Ölçme işleminin temel amacı, öğrencilerin kazanımlara ulaşma düzeylerini belirlemek ve yerinde, zamanında önlemler alarak eğitim sürecini iyileştirmektir. Bu sayede, eğitim programının işleyişine dair yeterli ve uygun dönüt sağlanmakta, gerekli düzeltmeler yapılabilmekte ve böylelikle eğitimin kalitesini kontrol etmek mümkün hale gelmektedir (Benzer ve Eldem 2013).

Ölçme ve Değerlendirmede önemli olan ölçme araçlarından bir tanesi de çoktan seçmeli testlerdir. Çoktan seçmeli testler, birden fazla seçenek arasından yalnızca bir doğru cevabın seçilmesi esasına dayanan maddelerden oluşan bir değerlendirme aracıdır. Eğitim

sisteminin farklı boyutlarında, sınıf içi değerlendirmelerden ulusal ve uluslararası sınavlara kadar geniş bir kullanım alanına sahiptir. Bu testler, objektif bir puanlama imkanı sunarak, öğrencilerin bilgi düzeylerini, yeteneklerini ve başarılarını etkili, güvenilir ve pratik bir şekilde ölçme olanağı sağlamaktadır (Yöntem 2022). Günümüzde, öğrenci seçme sınavlarında ve okullarda en sık tercih edilen ölçme araçlarından biri çoktan seçmeli testlerdir (Kuran ve Kanatlı 2009).

Şekil 1.1’de görüldüğü üzere, çoktan seçmeli test maddeleri, sorunun temel bileşeni olan madde kökü ve cevap alternatiflerinden oluşmaktadır. Madde kökü, sorunun tanıtımını yapmak için gerekli bilgileri içeren ve genellikle test maddesinin en uzun kısmını oluşturan ana bölümdür. Bu bölüm, bir soru cümlesi ya da tamamlanması gereken eksik bir ifade şeklinde yapılandırılabilir. Madde kökünü tamamlayan seçenekler, öğrencilerin doğru cevabı belirlemesi için sunulan ifadelerdir; yanlış cevapları temsil eden seçenekler ise çeldirici olarak adlandırılmaktadır. Çeldiriciler, öğrenme hedeflerine ulaşamayan öğrenciler için makul çözümler gibi görünerek dikkatlerini dağıtırken, öğrenme hedefini gerçekleştiren öğrenciler tarafından mantıksız olarak değerlendirilmeli ve yalnızca doğru seçenek anlamlı bulunmalıdır. Bu bağlamda, çoktan seçmeli madde yazımında en önemli unsurlardan biri, ayırt ediciliği yüksek çeldiriciler geliştirmektir (Gierl vd. 2017). Ayrıca, doğru cevabın konu hakkında bilgi sahibi öğrenciler tarafından kolayca tanınabilir olması gerektiği unutulmamalıdır. Ancak, doğru yanıtların doğrudan kaynak kitaplardan alıntı yapılarak yazılmaması önemlidir; aksi takdirde bu durum, öğrencilerin gerçek bilgisini ölçmek yerine ezber yeteneğini ön plana çıkarabilir (Yöntem 2022).



Şekil 1.1 Çoktan seçmeli maddenin yapısı.

Bir testin kapsamlı bir şekilde değerlendirilmesi için testin ortalaması, güvenilirlik katsayısı, ortalama güçlüğü, standart sapması, varyansı, merkezi yığılma ve dağılım ölçüleri gibi temel istatistiksel değerlerin hesaplanması gereklidir. Aynı şekilde, maddelerle ilgili değerlendirme yapabilmek için madde güçlük indeksi, madde ayırtıcılık gücü indeksi, madde varyansı, standart sapması, madde güvenilirlik katsayısı ve korelasyonu gibi ayrıntılı istatistiksel analizlerin gerçekleştirilmesi şarttır (Kim ve Park 2023). Bu tür hesaplamalar, testin genel yapısının ve maddelerin ölçme yeterliliğinin objektif bir şekilde değerlendirilmesine olanak tanır (Aydın 2013).

Tez çalışması kapsamında, bahsedilen istatistiksel değerlerin kağıt-kalem yöntemiyle hesaplanmasının zorlukları ve eğitim sistemimizde bu hesaplamaların tam anlamıyla gerçekleştirilememesi göz önünde bulundurularak (Kim vd. 2024), bu süreci kolaylaştırmak ve hesaplamaları otomatikleştirmek amacıyla bir web sayfası geliştirilmiştir. Bu web sayfası, kullanıcıların istatistiksel analizleri hızlı ve doğru bir şekilde yapabilmelerini sağlayarak test değerlendirme süreçlerini daha etkin hale getirmeyi hedeflemektedir.

Bu tez, çoktan seçmeli testlerin web ortamında güvenilirlik analizine odaklanarak, eğitimde ölçme ve değerlendirme süreçlerinin dijitalleşmesi açısından önemli bir boşluğu doldurmaktadır. Geleneksel yöntemlerle gerçekleştirilmesi zaman alıcı ve karmaşık olan istatistiksel analizlerin, geliştirilen web tabanlı platform aracılığıyla hızlı, doğru ve kullanıcı dostu bir şekilde uygulanabilmesi, eğitimde kaliteyi artırma ve değerlendirme süreçlerini modernize etme yönünde güçlü bir katkı sunmaktadır. Özellikle madde analizine ilişkin metriklerin (madde güçlük indeksi, ayırt edicilik gücü vb.) dijital ortamda otomatik olarak hesaplanabilmesi, öğretmen ve eğitimcilerin karar alma süreçlerini desteklemekte; böylece hem öğrencilerin başarı düzeylerinin daha doğru değerlendirilmesi hem de sınavların ölçme geçerliği açısından güçlendirilmesi sağlanmaktadır. Bu yönüyle tez, ölçme ve değerlendirme alanında teknolojik entegrasyonu teşvik eden özgün bir akademik katkı sunmaktadır.

Bu çalışma kapsamında aşağıdaki araştırma sorularına yanıt aranmıştır:

1. Web tabanlı olarak geliştirilen "sinavanalizlerim.com" sistemi, çoktan seçmeli testlerin güvenilirlik analizini ne ölçüde etkili bir şekilde gerçekleştirebilmektedir?

2. Geliştirilen sistem, kullanıcıların (öğretim elemanları/akademisyenler) sınav analiz süreçlerinde yaşadığı zorlukları gidermede ne derece işlevsel bir destek sunmaktadır?
3. Sistemin kullanıcı arayüzü, erişilebilirlik ve kullanılabilirlik açısından mevcut benzer ölçme-değerlendirme yazılımları ile karşılaştırıldığında ne tür avantajlar sağlamaktadır?
4. Kullanıcıların sistemle etkileşimleri sonucunda elde edilen ölçme verileri doğrultusunda, sistemin kullanılabilirlik düzeyi nedir ve bu düzey kullanıcı deneyimi açısından nasıl yorumlanabilir?
5. Sistem, öğretim elemanlarının ölçme-değerlendirme uygulamalarına ne tür pedagojik katkılar sunmaktadır ve bu katkılar geliştirilmeye açık hangi yönlere sahiptir?

1.1 Literatürde Mevcut Sistemler

Test ve madde analizine yönelik yazılımlar arasında, Türkiye’de Ali Aydın tarafından geliştirilen TestAn (C# tabanlı ve masaüstü), uluslararası düzeyde ise Ohio Üniversitesi’nden Gordon P. Brooks imzasını taşıyan TAP (ücretsiz Windows uygulaması) ve Assessment Systems Corporation tarafından geliştirilen ticari yazılım IteMan bulunmaktadır. TAP, klasik test teorisine dayalı güvenilirlik ve madde analizlerini kullanıcı dostu bir ara birimle desteklemekteyken, IteMan geniş grafik ve teknik raporlama özellikleriyle öne çıkmaktadır .

1.1.1 TestAn (Ali Aydın)

Afyon Kocatepe Üniversitesi’nden Ali Fikret Aydın tarafından C# tabanlı masaüstü uygulama olarak geliştirilen “TestAn”, çoktan seçmeli testlerin güvenilirlik hesaplamaları ve madde analizlerini gerçekleştirme özelliği taşımaktadır (Aydın 2010).

1.1.2 TAP – Test Analysis Program (Gordon P. Brooks, Ohio University)

TAP, Ohio Üniversitesi’nden Gordon P. Brooks tarafından geliştirilen, Delphi ile Windows ortamında çalışan ücretsiz bir test analiz paketidir. Yazılım, klasik test teorisine dayalı analizleri (madde ayırt edicilik, güçlük endeksi, Spearman–Brown gibi güvenilirlik

hesaplamaları) kolaylaştırmakta ve kullanıcı dostu bir arayüz sunmaktadır Program, eğitimde test tasarımı ve değerlendirme süreçlerinde öğretim elemanlarını destekleyecek niteliktedir (Brooks 2003).

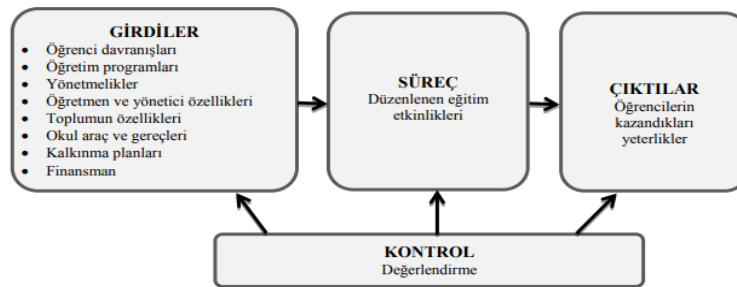
1.1.3 IteMan (Assessment Systems Corporation)

IteMan, Assessment Systems Corporation tarafından geliştirilen bir ticari psikometrik yazılımdır. Klasik test teorisine dayalı ölçümler üretir; mesela maddelerin P ve r-pbis değerlerini hesaplayarak MS Word biçiminde ayrıntılı grafik ve raporlar oluşturur. Yazılım güçlü analiz ve otomasyon özelliklerine sahip olsa da, ücretli olması ve yalnızca İngilizce arayüz sunması, özellikle Türk kullanıcılar için kullanım engeli teşkil etmektedir (ASC 2000).

2. LİTERATÜR TARAMA

Eğitim, bireylerin davranışlarını toplumsal ihtiyaçlarla uyumlu bir şekilde geliştirmeyi hedefleyen sistematik bir süreçtir. Toplumların nitelikli insan kaynağı oluşturma kapasitesinde eğitimin rolü merkezi bir öneme sahiptir. Ülkelerin kalkınma düzeyleri, eğitim almış bireylerin elde ettiği başarılarla doğrudan ilişkilendirilmekte ve bu durum, eğitim politikalarının stratejik ve uzun vadeli bir yaklaşımla uygulanmasının gerekliliğini ortaya koymaktadır. Ayrıca, bireylerin potansiyellerini tam anlamıyla gerçekleştirebilmeleri için fırsat eşitliğinin sağlanması temel bir ön koşul olarak değerlendirilmektedir. Eğitim süreci, bireyin kendini tanıması, ilgi alanları ve tercihleri doğrultusunda gelişim göstermesi için bir zemin sunmakta; bu sayede birey, toplumsal, kültürel ve ekonomik alanlarda anlamlı katkılar sağlayabilecek bir düzeye ulaşabilmektedir (Turan 2018).

Eğitim, bireylerde davranış değişikliği oluşturma süreci olarak, girdiler, süreç, çıktılar ve kontrol (değerlendirme) bileşenlerinden oluşan bir sistem olarak ele alınmaktadır. Bu sistemde, öğrencilerin eğitime başlamadan önce sahip oldukları davranışlar; öğretim programları, yasal düzenlemeler, öğretmen ve yönetici nitelikleri, toplumun özellikleri, okul ortamı, araç-gereçler, finansman kaynakları ve kalkınma planları gibi unsurlar, sistemin girdilerini oluşturmaktadır. Öğrencilerin belirlenen öğrenme hedeflerine ulaşabilmeleri için planlanan eğitim etkinlikleri, sistemin süreç bileşenini; bu süreç sonucunda öğrencilerin kazandıkları yeterlikler ise sistemin çıktıları temsil etmektedir. Sistemin işleyişine yönelik bir değerlendirme yapabilmek amacıyla gerçekleştirilen çalışmalar, sistemin kontrol (değerlendirme) bileşenini oluşturarak, diğer bileşenlere geribildirim sağlar ve bu sayede sürecin sürekli geliştirilmesine olanak tanır (Karadağ, 2014). Bu bileşenler Şekil 2.1’de belirtilmiştir.



Şekil 2.1 Eğitim sistemi bileşenleri arasındaki ilişki (diledebiyat.net, 2025).

Eğitim, yüzyıllardır farklı paradigmlar doğrultusunda şekillenmiş, her bir yaklaşımın katkısıyla yeni boyutlar kazanarak dinamik bir yapı haline gelmiştir (Baker vd. 2021). İdealistlere göre eğitim, bireyin Tanrı'ya ulaşma sürecinde gerçekleştirdiği etkinlikler olarak tanımlanırken, Realistler bu süreci bireyin toplumsal değerler doğrultusunda yetiştirilmesi olarak değerlendirmektedir (Erikawati 2023). Marksistler açısından eğitim, toplumsal çelişkilerin azaltılması ve üretime katkı sağlanması amacı taşırken; Pragmatistler, bireyde yaşantılar yoluyla istenen davranış değişikliklerinin oluşturulmasını hedefler. Varoluşçular ise eğitimi, bireyin varoluşsal sınırlarına ulaşmasını sağlayan bir süreç olarak ele almaktadır (Biesta 2014). Bu farklı tanım ve yaklaşımlara rağmen eğitimin ortak noktası, bireyin davranışlarında istenen değişimleri meydana getirmek ve onun potansiyelini gerçekleştirmesine olanak tanımadır. Bilgi, beceri ve tutumlar kazandıran eğitim, bireyin hem topluma anlamlı katkılar sunmasını hem de kendini gerçekleştirmesini sağlayan temel bir araçtır (Günel 2024).

Eğitim, insanlık tarihinin en köklü uğraş alanlarından biri olup, sosyal, politik, kültürel ve bireysel unsurları bünyesinde barındıran karmaşık bir yapıya sahiptir; bu nedenle, tanımlanması güç bir kavramdır. Genel anlamda eğitim, bireyin mutluluk ve refahını artırmayı amaçlayan, çok boyutlu ve kalıcı değişimlerin yaşandığı bir süreç olarak ifade edilebilir. Eğitim hedeflerinin gerçekleştirilmesi, bireylerin bu süreçte edindikleri temel beceriler ve değerlerle doğrudan ilişkilidir. Bu çerçevede, temel beceri ve değerlerin kazandırıldığı formel eğitimin ilk basamağı olan temel eğitim, süresi ülkeler arasında farklılık göstermekle birlikte, evrensel olarak zorunlu kabul edilmekte ve bir ülkenin insan kaynağının niteliğini belirleyen kritik bir eğitim aşaması olarak değerlendirilmektedir (Hanushek 2013). Toplumların refah düzeyi, büyük ölçüde eğitilmiş insan gücünün niteliğine bağlı olduğundan, temel eğitimin kalitesi bu bağlamda stratejik bir öneme sahiptir (Arıkan vd. 2017).

Nitelikli eğitim, öncelikle eğitim politikalarının ve uygulamalarının çağdaş, akılcı ve bilimsel temellere dayalı bir şekilde hayata geçirilmesine bağlıdır. Eğitim bilimlerinde gerçekleştirilen çalışmalar, kuramsal ve uygulamalı boyutlarda, bilimsel araştırmaların bulguları doğrultusunda şekillenmektedir. Eğitim-öğretim süreçlerinde temel hedef, eğitimin niteliğini artırmaktır ve bu hedefin gerçekleşmesinde öğretim, öğrenme ve

ölçme-değerlendirme boyutları belirleyici bir role sahiptir. Eğitim bilimlerinin farklı disiplinlerinde yapılan araştırmalar, özellikle öğrenme-öğretim ve ölçme-değerlendirme alanlarında önemli değişimlere yol açmıştır. Bu değişimlerin temelinde, son otuz ila otuz beş yıl içerisinde bilişsel öğrenme kuramları bağlamında yürütülen bilimsel çalışmaların etkisi büyüktür (Yılmaz 2011). Bilişsel öğrenme kuramlarının sunduğu bulgular, hem eğitim süreçlerinin niteliğini artırmış hem de yeni yaklaşım ve uygulamaların gelişmesine olanak sağlamıştır (Arıkan vd. 2017).

Dağılım ölçülerinin kavranması, öğrencilerin istatistiksel okuryazarlık düzeylerinin gelişmesine önemli katkı sağlamaktadır. Yapılan araştırmalar, öğrencilerin özellikle varyans ve standart sapma gibi temel istatistiksel kavramları anlamakta zorlandıklarını ve bu nedenle istatistiği genellikle zorlayıcı bir ders olarak algıladıklarını ortaya koymaktadır (Leng & Meng, 2023). Bu kavramsal yetersizlik, veri setlerindeki değişkenliği doğru biçimde yorumlamayı güçleştirmekte; bu da sınav sonuçlarının analizinden eğitim programlarının etkililiğinin değerlendirilmesine kadar pek çok süreçte sorunlara yol açabilmektedir. Öte yandan, dağılım ölçülerine yönelik geliştirilen ve öğrenme güçlüklerini hedef alan öğretimsel müdahalelerin, öğrencilerin hem istatistiksel başarılarını hem de bu kavramlara ilişkin bilişsel anlayışlarını artırmada etkili olduğu belirlenmiştir (delMas & Liu, 2005; Leng & Meng, 2023).

2.1 Ölçme ve Değerlendirme

2.1.1 Ölçme

Günümüz uygar toplumunda ölçme, hemen her alanda ve işlemde vazgeçilmez bir unsur olarak karşımıza çıkmaktadır. Örneğin, zamanı belirlemek, bir aracın hızını ölçmek ya da harcanan elektrik ve su miktarını tespit etmek için çeşitli ölçme araçları geliştirilmiştir. Çağdaş yaşamın düzenli ve sürdürülebilir bir şekilde ilerleyebilmesi, büyük ölçüde, farklı özellikleri ölçmeye olanak tanıyan bu araçların varlığına dayanmaktadır. Ölçme araçlarının bir an için yok olduğunu düşündüğümüzde, toplumsal düzenin bozulması kaçınılmaz hale gelir ve bireyler, modern yaşamın temel ihtiyaçlarını karşılayamaz duruma düşerek uygarlık düzeyinde ciddi bir gerileme yaşar (Tekin 1979).

Ölçme, bir betimleme süreci olarak tanımlanır ve geniş anlamda, bir nesnenin ya da nesne grubunun belirli bir özelliğe sahip olup olmadığını, eğer sahipse bu özelliğin derecesini gözlemleyerek tespit etme ve bu gözlem sonuçlarını sembollerle, özellikle sayısal ifadelerle temsil etme işlemidir (Briggs ve Maul 2022). Örneğin, bir bireyin cinsiyetinin (kadın-erkek), bir erkeğin medeni durumunun (evli-bekâr-dul-boşanmış) ne olduğunu belirlemek; belirli bir gündeki sıcaklığı derece cinsinden ifade etmek veya bir çocuğun ağırlığını kilogram cinsinden göstermek, ölçme süreçlerine somut örnekler olarak sunulabilir (Tekin 1979).

Ölçme işlemi, temel olarak bir özelliği belirleme ve değerlendirme sürecidir. Uzunluk, sıcaklık, derinlik, nem, yağış miktarı, nabız, depremin şiddeti, zeka ve başarı gibi kavramlar, varlıkların veya olayların sahip olduğu nitelikleri ifade eder. Ölçme, bu nitelikleri belirlemek ve "Ne kadar?" sorusuna yanıt aramak amacıyla gerçekleştirilir (Wilson 2024). Bu süreç, insanın doğasında bulunan "merak" duygusundan kaynaklanır. Bununla birlikte, ölçme işlemi, hayatı daha güvenli ve sağlıklı bir şekilde sürdürebilmek için de kritik bir öneme sahiptir. Varlıkların veya olayların çeşitli özelliklerini ve bu özelliklere sahip olma derecelerini tespit etme süreci genel anlamda "gözlem" olarak adlandırılabilir. Ancak, bir gözlemin "ölçme" olarak nitelendirilebilmesi için, elde edilen sonuçların mutlaka "sayı" veya "semboller" ile ifade edilmesi gerekmektedir (Yılmaz 1996).

Ölçme türleri, genel olarak doğrudan, dolaylı ve türetilmiş ölçme olmak üzere üç grupta sınıflandırılmaktadır. Doğrudan ölçme, ölçülmek istenen özelliğin, aynı nitelikte bir ölçüm aracı ile herhangi bir aracı değişken kullanılmaksızın belirlenmesini ifade eder. Örneğin, bir cismin uzunluğunun metreyle ya da ağırlığının teraziyle ölçülmesi doğrudan ölçmeye örnek teşkil eder (Luce ve Kirnan 2016). Buna karşılık, dolaylı ölçme, ölçülen özellik ile ölçüm aracının niteliğinin farklı olduğu, ölçümün geçerli varsayımlar temelinde dolaylı yollarla yapıldığı durumları kapsar. Havanın sıcaklığının cıvanın genleşme miktarıyla veya öğrencinin başarısının bir test aracılığıyla değerlendirilmesi bu tür ölçmelere örnektir. Dolaylı ölçme, özelliğin doğrudan gözlemlenememesi nedeniyle başka bir özellik ile ilişkilendirilmesini içerdiğinden, kullanılan varsayım ve araçların

geçerliliği ölçümün doğruluğu açısından kritik önem taşır (Yılmaz 1996, Uysal 2019). Üçüncü tür olan türetilmiş ölçme ise, bir değişkenin doğrudan veya dolaylı olarak ölçülemmediği, ancak diğer değişkenlerle olan aritmetik ilişkiler yoluyla tanımlandığı ölçmelerdir. Örneğin, hız, yoğunluk veya nüfus yoğunluğu gibi değişkenler; mesafe, zaman ya da nüfus ve alan gibi başka değişkenler arasındaki matematiksel ilişkiler aracılığıyla ölçülür. Bu nedenle, türetilmiş ölçme tanımlayıcı değişkenlerin ve aritmetik işlemlerin bütünlüğüyle gerçekleşir (Atılğan vd. 2014).

Ölçmede ölçek ise ölçme sonuçlarını ifade eden sayı veya sembollerin formal, yani matematiksel ve istatistiksel özelliklerini temsil eden değerdir (Turgut ve Baykul,2010).

2.1.2 Değerlendirme

Değerlendirme, karar verme sürecine dayalı bir işlem olup, bu yönüyle ölçme işleminden ayrılmaktadır. Ölçme, mevcut durumun, imkanlar dahilinde gerçeğe en uygun şekilde betimlenmesini hedeflerken, değerlendirme süreci, ölçme sonuçlarının belirlenmiş bir ölçütle karşılaştırılmasını içerir (Chikwe 2021). Bu süreçte, ölçme sonuçlarının ilgili ölçütle tanımlanan koşulları karşılayıp karşılamadığı analiz edilir. Nihai olarak, değerlendirme, ölçülen özelliğe dair bir karara ulaşmayı sağlar. Bu karar, ölçülen özelliğin belirli bir amaç doğrultusunda uygunluğunu veya işlevselliğini ortaya koyar (Özçelik 1998).

Değerlendirmenin temel unsurlarından biri olan ölçme, değerlendirme sürecinde ihtiyaç duyulan gözlem verilerinin sistematik bir şekilde elde edilmesi amacıyla gerçekleştirilir (Gömleksiz vd. 2008).

Ölçüt, ölçme sonuçlarının karşılaştırılmasında esas alınan ve karar verme süreci boyunca değişmeyen belirli standart veya referans değerlerdir (Gömleksiz vd. 2008). Ölçme sonucunun belirli bir ölçüt veya ölçütler takımıyla karşılaştırılmasının ardından, oluşan değer yargısına dayalı bir karara varılması gereklidir. Bu karar, değerlendirme sürecinin temel sonucunu ifade eder. Bu kararın bir isabetlilik derecesi mevcuttur ve bu derece 3 şekilde gösterilmektedir (Gömleksiz vd. 2008).

- Ölçme sonuçlarının hatasız, geçerli ve güvenilir olması,
- Ölçütlerin değerlendirme amacına uygun şekilde belirlenmesi,
- Kıyaslama işlemlerinin doğru bir şekilde gerçekleştirilmesi.

Kararın doğruluğunu ve sürecin güvenilirliğini sağlamak, değerlendirme sürecinde büyük önem taşır (Singh 2024). Kararın isabetsiz olması, hem sistem hem de bireyler hakkında yanlış yargılara varılmasına neden olarak süreci olumsuz etkileyebilir. Örneğin, bir programın hedef davranışları kazandırıp kazandırmadığını belirlemek amacıyla yürütülen bir değerlendirmede, değerlendirme alt sisteminde hata olmadığı varsayılırsa, aslında düzeltilmesi gereken bir program olmasına rağmen öğretimin yetersiz olduğu sonucuna ulaşılabilir ve sistem gereksiz şekilde eleyici bir niteliğe bürünebilir. Benzer şekilde, öğrencilerin seviyelerinin altında bir eğitim programında yüksek başarı göstermesi, sistem kaynaklarının gereksiz yere harcanmasına ve emeğin, zamanın ve maliyetin boşa gitmesine yol açabilir. Dolayısıyla, değerlendirme süreçlerinde dikkatli ve doğru kararlar almak, hem sistemin etkinliğini artırmak hem de bireylerin gerçek potansiyellerini ortaya koymak açısından bir öneme sahiptir (Baykul vd. 1993).

2.1.3 Ölçme Araçlarının Sahip Olması Gereken Temel Nitelikler

Bir ölçme aracında bulunması gereken en temel üç özellik, geçerlik, güvenirlik ve kullanılabilirlik (Prydz vd. 2023). Bu özellikler, ölçme aracının belirlenen amaca uygun ölçümler yapabilmesini, ölçümlerin tutarlılığını sağlamasını ve aracın kolaylıkla uygulanabilir olmasını ifade eder. Ancak, bu özelliklerin detaylı bir şekilde incelenmesine geçmeden önce, ölçme sürecinde önemli bir yer tutan hata kavramı ve hata türleri üzerinde durulması gerekmektedir (Arslan 2011).

Fiziksel bilimler ve eğitim bilimlerinde gerçekleştirilen her ölçme sonucunda kaçınılmaz olarak bir miktar hata bulunur. Ölçülen özelliğin gerçek değeri hakkında doğru ve güvenilir sonuçlara ulaşabilmek için, ölçme sürecine karışan hataların en aza indirilmesi gereklidir. Bu bağlamda, hata türlerinin ve kaynaklarının ölçme sonuçları üzerindeki etkilerinin ayrıntılı bir şekilde anlaşılması son derece önemlidir (Çepni vd. 2007). Ölçme hataları, sabit, sistematik (sistemli) ve rastgele (tesadüfi) olmak üzere üç temel kategoriye

ayrılmaktadır (Çepni vd. 2007). Bir ölçmeden diğerine tutarlı bir şekilde tekrarlanan, ölçme sonuçlarına sürekli olarak etki eden ve miktarı değişmeyen hatalar, sabit hata olarak tanımlanır (Gömleksiz vd. 2008).

Sistematik (Sistemli) hatalar, her zaman aynı oranda görülmez. Bir ölçmeden diğerine tekrarlanan, ancak miktarı ölçülen değişkenin büyüklüğüne veya boyutuna bağlı olarak farklılık gösteren hatalar şeklinde tanımlanır (Şimşek 2013).

Tesadüfi hatalar, ölçümü gerçekleştiren bireyin dikkatsizliği, ölçme ortamındaki koşullar, ölçme aracı ve çeşitli diğer faktörlerden kaynaklanan bir hata türüdür. Bu tür hataların kaynakları ve ölçme üzerindeki etkilerinin miktarları kesin olarak belirlenemez. Tesadüfi hatalar, ölçüm sonuçlarına yalnızca tek yönlü değil, bazen pozitif bazen de negatif yönde etkilerde bulunur (Wells ve Sireci 2020). Bu nedenle, tesadüfi hatalar kontrol edilemez ve sabit ya da sistematik hatalarda olduğu gibi cebirsel yöntemlerle düzeltilemez (Baykul 2000).

Klasik tanıma göre geçerlik, bir testin yalnızca ölçmek istediği değişkeni doğru bir şekilde ölçmesi ve diğer değişkenlerle karışıklık yaratmaması olarak ifade edilir (Sireci 1998). Cronbach, geçerliği dar anlamda, "bir test puanından elde edilen yordama veya kestirmenin doğruluğu" olarak tanımlamıştır. Aynı bağlamda, geçerlik, "test puanlarının, ölçüt olarak alınan diğer puanlarla karşılaştırılması" şeklinde de açıklanmaktadır (Cronbach ve Meehl 1955). Cronbach, geçerliği karar teorisi ile ilişkilendirerek, kullanıcının test puanlarından çıkarılabilecek yordamalara yönelik topladığı kanıtlar (evidence) olarak ele almıştır (Baykul 2000).

Bir testin geçerliğini belirlemek için kullanılabilecek basit ve tek bir yöntem bulunmamaktadır. Geçerlik, tek bir ifade veya açıklama ile kesin olarak değerlendirilemez (Cook ve Hatala 2016). Bir test, belirli bir amaç için geçerli olabilirken, başka bir amaç için aynı geçerliği sağlayamayabilir. Benzer şekilde, bir öğrenci grubu üzerinde ölçmek istenilen özelliği doğru bir şekilde ölçebilirken, farklı bir öğrenci grubu üzerinde aynı etkinliği göstermeyebilir. Bu nedenle, bir testin geçerliğinin değerlendirilmesi, testin kullanıldığı bağlam ve amaca uygun olarak çok yönlü bir analiz ve değerlendirme sürecini gerektirir (Micheels & Karnes 1950).

Çoktan seçmeli testler (ÇST'ler), nesnel puanlama imkânı ve uygulama açısından sundukları verimlilik nedeniyle farklı eğitim düzeyleri ve disiplinlerde yaygın olarak kullanılan değerlendirme araçları arasında yer almaktadır (McCoubrie 2004, Scott vd. 2006, Sayyah vd. 2011). Bu testlerin eğitimsel değerlendirmelerde tercih edilmesinde en önemli etkenlerden biri, yüksek güvenirlik düzeylerine sahip olmalarıdır. Güvenirlik, bir testin farklı zamanlarda ve çeşitli koşullarda benzer sonuçlar üretme kapasitesini ve elde edilen puanların tutarlılığını ifade eder (Ali vd. 2016). Yüksek güvenirliğe sahip testler, öğrencilerin bilgi ve beceri düzeylerini daha doğru şekilde yansıtarak eğitimcilerin bu sonuçlara dayalı bilinçli kararlar almalarına olanak tanır (Norcini vd. 1985). Güvenilir bir ölçme aracı, aynı birey grubuna tekrar uygulandığında tutarlı sonuçlar üretir ve bireylerin her uygulamada benzer puanlar alması beklenir (Tekin 1977).

Güvenirlik, geçerliğin bir alt unsuru olarak değerlendirilebilir. Bir testin, ölçmek istediği özelliği doğru bir şekilde ölçebilmesi için öncelikle bu özelliği tutarlı bir şekilde ölçmesi gereklidir. Bununla birlikte, güvenirlik, geçerliğin tek başına bir garantisi değildir. Çok güvenilir bir test, belirli bir amaç için düşük bir geçerliliğe sahip olabilir (Kimberlin ve Winterstein 2008). Öğretmen tarafından uygulanan sınavların, geçerlik ve güvenirliklerinin yanı sıra, kullanışlılık özelliklerinin de yüksek olması gereklidir. Kullanışlılık, sınavların hazırlanması, uygulanması ve değerlendirilmesi süreçlerinde ekonomik ve pratik olmasını ifade eder (Brown 2022). Bu bağlamda, sınav aracının hem ekonomik olması hem de öğretmene ve öğrenciye ek bir yük getirmemesi önemlidir. Öğretmenlerin, maliyeti düşük ve kolay uygulanabilir sınavlar hazırlaması bu açıdan gereklidir. Ayrıca, sınav türü, öğretmenin zaman olanaklarına uygun olarak seçilmelidir. Örneğin, test tipi sınavların hazırlanması daha fazla zaman alabilirken, puanlaması hızlıdır; yazılı yoklamalar ise kısa sürede hazırlanabilir, ancak puanlama süreci daha uzun sürer. Dolayısıyla, öğretmenler mevcut koşullarını ve kaynaklarını dikkate alarak, ihtiyaçlara en uygun sınav türünü belirlemelidir. Bu yaklaşım, sınavların hem kullanışlılığını hem de genel etkinliğini artıracaktır (İşman 1998).

Bir ölçme aracının kullanışlılığı, kullanım sürecinde bir dizi kolaylık sağlaması ile ilişkilidir. Ölçme aracının kullanımına yönelik olarak taşınması gereken temel özellikler şu şekilde sıralanabilir (Navarro-Cota vd. 2023, Gümüş 1975):

- Uygulanabilirlik: Ölçme aracı, uygulaması kolay bir yapıya sahip olmalıdır.
- Puanlama Kolaylığı: Puanlama işlemleri pratik bir şekilde gerçekleştirilebilmeli ve zaman açısından verimli olmalıdır.
- Anlaşılabilirlik: Ölçme aracı, açık ve net bir biçimde hazırlanmış olmalı, kullanıcıyı yanlış anlamaya yönlendirebilecek unsurlar içermemelidir.
- Tasarım ve Düzen: Soruların düzeni ve yerleşimi kullanıcı dostu bir tasarıma sahip olmalıdır.
- Ekonomiklik: Araç, maliyet açısından ekonomik olmalı ve uygulamada fazla masraf gerektirmemelidir.
- Zaman Yönetimi: Uygulama için ayrılan süre, amaca uygun bir şekilde belirlenmeli ve eksik ya da aşırı süre kullanımına neden olmamalıdır.
- Tutarlılık: Eğer ayrı bir cevap kağıdı kullanılacaksa, bu kağıt ile test defteri arasında tam bir uyum olmalı ve herhangi bir karışıklığa yol açmamalıdır.

Ölçme araçlarının sahip olması gereken nitelikler kapsamında kullanılabilirlik önemli bir unsur olarak öne çıkmakta; bu özellik, uygulama sürecinde hem değerlendiricilere hem de katılımcılara pratiklik ve anlaşılabilirlik sağlamaktadır. Ancak bir ölçme aracının sadece kullanışlı olması, değerlendirme sürecinde tek başına yeterli değildir. Ölçme işleminin geçerli ve güvenilir sonuçlar verebilmesi, aynı zamanda yapısal olarak güçlü testlerin ve bu testleri oluşturan maddelerin nitelikli bir biçimde tasarlanmasına bağlıdır. Bu nedenle, ölçme araçlarının değerlendirilmesinde test ve madde kavramlarının yapısal bileşenlerini ve bu unsurların analiz süreçlerini ele almak gerekmektedir (Ashraf Jaseem 2020).

2.2 Test ve Madde Analizi

2.2.1 Temel Kavramlar

Test kavramı, çoğunlukla belirli ve özel olarak tasarlanmış koşullar altında yerine getirilecek görevler ya da yanıtlanması gereken sorular bütünü şeklinde tanımlanmalıdır (Tekin 1977). Bir testin en küçük ölçülebilir ve puanlanabilir birimi, "madde" olarak adlandırılmaktadır (Hasançebi vd. 2019).

2.2.2 Test Analizi

Test uygulamasından elde edilen sonuçlar puanlandıktan sonra, eğitim-öğretim süreçlerinin etkinliğini desteklemek amacıyla, bu sonuçların kapsamlı bir şekilde analiz edilmesi ve değerlendirilmesi önem arz etmektedir (Yılmaz 1998). Test sonuçlarının düzenlenmesi, puanların azalan veya artan bir sıra ile sıralanması yoluyla sistematik bir biçimde gerçekleştirilir (Gömleksiz vd. 2008).

Bir grup öğrencinin sınav sonuçlarının analizi yapıldığında, genellikle öğrencinin öncelikle kendi puanına odaklanacağı görülür. Ardından, yakın arkadaşlarının notlarını incelemesi olasıdır. Bunun yanı sıra, en yüksek ve en düşük puanları hangi öğrencilerin aldığı gibi bilgiler de merak edilebilir. Sınav sonuçlarının detaylı bir şekilde incelenmesi, yüksek başarı gösterenler ile düşük performans sergileyenlerin belirlenmesini sağlar. Bu analiz, sınıfın genel akademik durumu, puanların dağılım özellikleri ve sınavın zorluk düzeyi hakkında değerli bilgiler sunar (Başol 2016). Bu açıklamaya örnek olarak Çizelge 2.1'de sınıf listesine göre 20 öğrencinin 40 maddelik bir testten aldığı puanlar gösterilmiştir. Çizelge 2.1 incelendiğinde en yüksek puanın 92 en düşük puanın 54 ve en çok tekrarlanan puanın 6 kez tekrar eden 65 olduğu tespit edilmektedir.

Merkezi eğilim ölçüleri, eğitimde özellikle öğrenci başarılarını değerlendirme ve sınav sonuçlarını analiz etme amacıyla yaygın biçimde kullanılmaktadır. Örneğin, bir sınıfın ortalama puanı, öğrencilerin genel konuya ne ölçüde hâkim olduğunu gösterebilirken; uç değerlerin bulunduğu veri setlerinde, medyan değeri öğrencilerin ortalama başarı düzeyini daha sağlıklı yansıtabilir (Honnor vd. 2022). Bu tür istatistiksel göstergeler, eğitimsel değerlendirmelerde önemli bir yer tutmakta; öğretmenlerin öğrenci performansındaki genel eğilimleri tanımlamalarına ve eksikliklerin bulunduğu alanları tespit ederek öğretim sürecini iyileştirmelerine katkı sağlamaktadır (Wilson 2024). Öte yandan, yapılan araştırmalar, öğrencilerin istatistik derslerine yönelik tutumlarının, merkezi eğilim kavramlarını anlama ve etkin bir şekilde kullanma becerilerini etkilediğini ortaya koymaktadır. Bu durum, eğitim ortamlarında olumlu istatistiksel tutumların geliştirilmesinin ne denli önemli olduğunu göstermektedir (Saidi ve Siew 2019, Peiró-Signes vd. 2020).

Çizelge 2.1 20 öğrencinin 40 maddelik bir sınavdan aldığı notlar.

Öğrenci	Puan	Öğrenci	Puan
Ahmet	86	Damla	65
Selim	72	Osman	89
Esra	75	Leyla	91
Müzeyyen	65	Seçil	65
Kadir	92	Nurgül	60
Uğur	90	Pelin	63
Ekrem	65	Tuğrul	57
Seda	74	İrfan	65
Ata	66	Kaan	65
Selma	61	Kürşat	58

Merkezi eğilim ölçütleri arasında en sık tercih edilen ölçüt, aritmetik ortalamadır. Bunun başlıca nedeni, aritmetik ortalamanın tüm verileri dikkate alarak hesaplanması ve ileri düzey istatistiksel analizler için en uygun ve kullanışlı ölçüt olarak kabul edilmesidir (Saha 2011). Aritmetik ortalama, genellikle "ortalama" terimiyle kısaca ifade edilmektedir (Baykul 1997).

Tepe değer, bir veri setinde en sık tekrar eden değeri ifade eder. Örneğin, Çizelge 2.1'deki verilere göre, en sık tekrar eden değer 65 olarak belirlenmiştir. Tepe değer, özellikle uç değerlerin bulunduğu durumlarda, merkezi eğilimi aritmetik ortalamaya kıyasla daha güvenilir bir şekilde temsil eden bir istatistiksel ölçüttür (Bickel ve Fruehwirth 2005). Bir dağılımda birden fazla tepe değer bulunabilir. İki tepe değer mevcut ise dağılım iki modlu, üç tepe değer mevcut ise üç modlu olarak tanımlanır (Başol 2016).

Bir veri grubunun aritmetik ortalaması, tüm veri değerlerinin toplamının, veri sayısına bölünmesiyle elde edilir. Bu durum, formül 2.1'de verilen ve $x_1 + x_2 + x_3 + \dots + x_n$ şeklinde ifade edilen veri değerleri ve n ile gösterilen toplam veri sayısı dikkate alınarak, matematiksel olarak gösterilebilir (Baykul 1997):

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (2.1)$$

Burada, $\bar{x}$ aritmetik ortalamayı, x_1 her bir veri değerini ve n toplam veri sayısını temsil etmektedir. Bunun dışında $\sum$ sembolü eklenen versiyonu ise formül 2.2 de gösterilmektedir.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.2)$$

Sıralanmış bir veri setinde tam ortada bulunan puan, “medyan” olarak adlandırılır. Diğer bir ifadeyle, medyan, sıralanmış veri setini %50’lik iki eşit gruba ayıran değeri temsil eder. Sıralanmış verileri iki eşit yarıya bölen bu ölçüt, dağılımın ağırlık merkezini moddan daha doğru bir şekilde yansıtmaktadır (Güler 2011).

Medyanın hesaplanma yöntemi, veri setindeki puan sayısının tek veya çift olmasına bağlı olarak formül 2.3 ve formül 2.4’deki gibi farklılık göstermektedir (Güler 2011). Puan dağılımı tek sayı ise, n toplam veri (ölçüm sonucu) sayısını ifade etmek üzere medyan (ortanca) aşağıdaki formül ile hesaplanır:

$$n \text{ değeri tek sayı ise : } \text{Medyan} = x_{\left(\frac{n+1}{2}\right)} \quad (2.3)$$

n : Veri setindeki toplam gözlem (veri) sayısıdır.

x : Küçükten büyüğe sıralanmış veri setinde belirli bir sıradaki değeri temsil eder.

$$n \text{ değeri çift sayı ise : } \text{Medyan} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} \quad (2.4)$$

n : Veri setindeki toplam gözlem sayısıdır (çift).

x : Bu sıradaki verilerin değerlerini gösterir.

Bir veri kümesinde, en büyük ölçüm sonucu ile en küçük ölçüm sonucu arasındaki fark, ranj olarak adlandırılır (Baykul ve Güzeller 2013).Ranj değerinin daha küçük olduğu veri setlerinde, değişkenliğin (yaygınlığın) daha düşük olduğu kabul edilmektedir. Bir veri setinde değişkenliğin olmadığı durumlarda (tüm değerlerin birbirine eşit olması halinde), ranj sıfır olarak ifade edilir (Alpar 2016).

Varyans, istatistiksel işlemlerde puanların dağılımını değerlendirmek için kullanılan bir istatistiksel ölçüttür. Formül 2.5’de belirtilen varyans, puanların ortalamadan olan

sapmalarının kareleri toplamının, evren için N'e, örneklem için ise (n-1)'e bölünmesiyle hesaplanır (Rennie 1997).

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (2.5)$$

s^2 : Varyans

$\bar{x}$: Aritmetik Ortalama

$X_1, X_2, X_3, \dots, X_n$: Puanlar

n: Ölçme Sayısı

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (2.6)$$

s : Standart sapma

$\bar{X}$: Aritmetik Ortalama

$X_1, X_2, X_3, \dots, X_n$: Puanlar

n: Ölçme Sayısı

Varyans, verilerin ortalamadan uzaklıklarının karelerinin ortalamasını ifade etmesi ve biriminin karesel olması nedeniyle, uygulamada genellikle karekökü alınarak kullanılır. Bu karekök, standart sapma olarak adlandırılır (Rakrak 2025). Standart sapma, verilerin yaygınlığını ölçmek için en yaygın kullanılan istatistiksel ölçütlerden biridir ve bir veri kümesindeki tüm değerlerin aritmetik ortalamaya olan uzaklıklarının ortalama bir temsilcisi olarak tanımlanır formül 2.6' da matematiksel ifadesi görülmektedir (Alpar 2016).

Standart sapma ve varyans, ortalamaya bağlı olarak hesaplandığından belirli sınırlılıklar içerir ve bu durum, özellikle farklı ortalamalara sahip grupların karşılaştırılmasında yanlışlık riskini artırır (Smith 2018). Varyans veya standart sapmalardaki küçük farklılıklar, bu yanlışlık nedeniyle yanlış yorumlamalara yol açabilir. Bu tür hataları önlemenin en etkili yöntemi, standart sapma ve aritmetik ortalamayı birlikte ele alan bağlı

değişim katsayısını kullanmaktır (Smith 2018). Bağlı değişim katsayısı, standart sapmanın aritmetik ortalamaya oranı olarak tanımlanır ve farklı ortalamalara sahip grupların daha tutarlı ve doğru bir şekilde karşılaştırılmasını sağlar (Doğan 2008). Bahsi geçen bağlı katsayısının matematiksel ifadesi formül 2.7 de verilmiştir.

$$V = \frac{S_x}{\bar{x}} \cdot 100 \quad (2.7)$$

V : Bağlı değişim katsayısı

S_x : Standart Kayma

$\bar{X}$: Ortalama

Bağlı değişim katsayısı, bir gruptaki değişkenliğin niteliğini analiz etmek ve yorumlamak amacıyla aşağıdaki ilkeler doğrultusunda değerlendirilir (Santos ve Dias 2021):

- Bağlı değişim katsayısı 25'ten büyükse ($V > 25$), dağılımın basık, heterojen ve çeşitlilik gösteren bir yapıya sahip olduğu kabul edilir.
- Bağlı değişim katsayısı 20 ile 25 arasında ise ($20 \leq V \leq 25$), dağılımın normal, simetrik ve dengeli olduğu düşünülür.
- Bağlı değişim katsayısı 20'den küçükse ($V < 20$), dağılımın sivri, homojen ve benzeşik bir yapıya sahip olduğu ifade edilir.

Bu ilkeler, gruplar arasındaki değişkenliğin yapısını daha doğru ve tutarlı bir şekilde yorumlamak için kullanılmaktadır (Doğan 2008).

Çarpıklık katsayısı: Formül 2.8'de gösterilen çarpıklık katsayısı bir dağılımın simetrik olup olmadığına ilişkin bilgi sağlar (Kasap 2012).

$$\zeta k = \frac{3 \cdot (\bar{x} - median)}{S_x} \quad (2.8)$$

ζk : Çarpıklık katsayısı

s_x : Standart kayma

$\bar{x}$: Ortalama

$\zeta k = 0$ Simetrik eğri

$\zeta k < 0$ Sola çarpık eğri negatif kayışlı

$\zeta k > 0$ Sağa çarpık eğri pozitif kayışlı

Çarpıklık katsayısı, pozitif, negatif veya sıfır değer alabilir. Katsayının pozitif bir değer alması, dağılımın sağa çarpık (pozitif çarpıklık) olduğunu ve ortalamadan ortancadan büyük olduğunu ifade eder. Negatif bir değer, dağılımın sola çarpık (negatif çarpıklık) olduğunu ve ortancanın ortalamadan büyük olduğunu gösterir (Moral de la Rubia 2023). Çarpıklık katsayısının sıfır olması ise dağılımın ortalamaya göre simetrik bir yapıya sahip olduğunu belirtir (Gömleksiz vd. 2008).

$$K20 = \frac{K}{K-1} \left(1 - \frac{\sum p \cdot q}{S_x^2} \right) \quad (2.9)$$

K: Testte bulunan madde sayısı

p : Madde güçlük indeksi

q : 1 – p

S_x^2 :Test puanları dağılım varyansı

KR20: Formül 2.9’da matematiksel ifadesi gösterilen Kuder-Richardson 20 (K-R20) Bir testin iç tutarlılığını değerlendirmek için kullanılan bir yöntemdir. 1-0 puanlama sistemine uygun yapısı nedeniyle, özellikle çoktan seçmeli ve doğru-yanlış formatındaki testlerde yaygın olarak tercih edilmektedir (Ntumi vd. 2023). Güvenirlik katsayısı, testin tek bir uygulamasıyla elde edilen puanların ilgili formüle yerleştirilmesiyle hesaplandığından, bu yöntem tek uygulamalı bir analiz tekniği olarak kabul edilir (Kasap 2012).

KR21: KR20 formülü, doğru cevapların 1 puan, yanlış veya boş bırakılan cevapların ise 0 puanla değerlendirildiği testlerde veya soruların doğru cevabının "evet" ya da "hayır" gibi iki seçenektan biri olduğu durumlarda güvenilirliği belirlemek için kullanılır. Ancak, test maddelerinin farklı ağırlıklarla puanlandığı ya da test puanlarının şans başarısına göre düzeltildiği durumlarda bu formül uygun değildir. Bunun nedeni, K-R20 formülünün, maddeler arası kovaryansların eşit olduğu varsayımına dayanmasıdır (Ntumi vd. 2023). Ek olarak, testteki tüm maddelerin aynı güçlük düzeyine sahip olduğu varsayıldığında, K-R20’nin özel bir versiyonu olan Kuder-Richardson 21 (K-R21) formülü kullanılır. Bu formül 2.10’da belirtilmektedir.(Özen vd. 2006).

Kuder-Richardson 21 (K-R21) formülü, test puanlarının ortalaması, standart sapması ve madde sayısına bağlı olarak hesaplandığı için madde analizi yapılmamış testlerde de kullanılabilir (Setyaedhi 2024). Buna karşılık, Kuder-Richardson 20 (K-R20) formülü, madde düzeyinde analiz yapılmasını gerektirdiğinden, madde analizi yapılmamış testlere uygulanamaz (Özen vd. 2006).

$$K21 = \frac{K}{K-1} \left(1 - \frac{K\bar{X} - \bar{X}^2}{k \cdot S_x^2} \right) \quad (2.10)$$

K21: Testin güvenilirlik katsayısı

K: Testte bulunan madde sayısı

$\bar{X}$: Testin ortalaması

S_x^2 :Test puanları dağılım varyansı

Formül 2.11’de gösterilen bir testin ortalama güçlüğü, testteki maddelerin güçlük indekslerinin aritmetik ortalaması olarak tanımlanır (Tekin 1977). Testlerin farklı amaçlar doğrultusunda kullanılabileceği daha önce ifade edilmiştir. Bu doğrultuda, bir testin ortalama güçlük düzeyi ile test maddelerinin güçlük dağılımı, testin kullanım amacına uygun olarak tasarlanmalıdır. Özellikle, testin çok sayıda öğrenci arasından az sayıda öğrenciyi seçmek amacıyla kullanılması durumunda, testin güçlük düzeyi bu hedefe uygun şekilde düzenlenmelidir (Baykul 2000).

$$\bar{P} = \frac{\bar{X}}{K} \quad (2.11)$$

$\bar{P}$: Testin ortalama güçlüğü

$\bar{X}$: Güçlük indekslerinin aritmetik ortalaması

K : Testten alınabilecek en yüksek puan

Bahsi geçen maddelerin güçlük düzeyinin her madde için hesaplanması gerekmektedir. Madde analizi bölümü buradaki değerlerin nasıl hesaplanması gerektiğini ele almaktadır. Yukarıdaki hesaplamalar doğrultusunda bir değerlendirme yapılacak olursa (Aydın

2013): Testin ortalama güçlük düzeyi 0.50'den küçükse, test zordur. Sınıf genellikle başarısız öğrencilerden oluşmaktadır. Test, öğrencilere zor gelmiştir ve eğitim seviyesi yeterli olmayabilir. Testin ortalama güçlük düzeyi 0.50'den büyükse, öğrenciler, konuları kavrama düzeyine sahiptir. Eğitim seviyesi yeterli düzeydedir ve test kolay sorulardan oluştuğu için güçlük düzeyi düşük olarak değerlendirilir.

2.2.3 Madde Analizi

Madde analizi, belirli niteliklere sahip olması beklenen bir testin içeriğinde yer alacak maddeleri değerlendirmeyi amaçlayan sistematik bir süreçtir. Bu sürecin etkin bir şekilde gerçekleştirilebilmesi için, geliştirilmekte olan testin nihai formuna dahil edilecek maddelerin analiz edilmesinden önce bir ön uygulama yapılması ve bu uygulamanın sonuçlarının elde edilmesi gereklidir (Danuwijaya 2018). Madde analizi, test maddelerinden elde edilen sonuçların belirlenen ölçütlere göre uygunluğunu değerlendirmek, uygun bulunmayan maddelerin nedenlerini belirlemek, testin amacına uygun olmasını sağlamak için gerekli düzenlemeleri planlamak ve bu düzenlemeleri hayata geçirmekten oluşur (Tomak 2013).

Madde analizine ilişkin önceki açıklamalarda sürecin temel işlevi, uygulanma amacı ve değerlendirme boyutu ele alınırken, aşağıda sunulan ifade ise madde analizinin uygulama düzeyine ilişkin planlama aşamasına odaklanmaktadır. Bu çerçevede, test madde analizi; belirli niteliklere sahip olması beklenen bir teste dahil edilecek maddelerin seçimini amaçlayan sistematik bir süreçtir. Belirli bir amaç için geliştirilecek bir testin nihai formuna eklenecek maddeler, testin uygulanacağı gruba benzer özelliklere sahip bir örneklem üzerinde, gerçek uygulama koşullarına yakın şartlarda gerçekleştirilen maddeler arasından seçilir (Taşdemir 2003).

Bir öğrenci grubunun bir testin maddelerine verdikleri yanıtların analizi, hem testin genel yapısı hem de tek tek maddelerin özellikleri hakkında değerli bilgiler sunar. Bu analiz sayesinde testin istatistiksel özellikleri tahmin edilebilir. Elde edilen bulgular doğrultusunda, istenilen niteliklere sahip maddeler bir araya getirilerek yeni bir test oluşturulabilir veya maddelerde tespit edilen kusurlar düzeltilerek testin kalitesi artırılabilir (Taşdemir 2003).

Test madde analizi süreci, testin uygulanacağı grubun özelliklerini taşıyan bir örneklem

üzerinde gerçekleştirilen test sonuçlarının aşağıdaki işlemlere tabi tutulmasıyla gerçekleştirilir (Taşdemir 2003):

- a. Veriler, büyüklük sırasına göre sıralanır.
- b. Toplam test (öğrenci) sayısı ve bu toplamın %27'lik oranı hesaplanır. Bu hesaplama, $(27N / 100)$ formülü ile yapılır.
- c. Sıralanmış verilerden üst %27'lik ve alt %27'lik dilimler belirlenir. Eğer %27'lik dilimin sınır değerlerine denk gelen ölçümlerde tekrar eden puanlar varsa, bu puanlardan yalnızca ihtiyaç duyulan sayı kadar öğrenci (veya kağıt) rastgele seçilir. Eşit puana sahip diğer kağıtlar analize dahil edilmez.
- d. Alt ve üst %27'lik dilimlere giren öğrencilerin her bir maddeye verdikleri yanıtlar, hazırlanan bir form üzerinde kaydedilir ve bu veriler istatistiksel analizlere tabi tutulur.
- e. Bu analiz verileri, her test maddesi için ayrı ayrı tablolar şeklinde düzenlenebileceği gibi tüm test maddelerini içeren tek bir tablo halinde de sunulabilir.

Bu yöntem, test maddelerinin istatistiksel açıdan değerlendirilmesini ve testin genel özelliklerinin analizini sağlar (Aydın 2013).

Bir testteki maddelerin zorluk dereceleri, belirli sayısal verilere dayanarak hesaplanır. Madde güçlük indeksi, bir maddenin uygulandığı grupta doğru yanıtlanma oranını ifade eder ve "P" ile gösterilir. Bu indeksin değerleri 0 ile 1 arasında değişir (Dündar 2019).

Her bir test maddesi, değerlendirdiği öğrenme hedefi ve öğretim üyesinin pedagojik amacı doğrultusunda analiz edilebilir (Sowdani vd. 2018). Eğer öğrenme hedefi temel düzeyde bir kavramı içeriyorsa, öğretim üyeleri, öğrencilerin %90 ila %100'ünün maddeye doğru yanıt vermesini bekleyebilir. Ancak, hedef daha karmaşık veya zorlayıcı nitelikteyse, %30'un altında bir madde zorluk düzeyi öğretim üyeleri tarafından kabul edilebilir olarak değerlendirilebilir (Rush vd. 2016). Genel bir değerlendirme kuralı olarak, %75'in üzerindeki bir zorluk düzeyi "kolay", %25 ile %75 arasındaki düzey "orta", %25'in altındaki düzey ise "zor" olarak sınıflandırılmaktadır (Sowdani vd. 2018). Bununla birlikte, aşırı kolay veya aşırı zor maddeler, madde ayırt ediciliği konusunda sınırlı bilgi sunmakta ve genellikle ölçme açısından daha az faydalı olmaktadır (Towns, 2014).

P değerinin 0'a yakın olması, maddenin zor olduğunu (Arıkan vd. 2014),

P değerinin 1'e yakın olması, maddenin kolay olduğunu (Arıkan vd. 2014),

P değerinin 1 olması, maddenin tüm katılımcılar tarafından doğru yanıtlandığını (Arıkan vd. 2014),

P değerinin 0 olması, maddenin hiçbir katılımcı tarafından doğru yanıtlanmadığını gösterir (Arıkan vd. 2014).

Bu değerler, maddelerin zorluk seviyelerinin analiz edilmesinde temel bir gösterge olarak kullanılır (Arıkan vd. 2014). Bahsi geçen güçlük indeksinin matematiksel ifadesi 2.12'de gösterilmiştir.

$$p_j = \frac{n(D)}{N} \quad (2.12)$$

p_j : Maddenin güçlük indeksi

$n(D)$: Maddeyi doğru yanıtlayanların sayısı

N : Testi alan tüm öğrencilerin sayısı

Eğer bir gruplama mevcut ise 2.13'de belirtilen formül kullanılır.

$$p_j = \frac{Dü + Da}{2N} \quad (2.13)$$

p_j : Maddenin güçlük indeksi

$Dü$: Üst grupta maddeyi doğru cevaplayanların sayısı

Da : Alt grupta maddeyi doğru cevaplayanların sayısı

N : Bir gruptaki öğrenci sayısı

Bir sorunun ayırt ediciliği, yoklanan davranışa sahip olan öğrenciler ile bu davranışa sahip olmayan öğrencileri ayırma gücünü ifade eder ve bu güç, korelasyon istatistikleri ile ölçülür. Bu araştırmada, madde ve test puanlarının sürekli değişkenlerden oluşması nedeniyle madde ayırt edicilik indeksleri Pearson momentler çarpımı korelasyon katsayısı (r_{jx}) ile hesaplanmıştır.

Formül 2.14 de bahsi geçen bu yöntemde, değerlendirme yapan kişi, öğrencileri bireysel puan tablosuna göre iki gruba ayırmaktadır: "yüksek başarı düzeyine sahip" ve "düşük başarı düzeyine sahip" gruplar. Bu sınıflandırma çerçevesinde, öğrencilerin %27'si

“güçlü grup” olarak, %27’si ise “zayıf grup” olarak tanımlanmaktadır. Bazı yaklaşımlar, öğrencilerin en üst üçte birlik ve en alt üçte birlik dilimler temelinde gruplandırılmasını önermekle birlikte, yapılan araştırmalar, %27’lik gruplar temelinde gerçekleştirilen ayırımın, D değerinin duyarlılığını ve doğruluğunu artırdığını ortaya koymaktadır (Tavakol ve Dennick 2011).

$$r_{jx} = \frac{D\ddot{u} - Da}{N} \quad (2.14)$$

r_{jx} : Madde ayırt edicilik indeksi

$D\ddot{u}$: Üst grupta maddeyi doğru cevaplayanların sayısı

Da : Alt grupta maddeyi doğru cevaplayanların sayısı

N : Bir gruptaki öğrenci sayısı

(r_{jx}) bir korelasyon katsayısı olarak (-1.00) ile (+1.00) arasında değerler alır. Pozitif değerler, maddenin testin bütünü ile aynı yönde bir ilişki içinde olduğunu; (0.00) değeri, madde ve test arasında herhangi bir ilişkinin bulunmadığını; negatif değerler ise madde ile testin ölçtüğü değişkenler arasında ters yönlü bir ilişki olduğunu gösterir. (r_{jx}) korelasyonunun (+1.00)’e yakın olması tercih edilir, çünkü bu durum, maddenin ölçülen özellik bakımından öğrenciler arasında etkili bir ayırım yapabildiğini gösterir (Acar, 2018). Çizelge 2.2’de korelasyon katsayısına göre elde edilen sonuçların örnekleri yer almaktadır.

Çizelge 2.2 Test Maddelerinin Ayırcılık Gücü İndeksine Dayalı Değerlendirilmesi (Yılmaz 2012).

Maddenin Ayırcı Gücü	Maddenin Değerlendirilmesi
0,40 ve daha yüksek	Oldukça yüksek ayırt ediciliğe sahip bir madde
0,30 – 0,39 arası	Yüksek ancak iyileştirilmeye açık bir madde
0,20 – 0,29 arası	Revize edilmesi ve geliştirilmesi gereken bir madde
0,19 ve daha düşük	Düşük ayırt ediciliğe sahip ve testten çıkarılması önerilen bir madde

Madde varyansı ve standart sapması, bir maddeye verilen yanıtların (0-1 puanlarının) değişim derecesini ifade eden istatistiksel değerlerdir. Kısaca, madde puanlarının değişkenliğini ölçen göstergelerdir. Bu göstergelerin matematiksel ifadesi formül 2.15 ve formül 2.16'da gösterilmiştir. (Güler 2011).

$$s_j^2 = p_j(1 - p_j) \quad (2.15)$$

s_j^2 : Madde varyansı

p_j : Madde güçlük indeksi

$$s_j = \sqrt{p_j(1 - p_j)} \quad (2.16)$$

s_j : Madde standart sapması

p_j : Madde güçlük indeksi

Madde güvenilirliğini ifade eden ve formül 2.17 de belirtilen madde güvenirlik katsayısı, madde ayırt edicilik gücü ile madde standart sapmasının çarpımına eşittir. Madde güvenirlik katsayısı (indeksi) aşağıdaki formül kullanılarak hesaplanabilir (Baykul ve Güzeller 2013):

$$r_j = r_{jx} \cdot s_j \quad (2.17)$$

r_j : Madde güvenirlik katsayısı

r_{jx} : Madde ayırt edicilik indeksi

s_j : Madde standart sapma

Formülden açıkça görüleceği üzere, madde güvenirlik indeksi, madde ayırt edicilik gücü (r_{jx}) ve madde standart sapmasının (s_j) bir fonksiyonudur. Bu nedenle, madde standart sapması ve madde ayırt edicilik gücü arttıkça madde güvenirliği de artış gösterir (Alıcı vd. 2014).

Madde güvenirliğinin değerlendirilmesinde, bir test maddesinin ölçmek istediği özelliği ne derece doğru ve tutarlı şekilde ölçtüğü önemlidir. Bu değerlendirme yapılırken, yalnızca madde düzeyinde ayırt edicilik ve standart sapma gibi ölçütler yeterli olmayabilir. Çünkü testin genel güvenirliği, yalnızca tekil maddelere değil, aynı zamanda bu maddelerin birbirleriyle ve toplam puanla olan ilişkilerine de bağlıdır. Bu

noktada korelasyon katsayısı devreye girer. Korelasyon katsayısı, bir maddenin ya da testin geneliyle ne ölçüde ilişkili olduğunu sayısal olarak ortaya koymakta ve böylece güvenilirlik analizlerinin daha bütüncül bir çerçevede ele alınmasına katkı sağlamaktadır (Krus ve Ney 1987).

Korelasyon katsayısı, istatistikte iki değişken arasındaki ilişkinin derecesini gösteren bir sayısal değer olarak tanımlanır. Bu katsayı, incelenen değişkenler arasındaki benzerlik derecesine bağlı olarak yüksek veya düşük; pozitif (+) ya da negatif (-) olabilir (Çağlar 1970).

Örneğin:

- Aynı teste bir gün arayla verilen puanlar arasındaki korelasyon katsayısının (0.86) veya (0.90) olması, ilişkinin pozitif ve yüksek düzeyde olduğunu ifade eder.
- (0.30) ile (0.50) arasındaki değerler, orta düzeyde bir ilişkiyi gösterir.
- (0.25) ve altındaki değerler, düşük düzeyde bir ilişkiyi işaret eder, ancak pozitif (+) ise hâlâ olumlu olarak değerlendirilir.

Negatif korelasyon katsayıları (-) da aynı şekilde yorumlanır, ancak bu durumda ilişki ters yönlüdür. Matematiksel olarak korelasyon katsayısı, (-1.00) ile (+1.00) arasında değişir. Diğer bir ifadeyle, bu değer (+1.00)'den başlayarak (-1.00)'ye kadar uzanabilir (Çağlar 1970).

$$r_{pb} = \left(\frac{\bar{Y}_p - \bar{Y}_q}{s_y} \right) \cdot \sqrt{pq} \quad (2.18)$$

r_{pb} : Nokta Çift Serili Korelasyon Katsayısı

$\bar{Y}_p$: Sürekli değişkene ait ölçümlerin ortalaması

$\bar{Y}_q$: İki kategorili değişkene ait ölçümlerin ortalaması

s_y : Standart kayma

p : Sürekli değişkende dikkate alınan durum

q : $1 - p$

Korelasyon, nokta çift serili korelasyon ve biserial korelasyon olmak üzere iki temel kategoriye ayrılmaktadır. Madde ayırt edicilik indeksini hesaplamak için yaygın olarak

kullanılan bir diğer yöntem, matematiksel olarak ifade edilen nokta çift serili korelasyon katsayısıdır. Bu istatistik, bir testteki belirli bir sorunun (doğru ya da yanlış) toplam test puanı ile olan ilişkisini gösterir. Sorular, "doğru" yanıtlar için 1, "yanlış" yanıtlar için ise 0 olarak puanlanır. Testteki doğru cevapların toplamı, öğrencinin toplam puanını oluşturur. Her bir soru için madde ayırt edicilik indeksini hesaplamak amacıyla 2.18 de belirtilen formül kullanılır (Büyüköztürk vd. 2013). Bu yöntem, soruların öğrenciler arasında ayırt edicilik düzeylerini değerlendirmek için etkili bir araç sunar.

Örnek:

Nokta çift serili korelasyon katsayısı (r_{pb}), bir test maddesinin toplam test puanı ile olan ilişkisini ölçerek, maddenin ayırt edicilik gücünü belirlemek için kullanılır. Bu katsayı, özellikle doğru/yanlış şeklinde puanlanan testlerde, her bir maddenin ne derece ayırt edici olduğunu değerlendirmek amacıyla hesaplanır (Attali ve Fraenkel 2000).

Bir sınavda 10 öğrenciye 5 maddelik bir test uygulanmıştır. Aşağıda, her öğrencinin 3. maddeye verdiği yanıt (1=doğru, 0=yanlış) ve toplam test puanı verilmiştir:

Çizelge 2.3 Nokta Çift Serili Korelasyon Örnek Veriler

Öğrenci	3. Madde Yanıtı	Toplam Puan
A	1	5
B	1	4
C	0	3
D	1	5
E	0	2
F	0	3
G	1	4
H	0	2
I	1	5
J	0	1

Bu verilerle, 3. madde için r_{pb} katsayısı hesaplanabilir. Yüksek bir r_{pb} değeri, maddenin yüksek puan alan öğrenciler tarafından doğru yanıtlandığını ve düşük puan alan öğrenciler tarafından yanlış yanıtlandığını gösterir, bu da maddenin iyi bir ayırt ediciliğe sahip olduğunu belirtir (Attali ve Fraenkel 2000).

Formül 2.19 da matematiksel ifadesi geçen çift serili korelasyon katsayısı, sürekli bir değişkenin yapay olarak iki kategorili hale getirildiği durumlarda kullanılan bir yöntemdir ve özellikle eğitim ve psikoloji alanındaki test geliştirme çalışmalarında yaygın bir şekilde uygulanmaktadır. Test geliştirme sürecinde, madde olarak adlandırılan sorular doğru yanıtlar için 1, yanlış yanıtlar, yanıtı bırakılan sorular veya birden fazla seçeneğin işaretlenmesi durumunda ise 0 puan alacak şekilde puanlanır. Bu korelasyon katsayısı, Pearson momentler çarpımı korelasyon katsayısının bir türevi olarak geliştirilmiştir ve test maddelerinin ayırt edicilik gücünü belirlemede etkili bir araçtır (Baykul ve Güzeller 2013).

$$r_b = \left(\frac{\bar{Y}_p - \bar{Y}_q}{s_y} \right) \cdot \frac{pq}{y} \quad (2.19)$$

r_b : Çift serili korelasyon katsayısı

$\bar{Y}_p$: Sürekli değişkene ait ölçümlerin ortalaması

$\bar{Y}_q$: İki kategorili değişkene ait ölçümlerin ortalaması

s_y : Standart kayma

p : Sürekli değişkende dikkate alınan durum

q : $1 - p$

y : p ve q 'yu ayıran ordinat yüksekliği

2.3 Çoktan Seçmeli Testler

Çoktan seçmeli testler, cevaplayıcının sorunun doğru cevabı olma ihtimali bulunan seçenekler arasından birini seçerek yanıt vermesini gerektiren test türleridir. Doğru-yanlış testleri ise, seçenek sayısının iki ile sınırlı olması nedeniyle çoktan seçmeli testlerin bir alt kategorisi olarak değerlendirilmektedir (Aslan 2018). Çoktan seçmeli testlerde dikkat edilmesi gereken hususlar (Aslan 2018, İşeri 2023);

- Maddeler açık, yalın ve kolay anlaşılır bir dille yazılmalıdır.
- Her madde, ölçmek istediği davranışı doğrudan yoklamalı ve bu davranışı sergileyen kişilerin doğru cevabı verebilmeleri için gerekli bilgileri içermelidir.
- Madde kökü net, açık ve yoruma yer vermeyecek şekilde ifade edilmelidir.
- Maddede zorluk yaratmak amacıyla kökte belirsizlik veya yanıltıcı ifadeler kullanılmamalıdır. Eğer soru zorlaştırılacaksa, bu durum seçeneklerin doğru cevaba yakınlaştırılmasıyla sağlanmalıdır.
- Madde kökünde gereksiz ifadelerden kaçınılmalı ve yalnızca ölçülmek istenen davranışı açıkça ortaya koyan ifadeler yer almalıdır.
- Seçeneklerden herhangi birine ipucu oluşturabilecek ifadeler madde kökünde kullanılmamalıdır.
- Tekrar eden ifadeler madde kökünde toplanmalı, seçenekler mümkün olduğunca kısa ve öz yazılmalıdır.
- Seçeneklerin uzunlukları, kapsamaları ve ifade tarzları birbirine uyumlu olmalıdır.
- Çeldiriciler (yanıltıcı seçenekler), doğru cevaba yakın olacak şekilde tasarlanmalı ve maddenin hedeflenen zorluk düzeyiyle uyumlu olmalıdır.
- Doğru cevap seçenekleri, sorulara yaklaşık olarak eşit dağıtılmalıdır.
- Aynı doğru cevaba (örneğin "A" seçeneği) art arda çok kez yer verilmemelidir.
- Seçeneklerin yerleşimi (örneğin "A, A, B, B, C, C" gibi bir örüntü) belirli bir düzeni takip etmemelidir. Aksi takdirde bazı cevaplayıcılar bu örüntüyü fark ederek soruyu okumadan doğru cevabı bulabilir.
- "Yukarıdakilerin hepsi" seçeneği: Seçenekler birbirine çok yakınsa kullanılabilir. Ancak bir seçeneğin yanlış olduğu açıksa bu seçenek elenmelidir.
- "Yukarıdakilerin hiçbiri" seçeneği: Bu seçenek doğru cevap olarak kullanılmamalıdır, çünkü bu durumda cevaplayıcının davranışı sergileyip sergilemediği anlaşılamaz. Ancak kesin bir doğru cevabı olan sorularda çeldirici olarak kullanılabilir.
- Aynı madde içerisinde "Yukarıdakilerin hepsi" ve "Yukarıdakilerin hiçbiri" seçenekleri birlikte yer almamalıdır. Mantıksal olarak bu iki seçeneğin aynı anda doğru olması mümkün değildir.
- Çeldiriciler, davranışı sergilemeyen cevaplayıcıları yanıltmalı ancak davranışı sergileyenleri yanıltmamalıdır.

- Çeldirici oluřturulmasında güçlük yařanıyorsa, madde formu yeniden gözden geçirilmeli veya deęiřtirilmelidir.
- Olumsuz ifadelerin olumlu algılanma riski göz önünde bulundurulmalıdır. Bu tür ifadeler altı çizilerek, koyu yazılarak veya başka yollarla dikkat çekici hale getirilmelidir.
- Maddeler arasında ve madde kökü ile seçenekler arasında yeterli boşluk bırakılmalıdır.
- Ortak köklü veya ortak seçenekli maddeler, başlangıç ve bitişlerinde dięer maddelerden ayrılmalı ve aynı sayfa üzerinde düzenlenerek dikkat daęınıklığı önlenmelidir.
- Maddelerdeki ifadeler, bilinen bir kaynaktan doğrudan alınmamalı, özgün bir şekilde oluřturulmalıdır.
- Seçenek sayısı, testin uygulandıęı grubun düzeyine uygun olmalıdır. Örneęin, ilkokullarda 4, ortaokul ve üzerindeki düzeylerde 5 seçenek tercih edilmelidir.
- Testteki tüm maddelerin seçenek sayısı aynı olmalıdır.
- Testin ilk soruları kolay seçilerek cevaplayıcının güdülenmesi saęlanmalıdır.
- Yazı boyutu, cevaplayıcıların düzeyine uygun büyüklükte olmalıdır.
- Bir test maddesi dięerlerinden bağımsız olarak cevaplanabilecek şekilde tasarlanmalıdır.
- Her test maddesi önemli bir davranışı yoklamalıdır.
- Birbirinin zıddı olan seçenekler aynı maddede birlikte yer almamalıdır.
- Seçenekler arasında büyüklük sırası varsa, bu sıralama mantıksal olarak düzenlenmelidir.

2.4 Kullanılabilirlik Kavramı ve Tespiti

Kullanılabilirlik, bir ürünün kalitesi, verimlilięi ve kullanıcı memnuniyeti açısından önemli bir kavramdır ve genellikle "kullanıcı dostu" ile "kullanım kolaylığı" terimleriyle ilişkilendirilir. Kullanıcı ihtiyaçlarına ve bireysel deęerlendirmelere odaklanan kullanılabilirlik, uluslararası standartlarda üç ana bileşenle tanımlanmaktadır: “etkililik”, “verimlilik” ve “tatmin” (Telek 2013).

Etkililik, bir ürün veya sistemin belirlenen hedeflere ulaşmada gerekli işlevleri ne ölçüde yerine getirdiğini ifade eder. Verimlilik, kullanıcıların bilişsel ve fiziksel yeterliliklerini dikkate alarak bu hedeflere etkili bir şekilde ulaşmalarını saęlar. Tatmin ise kullanıcıların

ürünle etkileşimlerinde beklentilerinin karşılanmasını ve olumlu bir deneyim yaşamalarını amaçlar. Bu üç bileşen, kullanılabilirliğin temelini oluşturarak kullanıcı odaklı tasarımlara rehberlik eder (Telek 2013).

2.4.1 Kullanılabilirliğin Bileşenleri

Kullanılabilirlik, tanımlanması ve ölçülmesi güç olan karmaşık bir kavram olup; kullanıcı, ürün, bağlam ve çevresel koşulların etkileşimiyle şekillenmektedir. Çok boyutlu bir yapıya sahip olan bu kavram, farklı disiplinlerce çeşitli açılardan ele alınmaktadır. Uluslararası Standardizasyon Örgütü (ISO), kullanılabilirliği; etkililik, verimlilik ve kullanıcı memnuniyeti bileşenleriyle ilişkilendirmektedir (Bevan 1995). Nielsen (1993) ise sistemlerin daha kullanılabilir şekilde tasarlanabilmesi için beş temel faktörün göz önünde bulundurulması gerektiğini vurgulamaktadır: öğrenme kolaylığı, kullanım etkinliği, hatırlanabilirlik, hata oranının düşüklüğü ve kullanım tatmini. Bu doğrultuda, kullanılabilirlik yalnızca tek bir özelliğe indirgenemeyip, söz konusu faktörlerin bütünsel olarak değerlendirilmesiyle anlam kazanmaktadır (Nielsen 1993).

2.4.2 Kullanılabilirlik Ölçümleri

Kullanılabilirliğin değerlendirilmesinde üç temel yöntem öne çıkmaktadır: sorgulamaya dayalı yaklaşımlar, uzman görüşüne dayalı inceleme yöntemleri (sezgisel değerlendirme) ve doğrudan kullanıcılarla gerçekleştirilen testler. Bu sınıflandırma, farklı veri toplama teknikleri ve değerlendirme perspektiflerini bir araya getirerek, sistemin kullanıcı deneyimi açısından çok yönlü biçimde analiz edilmesine olanak tanır (Gürses 2005).

Sorgulamaya Dayalı Yöntemler: Bu yöntemde, anket, mülakat ve odak grup görüşmeleri gibi teknikler kullanılarak kullanıcıların ürün ya da sistem hakkındaki düşünceleri toplanır. Uygulaması kolay ve düşük maliyetli olması avantaj sağlarken, gerçek kullanım koşullarını yansıtmadığı için elde edilen verilerin güvenilirliği sınırlı olabilir (Chao ve Chiu 2010).

İncelemeye Dayalı Yöntemler: Kullanılabilirlik değerlendirmesi, deneyimli uzmanlar tarafından sistemin kullanıcı gözüyle incelenmesi yoluyla gerçekleştirilir. Bu değerlendirme, genellikle sistemin tasarım ve geliştirme aşamalarında uygulanır.

Uzmanlar kendi bilgi birikimlerine dayanarak muhtemel kullanılabilirlik sorunlarını belirler. Değerlendirme kriterlerinin açık ve nesnel olması, birden fazla uzmanın katılımıyla yapılması yöntemin güvenilirliğini artırır (Nielsen & Molich 1990).

Kullanıcı testleri: Sistemi temsil eden bireylerle gerçek ortamda yürütülerek kullanıcı-arayüz etkileşimine dair veriler toplanmasını sağlar. Katılımcılara belirli görevler verilerek gözlemler yapılır, yazılım kullanım süreçleri izlenir ve performans verileri toplanır. Gelişmiş laboratuvar ortamlarında gerçekleştirilen bu testlerde, göz izleme cihazları gibi teknolojilerle kullanıcının dikkat dağılımı ve görev başarımı ölçülür. Elde edilen sonuçlar, sistemin güçlü ve zayıf yönlerinin nesnel biçimde değerlendirilmesine imkân tanır. Ayrıca, bu testlerin güvenilirliği ve geçerliliği değerlendirme sürecinin bütünlüğü açısından önemlidir (Szekely vd. 2023).

2.5 Konu İle İlgili Daha Önce Yapılmış Çalışmalar

Rehman, Aslam ve Hassan (2018) tarafından yürütülen çalışmada, diş hekimliği öğrencilerine yönelik uygulanan çoktan seçmeli test maddelerinin kalitesi madde analizi yoluyla değerlendirilmiştir. Araştırma sonuçlarına göre, testte yer alan maddelerin %52,5'i "çok zor" kategorisinde yer almakta, %67,5'i ise düşük ayırt edicilik katsayısına sahiptir. Ayrıca çeldiricilerin %51,6'sı işlevsel bulunmamıştır. Bu bulgular, test maddelerinin ölçme gücünü artırmak için madde analizinin düzenli biçimde uygulanmasının önemini göstermektedir. Araştırma, özellikle çoktan seçmeli testlerin güvenilirliğini ve geçerliğini artırmak için içeriksel kalitenin ve madde yapısının titizlikle incelenmesi gerektiğini ortaya koymaktadır (Rehman vd. 2018).

Tavakol ve Dennick (2011) tarafından yapılan çalışmada, çoktan seçmeli testlerin geçerlik ve güvenilirliğinin artırılması amacıyla sınav sonrası analiz süreçlerinin önemine dikkat çekilmiştir. Araştırmacılar, özellikle Cronbach alfa katsayısının testin iç tutarlılığını değerlendirmede etkili bir araç olduğunu ve madde-toplam korelasyonlarının bireysel maddelerin ayırt ediciliğini belirlemek açısından kritik olduğunu vurgulamaktadır. Çalışma, madde analizinin yalnızca zayıf maddelerin elenmesini değil, aynı zamanda testin genel yapısal bütünlüğünü ve ölçme gücünü artırmaya yönelik

stratejik bir araç olarak kullanılmasını önermektedir. Bu bağlamda, sınav sonrası analizlerin sistematik biçimde uygulanması, testlerin hem geçerlik hem de güvenilirlik düzeylerini artırarak eğitimde daha sağlıklı ölçme-değerlendirme süreçlerine katkı sunmaktadır (Tavakol ve Dennick 2011).

Imtiazuddin ve arkadaşları (2020) tarafından yürütülen çalışmada, farmakoloji alanında uygulanan çoktan seçmeli testler, madde analizine tabi tutularak ölçme aracının niteliği değerlendirilmiştir. Araştırmada yer alan 50 maddenin %66'sının orta güçlükte, %30'unun ise yüksek güçlük düzeyinde olduğu tespit edilmiştir. Yüksek güçlük düzeyinin başlıca nedenleri arasında madde yazım hataları (%41), öğretim eksiklikleri (%36) ve üst düzey bilişsel hedeflerin ölçülmeye çalışılması (%23) gösterilmiştir. Discrimination Index sonuçları, maddelerin %52'sinin yüksek ayırt edicilik seviyesinde olduğunu ortaya koyarken, %16'sının ise düşük ayırt ediciliğe sahip olduğu belirlenmiştir. Ayrıca, çeldirici analizinde maddelerin %52'sinde tüm seçeneklerin işlevsel olduğu görülmüş, bu da testin yapısal açıdan genel olarak yeterli olduğunu göstermektedir. Araştırmacılar, testlerin kalite düzeyinin artırılması için sistematik madde analizinin düzenli biçimde uygulanması gerektiğini vurgulamışlardır (Imtiazuddin vd. 2020).

Rahma ve arkadaşları (2017) tarafından yapılan çalışmada, çoktan seçmeli testlerde üç ve dört seçenekli soru formatlarının psikometrik özellikleri karşılaştırılmıştır. Araştırmada, dört seçenekli sorularda daha fazla çeldirici işlevsiz bulunurken, üç seçenekli soruların ayırt edicilik düzeylerinin daha yüksek olduğu tespit edilmiştir. Testlerin iç tutarlılıkları Kuder-Richardson (KR-20) formülüyle değerlendirilmiş ve her iki formatın da benzer düzeyde güvenilirliğe sahip olduğu (0.82 ve 0.83) belirlenmiştir. Ayrıca, üç seçenekli testlerde, öğrencilerin ayırt edilme düzeyinin daha yüksek olduğu, ancak güçlük indeksinin de daha düşük seviyede gerçekleştiği gözlemlenmiştir. Bu bulgular, fazla sayıda işlevsiz çeldirici yerine daha az ama nitelikli seçenekler içeren maddelerin kullanılmasının test kalitesini artırabileceğini göstermektedir (Rahma vd. 2017).

Kar, Lakshminarayanan ve Mahalakshmy (2015) tarafından gerçekleştirilen çalışmada, çoktan seçmeli soruların geçerli ve güvenilir olabilmesi için dikkat edilmesi gereken

temel ilkeler sistematik şekilde ele alınmıştır. Araştırma, etkili MCQ yazımında çeldiricilerin işlevselliği, madde zorluğu ve ayırt ediciliğinin testin ölçme gücünü doğrudan etkilediğini vurgulamaktadır. Ayrıca, hatalı yazılmış soruların ölçme sürecini olumsuz etkilediği, bu nedenle madde analizlerinin hem ön test (pre-validation) hem de son test (post-validation) aşamalarında uygulanmasının gerekliliği ortaya konmuştur. Çalışma ayrıca, üst düzey bilişsel becerilerin (analiz, sentez, değerlendirme) ölçülmesine yönelik olarak Bloom Taksonomisi temelli soru yazımına ilişkin örnekler sunmuş ve bu yaklaşımın eğitimde derin öğrenmeyi teşvik edici olduğunu ifade etmiştir (Kar vd. 2015).

Haladyna (2004) tarafından hazırlanan kapsamlı kaynakta, çoktan seçmeli testlerin etkili biçimde geliştirilmesi, yapılandırılması ve geçerlik-güvenirlilik düzeylerinin artırılmasına yönelik sistematik ilkeler sunulmaktadır. Çalışmada, madde yazımında yapılan yaygın hataların testin ölçme kalitesini düşürdüğü; işlevsel olmayan çeldiricilerin hem testin zorluk derecesini yanlış yansıttığı hem de ayırt ediciliği azalttığı belirtilmiştir. Haladyna, madde analizine dayalı geliştirmenin, testlerin eğitimsel amaçlarla tutarlı ve psikometrik açıdan güçlü araçlara dönüşmesi için vazgeçilmez olduğunu vurgulamaktadır. Ayrıca testlerin, sadece bilgi düzeyini değil, analiz ve değerlendirme gibi yüksek düzey bilişsel süreçleri de ölçebilecek biçimde yapılandırılması gerektiği ifade edilmektedir. Bu yönüyle çalışma, ölçme aracı kalitesini artırmaya yönelik teorik ve uygulamalı çerçeveler sunmaktadır (Haladyna 2004).

Kaur, Singla ve Mahajan (2016) tarafından yapılan çalışmada, farmakoloji dersine ait çoktan seçmeli sorular, zorluk düzeyi (DIF I), ayırt edicilik (DI) ve çeldirici etkinliği (DE) açısından analiz edilmiştir. Araştırma sonuçlarına göre, testteki 50 maddeden %76'sı kabul edilebilir zorluk düzeyinde, %22'si fazla kolay ve yalnızca %2'si fazla zor olarak sınıflandırılmıştır. Ayırt edicilik açısından değerlendirildiğinde ise, maddelerin %62'sinin yüksek ayırt ediciliğe, %24'ünün orta düzeyde ayırt ediciliğe, %14'ünün ise düşük ayırt ediciliğe sahip olduğu saptanmıştır. Çeldirici etkinliği analizinde toplam 150 çeldiriciden %82'si işlevsel bulunmuştur. Bu verilere dayanarak 6 madde elenmiş, 17 madde revize edilmiş ve 27 madde daha sonraki kullanımlar için uygun görülmüştür. Çalışma, madde analizinin test kalitesini artırmada ve soru bankalarının geliştirilmesinde önemli bir araç olduğunu vurgulamaktadır (Kaur vd. 2016).

Bhattacharjee ve arkadaşları (2022) tarafından yürütülen çalışmada, COVID-19 pandemisi sırasında çevrim içi olarak uygulanan çoktan seçmeli bir sınavın madde analizine tabi tutulması yoluyla, sınav maddelerinin psikometrik özellikleri değerlendirilmiştir. Araştırmada kullanılan 60 madde, zorluk indeksi (Dif I), ayırt edicilik indeksi (DI) ve çeldirici etkinliği (DE) açısından analiz edilmiştir. Bulgular, maddelerin %15'inin zor, %46.67'sinin kolay ve geri kalanının orta düzeyde olduğunu ortaya koymuştur. Maddelerin %70'i düşük ayırt edicilik göstermiş, %6.67'sinde ise negatif ayırt edicilik tespit edilmiştir. Çeldirici analizi ise 180 çeldiricinin %51.66'sının işlevsiz olduğunu göstermiştir. Çalışma, madde analizinin yalnızca geçerlik ve güvenilirliği artırmakla kalmayıp, öğretim üyelerine soru hazırlama konusunda geri bildirim sunduğunu ve ölçme aracının geliştirilmesine katkı sağladığını vurgulamaktadır (Bhattacharjee vd. 2022).

Christian ve arkadaşları (2017) tarafından yapılan çalışmada, tıp fakültesi stajyerlerinin katılımıyla uygulanan çoktan seçmeli test maddeleri, madde analizi yöntemleri kullanılarak değerlendirilmiştir. Çalışma kapsamında madde güçlük indeksi (p-değeri), ayırt edicilik indeksi ve çeldirici etkililiği gibi temel ölçütler incelenmiştir. Bulgular, yüksek ayırt ediciliğe ve işlevsel çeldiricilere sahip ortalama güçlükteki soruların, testin geçerliğini artırma potansiyeli taşıdığını göstermiştir. Ayrıca, negatif ayırt ediciliğe sahip maddelerin testin geçerliği açısından sakıncalı olduğu ve değerlendirme sürecinden çıkarılması gerektiği vurgulanmıştır. Bu yönüyle çalışma, çoktan seçmeli test maddelerinin değerlendirilmesinde sistematik analizlerin uygulanmasının önemine dikkat çekmektedir (Christian 2017).

Kumar ve arkadaşları (2021) tarafından yapılan bir çalışmada, çoktan seçmeli testlerin değerlendirilmesinde kullanılan madde analizinin geçerlik ve güvenilirlik açısından önemli bir kalite güvence aracı olduğu vurgulanmıştır. Araştırmada 150 tıp fakültesi öğrencisinin katıldığı üç farklı sınavdaki toplam 90 soruya ilişkin güçlük, ayırt edicilik ve çeldirici etkinliği analiz edilmiştir. Elde edilen bulgulara göre soruların %82'si uygun güçlük düzeyine sahipken, %80'i iyi ya da kabul edilebilir ayırt ediciliğe sahip bulunmuştur. Çeldirici analizlerinde ise işlevsel çeldirici oranının %73 olduğu belirlenmiştir. Çalışmanın sonucunda, kaliteli soru havuzları oluşturmak ve değerlendirme araçlarının

geçerliliğini artırmak amacıyla tüm çoktan seçmeli sınavlara madde analizi uygulanması önerilmiştir. Bu çalışma, çoktan seçmeli test analizinde kullanılan istatistiksel yöntemlerin ve ölçütlerin etkinliğini ortaya koyarak, ölçme araçlarının geliştirilmesine yönelik önemli katkılar sağlamaktadır (Kumar vd. 2021).

Date ve arkadaşları (2019) tarafından yürütülen çalışmada, farmakoloji dersinde kullanılan çoktan seçmeli sorular madde analizi yoluyla değerlendirilmiş ve elde edilen veriler doğrultusunda soru bankasının yapılandırılması hedeflenmiştir. Çalışmada kullanılan 40 sorunun %70'i kabul edilebilir güçlük düzeyinde bulunurken, %77.5'inin ayırt edicilik düzeyi "kabul edilebilir" ile "mükemmel" arasında sınıflandırılmıştır. Çeldirici etkinliği açısından değerlendirildiğinde, toplam 120 çeldiricinin %70'i işlevsel bulunmuş, 16 soruda çeldiricilerin tamamı işlevsel olarak değerlendirilmiştir. Araştırma sonuçları, madde analizinin yalnızca düşük nitelikli soruların elenmesi değil, aynı zamanda yüksek kaliteli soru bankalarının oluşturulması sürecinde de kullanılabilecek geçerli bir yöntem olduğunu ortaya koymuştur (Date vd. 2019).

Yükseköğretimde çoktan seçmeli testlerin etkinliği uzun süredir tartışma konusu olmuştur. Fozzard ve arkadaşları tarafından yapılan kapsamlı bir çalışmada, beş seçenekli sorulardan dört seçenekliye geçişin öğrenci başarıları ve madde analizleri üzerindeki etkisi incelenmiştir. Araştırma sonucunda, seçenek sayısındaki azalmanın, soru güclüğü ve ayırt ediciliği üzerinde anlamlı bir olumsuz etkisi olmadığı belirlenmiştir. Ayrıca, çoğu test maddesinin dört etkili çeldirici içermediği görülmüş ve bu durum, geleneksel beş seçenekli soruların sorgulanmasına yol açmıştır. Bu bulgular, daha az sayıda fakat etkili çeldirici ile oluşturulan soruların da geçerli ve güvenilir sonuçlar verebileceğini ortaya koymuştur (Fozzard vd. 2018).

Barmanray ve arkadaşları (2022) tarafından gerçekleştirilen çalışmada, tıp öğrencilerine yönelik bir patoloji sınavında kullanılan çoktan seçmeli sorular; zorluk düzeyi, ayırt edicilik katsayısı, çeldirici işlevselliği ve içerik uygunluğu gibi ölçütlerle sistematik biçimde analiz edilmiştir. Araştırma bulgularına göre, maddelerin %52'si istenilen zorluk aralığında yer alırken, %38'i kolay ve %10'u zor kategorisinde değerlendirilmiştir. Ayırt edicilik indeksine göre soruların %58'i kabul edilebilir düzeyde bulunmuş, %24'ünün ise

düşük ayırt edici güce sahip olduğu belirlenmiştir. Ayrıca, soruların yalnızca %66'sı öğrenme hedefleri ile tamamen uyumlu bulunmuştur. Bu sonuçlar, MCQ'ların eğitimsel hedeflere uygun biçimde yapılandırılmasının ve madde analizlerinin sistematik uygulanmasının, sınav kalitesini artırmak için kritik bir öneme sahip olduğunu ortaya koymuştur (Barmanray vd. 2022).

Kheyami ve arkadaşları (2018) tarafından yapılan bir çalışmada, Bahreyn'deki Arabian Gulf Üniversitesi'nde uygulanan pediatri sınavlarındaki çoktan seçmeli test maddeleri analiz edilmiştir. Çalışma sonucunda, soruların zorluk (DIFI), ayırt edicilik (DI) ve çeldirici etkililiği (DE) açısından büyük oranda kabul edilebilir düzeyde olduğu, ancak bazı maddelerin yeniden düzenlenmesi veya çıkarılması gerektiği belirtilmiştir. Ayrıca, beş seçenekli sorular yerine üç ya da dört seçenekli sorular kullanılmasının, işlevsiz çeldirici oranını önemli ölçüde azalttığı ve test kalitesini artırdığı sonucuna ulaşılmıştır (Kheyami vd. 2018).

Towns (2014), çoktan seçmeli test maddelerinin oluşturulmasında akademik geçerlik ve güvenilirliğin sağlanmasına yönelik kapsamlı ilkeler sunmaktadır. Yazara göre, etkili test maddeleri oluşturulabilmesi için her bir sorunun belirli bir öğrenme hedefini ölçmesi gereklidir ve bu hedeflerin açıkça tanımlanmış olması gerekir. Ayrıca, madde yazımında yaygın yapılan hatalar, ölçme sonuçlarının geçerliliğini zedeleyebileceğinden dolayı, bu tür hataların önlenmesi adına yapılandırılmış madde yazım yönergeleri takip edilmelidir. Özellikle işlevsel olmayan çeldiricilerin testin ayırt ediciliğini azalttığı, bu nedenle her seçeneğin öğrenci açısından makul ve anlamlı olması gerektiği vurgulanmıştır. Ayrıca, yapılan araştırmalar üç seçenekli soruların hem madde kalitesi hem de testin genel ölçme gücü açısından yeterli olduğunu göstermiştir (Towns 2014).

Solak, Uçar ve Albadwieh (2020) tarafından geliştirilen CBE-MCQs adlı bilgisayar destekli değerlendirme yazılımı, çoktan seçmeli sorulara dayalı sınavları otomatik olarak analiz edebilmektedir. Yazılım, madde güçlüğü, ayırt edicilik ve distraktör etkinliği gibi klasik test kuramına dayalı analizleri gerçekleştirebilmekte; ayrıca ölçme sonuçlarının öğrenme çıktılarıyla ilişkisini değerlendirerek öğretim sürecine geribildirim

sağlamaktadır. Yapılan kullanılabilirlik anketi sonucunda, yazılım %87.4 oranında öğretim elemanları tarafından etkili ve kullanıcı dostu bulunmuştur (Solak et al 2020). Benzer şekilde, Moore ve arkadaşları (2024) tarafından geliştirilen SAQUET adlı sistem, çoktan seçmeli soruların kalitesini değerlendirmek amacıyla GPT-4 gibi büyük dil modellerini kullanarak, soru yazım hatalarını %94 doğrulukla tespit edebilmektedir. Bu sistem, insan değerlendirmeleriyle yüksek düzeyde uyum göstererek geleneksel otomatik değerlendirme metotlarına kıyasla daha derinlemesine analiz olanağı sunmaktadır (Moore et al 2024).

Vieira (2020) tarafından geliştirilen bulut tabanlı sınav üretim aracı, öğretmenlerin konu ve zorluk seviyesine göre sınavları otomatik biçimde oluşturabilmesini sağlamakta, sistemin kullanılabilirliği ise WEBUSE yöntemiyle değerlendirilmiştir. Kullanıcılar genel anlamda sistemi işlevsel bulmuş; ancak arayüz ve tasarım bağlamında bazı iyileştirme gereksinimlerine işaret etmişlerdir (Vieira 2020).

Goncher, Boles ve Jayalath (2014), çoktan seçmeli sorulara verilen yazılı açıklamaların otomatik olarak analiz edilmesini sağlayan bir model üzerinde çalışmışlardır. Bu model, öğrencilerin kavramsal anlayışını değerlendirirken açık uçlu yanıtları da hesaba katarak klasik testlerin sınırlılıklarını aşmayı hedeflemektedir (Goncher et al 2014).

Das ve arkadaşları (2021), öğrenme materyallerinden çoktan seçmeli sorular ve uygun distraktörler üretebilen bir sistem geliştirmiştir. Bu sistem, anahtar kavramlara dayalı olarak doğru ve yanlış seçenekler oluşturmakta ve e-öğrenme uygulamalarında değerlendirme sürecini hızlandırmaktadır (Das et al 2021).

Şahhüseyinoğlu (1998) tarafından yapılan çalışmada, sayısal yetenek testlerinde seçenek sayısındaki değişimin madde ve test istatistikleri üzerindeki etkileri incelenmiş ve şans başarısının bu etkiler üzerindeki rolü araştırılmıştır. Çalışmada, şans başarısının test istatistiklerinden arındırılması sonucunda güvenilirlik ve geçerlik katsayılarında genel bir artış gözlemlenmiştir. Ancak, şans başarısından arındırılmadan önce yapılan karşılaştırmalarda olduğu gibi, iki testin güvenilirlik ve geçerlik katsayıları arasında anlamlı bir fark bulunamamıştır. Testlerin ortalamaları, şans başarısının etkisinin ortadan kaldırılmasıyla bir miktar azalmış; buna karşın, seçenek sayısındaki azalmanın

ortalamlar üzerindeki artış eğilimini koruduğu belirlenmiştir. Madde güçlük indekslerinde ise şans başarısının arındırılmasıyla beklenen bir azalma meydana gelmiş ve bu bulguların, şans başarısının matematiksel bağıntıları ile uyumlu olduğu saptanmıştır (Şahhüseyinoğlu 1998).

Arslan (2011) beden eğitimi öğretmen adaylarının beden eğitimi derslerinde alternatif ölçme ve değerlendirme yaklaşımlarına yönelik genel algılarını, bu yaklaşımlar konusundaki yeterlik algılarını ve genel olarak eğitimde ölçme ve değerlendirme konusundaki yeterlik algılarını incelemeyi amaçlamaktadır. Araştırmanın bir diğer amacı ise, alternatif ölçme ve değerlendirme araçlarının kullanıldığı beden eğitimi derslerinin, öğrencilerin beden eğitimi derslerinde ölçme ve değerlendirme konusundaki algıları üzerinde bir etkisinin olup olmadığını belirlemektir. Araştırma sonuçlarına göre, Ölçme ve Değerlendirme Gelişim Programı'na (ÖDGP) katılan öğretmen adaylarının (deney grubu), beden eğitimi derslerine yönelik alternatif ölçme ve değerlendirme algı puanlarında, deneysel işlem öncesine kıyasla işlem sonrasında anlamlı bir artış tespit edilmiştir. Ayrıca, bu değişim, kontrol grubunda yer alan öğretmen adaylarının algı puanlarındaki değişimle karşılaştırıldığında istatistiksel olarak anlamlı bir farklılık göstermiştir. Elde edilen anlamlı fark, deney grubunun lehine bulunmuştur (Arslan 2011).

Aktürk (2012) tarafından gerçekleştirilen bu araştırma, sosyal bilgiler öğretmenlerinin sürece dayalı ölçme ve değerlendirme yöntemlerini kullanabilme durumlarını belirlemeyi amaçlamaktadır. Araştırma bulguları, öğretmenlerin bu yöntemleri kullanma sıklığının mezun oldukları bölüm veya ders verdikleri sınıfların öğrenci sayısına göre değişiklik göstermediğini, ancak kıdem ve mezun olunan fakülte faktörlerine bağlı olarak farklılık gösterdiğini ortaya koymuştur. Ayrıca, sosyal bilgiler öğretmenlerinin sürece dayalı ölçme ve değerlendirme yöntemlerini birbirleriyle kullanım durumlarının da değişkenlik gösterdiği belirlenmiştir (Aktürk 2012).

3. MATERYAL ve METOT

3.1 Araştırmanın Genel Uygulama Süreci

Bu çalışmanın temel amacı, çoktan seçmeli testlerin güvenilirlik analizini gerçekleştirebilecek web tabanlı bir sistemin geliştirilmesi ve geliştirilen sistemin kullanılabilirliğinin değerlendirilmesidir. Bu doğrultuda araştırma süreci iki temel aşamada ele alınmıştır: birincisi, sistemin yazılımsal ve işlevsel bileşenlerinin tasarlanarak geliştirilmesi süreci; ikincisi ise sistemin hedef kullanıcı kitlesi tarafından değerlendirilerek kullanılabilirliğine ilişkin verilerin toplanmasıdır (Shahid vd. 2023).

Geliştirme sürecinde, test güvenilirlik katsayılarının yanı sıra madde güçlüğü ve ayırt edicilik indekslerinin dijital ortamda hesaplanmasına olanak tanıyan bir yapı kurulmuştur. Web platformunun arayüz tasarımı, veri işleme algoritmaları ve kullanıcı deneyimini artırmaya yönelik bileşenler, araştırmanın birinci aşamasında planlı bir şekilde oluşturulmuştur. Bu kapsamda, sistemin hem teknik yeterliliği hem de kullanıcıya sunduğu işlevsellik esas alınarak bir yazılım altyapısı oluşturulmuştur (Yakunin ve Bodrunova 2022).

İkinci aşamada ise, geliştirilen sistemin kullanıcılar açısından ne ölçüde etkili, anlaşılır ve erişilebilir olduğuna yönelik analizler yapılmıştır. Kullanılabilirlik kriterleri çerçevesinde, sistemin uygulama sürecindeki performansı değerlendirilmeye çalışılmıştır. Tüm bu süreçte elde edilen bulgular, çalışmanın hem ürün geliştirme hem de kullanıcı geri bildirimi açısından bütüncül bir değerlendirme sunmasını sağlamıştır (Yakunin ve Bodrunova 2022).

3.2 Materyaller

- Web sayfasının hazırlanması sürecinde HTML 5, CSS 3, BOOTSTRAP 5 (Css Framework), JavaScript, Laravel 10 (PHP Framework), jQuery (JS Framework) ve MySQL tasarım ve programlama araçlarından faydalanılmıştır.

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.

3.2.1 HTML

HTML (Hyper Text Markup Language/Hiper Metin İşaretleme Dili), kökenleri internetin ilk yıllarına dayanan ve 1993 yılından itibaren kullanılan en eski ve köklü metin işaretleme dillerinden biridir. 1999 yılına kadar sırasıyla 2.0, 3.0 ve 4.0 sürümlerini geçiren HTML, son olarak 1999 yılında 4.0.1 sürümüne ulaşmıştır. Bu sürümler, World Wide Web Consortium (W3C) bünyesindeki HTML Working Group tarafından geliştirilmiştir. Aynı dönemde, HTML Working Group çalışmalarını sonlandırarak XML (Extensible Markup Language) ve bu standart tabanlı bir dil olan XHTML'e yönelmiştir (Danuwijaya 2018). Ancak, yeni bir web standardına duyulan ihtiyaç nedeniyle, 2004 yılında küçük bir grup Web Hypertext Application Working Group (WHATWG) adı altında bir araya gelerek HTML5'in şartnamesini oluşturmuştur (Uslu, 2015). Her bir web sayfası, bir HTML dosyasından oluşmaktadır. HTML, web sayfalarını oluşturmak için kullanılan standart bir biçimlendirme dili olup, web sayfalarının temel iskeletini oluşturmaktadır.

3.2.2 CSS

Web sayfalarının görsel biçimlendirmesi ise CSS (Cascading Style Sheets/Basamaklı Stil Sayfaları) kullanılarak gerçekleştirilmektedir. CSS, bir web sayfasının nasıl görüneceğini belirlemek için kurallar tanımlamaya olanak sağlayan web yardımcı programıdır (Jain Shrivastava ve Pandey Sharma 2024). Örneğin, CSS ile sayfa menülerinin renkleri, yazı tipleri veya arka plan renkleri gibi özellikler kolaylıkla tanımlanabilir. CSS 3.0 sürümüyle birlikte, CSS'in sağladığı olanaklar önemli ölçüde genişlemiştir. CSS 3.0 öncesinde, kare veya daire gibi geometrik şekiller sayfaya eklenirken görseller kullanılması gerekirken, CSS 3.0 ile bu şekiller, yalnızca birkaç satır CSS kodu kullanılarak görsel eklemeye gerek kalmaksızın oluşturulabilmektedir (Özbahçe, 2019).

3.2.3 BOOTSTRAP

Bootstrap, başlangıçta Twitter'da görev yapmakta olan Mark Otto ve Jacob Thornton tarafından geliştirilen açık kaynaklı bir yazılımdır. İlk yayımlandığı dönemde, şirket bünyesindeki mühendislerin ön yüz geliştirme araçlarını standartlaştırma ihtiyacına yanıt vermek amacıyla oluşturulmuştur (Otto ve Thornton 2011). Ağustos 2011'de piyasaya sürülmesinden bu yana büyük bir popülerlik kazanmış olan Bootstrap, başlangıçta tamamen CSS tabanlı bir proje olarak tasarlanmış, zamanla JavaScript eklentileri ve formlar ile düğmelerle entegre çalışan simgeleri içeren daha kapsamlı bir yapıya evrilmiştir (Jhawar ve Katyar 2017). Temel özellikleri arasında, duyarlı (responsive) web tasarımını desteklemesi ve 12 sütunlu, 940 piksel genişliğinde sağlam bir ızgara sistemine sahip olması bulunmaktadır (Spurlock 2013).

Bootstrap'ın dikkat çeken unsurlarından biri, kullanıcıların ihtiyaçlarına uygun bir yapı oluşturmaya olanak tanıyan web tabanlı yapılandırma aracıdır. Bu araç, kullanıcıların belirli CSS ve JavaScript özelliklerini seçerek projelerine dahil etmelerini sağlamaktadır (López-Gorozabel ve Cedeño Palma 2022). Bu özellikler, ön yüz geliştirme süreçlerini hızlandırarak modern bir tasarım ve geliştirme altyapısı üzerine inşa edilmesine olanak tanımaktadır. Bootstrap'ın kullanımı oldukça basittir; temel düzeyde, site köküne birkaç CSS ve JavaScript dosyasının eklenmesiyle kolayca çalışmaya başlanabilir (Spurlock, 2013).

3.2.4 JavaScript

JavaScript, ilk olarak 1995 yılında Brendan Eich tarafından Netscape Communicator tarayıcısı üzerinde çalışmak üzere LiveScript adıyla geliştirilmiştir. Günümüzde JavaScript, neredeyse tüm web uygulamalarında yaygın bir şekilde kullanılmaktadır (Eich ve Wirfs-Brock 2020). 2019 yılının Haziran ayında yapılan bir araştırmaya göre, programcılarının sıklıkla kullandığı GitHub (GitHub 2019) ve Stack Overflow (Stack Overflow 2019) platformlarında JavaScript, en popüler programlama dili olarak gösterilmektedir (Yılmaz 2019).

JavaScript, kullanıcıların tarayıcılarında çalışan “İstemci Tarafı” sürümü ve Netscape sunucularında çalışabilen “Sunucu Tarafı” sürümü olmak üzere iki temel türde

kullanılabilmektedir (Kappagantula 2024). İstemci Tarafı JavaScript, Netscape Navigator (2. sürüm ve sonrası), Microsoft Internet Explorer (3. sürüm ve sonrası) ve Opera tarayıcıları tarafından desteklenmektedir. Tarayıcılar, JavaScript dilini görüntüler, pencereler ve çerçeveler gibi ek nesnelerle genişleterek, yöntemler ve özelliklerle birlikte Document Object Model (DOM) adı verilen bir yapı oluşturur (Lee ve Smith 2008). Bununla birlikte, her tarayıcının DOM uygulamasında kendine özgü küçük farklılıklar bulunmakta olup, tarayıcıların olgunlaşması ve yeni sürümlerin piyasaya sürülmesiyle birlikte DOM'a sürekli olarak yeni özellikler eklenmektedir (Tuncay 2002).

3.2.5 Laravel

Laravel Framework, sunduğu yüksek esneklik, kapsamlı kütüphaneler ve kolayca entegre edilebilen ek paketler nedeniyle web sitesi geliştirme sürecinde tercih edilmektedir (Laaziri ve Benmoussa 2019). Framework, kullanıcı arayüzü ile program mantığını birbirinden ayırarak geliştirme hızını artıran ve arayüz sunumunda esneklik sağlayan Model-View-Controller (MVC) mimari modelini temel almaktadır. MVC, genellikle web tabanlı uygulamalarda yaygın olarak kullanılan bir yazılım tasarım deseni olup, üç temel katman sunmaktadır: model, görünüm ve denetleyici. Geliştiriciler tarafından standart bir web tasarım deseni olarak benimsenen MVC, Java, PHP, Python ve C# gibi birçok programlama dili ile uygulama geliştirme süreçlerinde kullanılmaktadır (Necula 2023). Örneğin, Java'da Spring-MVC Framework, PHP'de CakePHP ve Microsoft tarafından ASP.Net MVC Framework olarak uygulanmaktadır (Sabah 2023).

3.2.6 jQuery

jQuery, 2006 yılında John Resig tarafından "Daha az yaz, daha fazlasını yap" sloganıyla geliştirilmiş hızlı ve özlü bir JavaScript kütüphanesidir. Kısa sürede geniş bir tanınırlık kazanarak web uygulamaları geliştirme süreçlerinde en popüler teknolojilerden biri haline gelen jQuery, sade sözdizimi sayesinde karmaşık işlemleri basitleştirerek kullanıcı dostu bir deneyim sunmaktadır (Tran 2020). Günümüzde jQuery, Twitter ve Facebook gibi sosyal medya platformlarından Yahoo! ve Google gibi arama motorlarına, şık efektlerle donatılmış modern web sitelerine kadar birçok alanda kullanılmaktadır (Desai 2021). Çoğu modern JavaScript web uygulaması, temel tarayıcı API'lerinin üzerine işlevsellik eklemek için kütüphanelerden faydalanmakta ve bu kütüphaneler,

dinamik dil özelliklerini yoğun bir şekilde kullanarak geliştiricilere kullanım kolaylığı sağlamaktadır (Baştuğ, 2019).

3.2.7 MySQL

MySQL, açık kaynaklı, kararlı, verimli ve kullanımı kolay yapısı sayesinde farklı sektörlerde en çok tercih edilen ve güvenilen veritabanı yönetim sistemlerinden biridir (Salunke ve Ouda 2024). Ancak, verilerin karmaşık yapısı ve sorguların çeşitliliği nedeniyle sorgu optimizasyonunda önemli zorluklarla karşılaşmaktadır. Genellikle basit OLTP (Çevrimiçi İşlem İşleme) türü sorgular için tasarlanmış olan MySQL sorgu optimize edicisi, daha karmaşık OLAP (Çevrimiçi Analitik İşleme) türü sorgularla başa çıkarken sınırlamalar göstermektedir (Rahman ve Islam 2024). Sorgu performansını artırmaya yönelik çeşitli çalışmalar, SQL ifadelerinin, indekslerin, şema tasarımının, mimarının veya sunucu tarafının optimize edilmesi gibi yöntemlere odaklanmıştır. Ancak bu yöntemler, performansı artırsa da geliştiricilerden veya veritabanı kullanıcılarından daha fazla çaba gerektirdiği için genellikle pratik ve uygulanabilir bulunmamaktadır (Ramu 2023). Bunun yanı sıra, bazı çalışmalar sorgu performansını artırmak için plan ön bellekleme gibi içsel güncellemeler eklemeyi hedeflemektedir (Sarı, 2024).

3.3 Test Ve Madde Analizi WEB Sitesi

Tasarlanan sitenin internet adresi, amaçladığı hedef göz önünde bulundurularak ve adının akılda kalıcılığının kolay olması açısından “sinavanalizlerim.com” olarak belirlenmiştir.

3.3.1 Web Sitesi Kullanılabilirlik Ölçeği: Yapı, Uygulama ve Değerlendirme Yöntemi

Bu çalışmada geliştirilen sistemin kullanılabilirliğini ölçmek amacıyla kullanıcı görüşlerine dayalı bir değerlendirme aracı kullanılmıştır. Kullanılan ölçme aracı, beşli Likert tipi maddelerden oluşmaktadır. Ölçekte, kullanıcılar her bir ifadeye 1 (Kesinlikle katılmıyorum) ile 5 (Kesinlikle katılıyorum) arasında puan vererek görüş bildirmiştir. Ölçekte toplam 29 madde bulunmaktadır. Bu maddeler; sistemin gezinme kolaylığı, tasarımı, erişilebilirliği ve genel kullanım kolaylığı gibi boyutlara yöneliktir.

Katılımcıların verdiği puanlar her madde için ayrı ayrı değerlendirilmiş, toplam ve ortalama puanlar hesaplanmıştır. Puanların değerlendirilmesinde genel eğilimler, alt boyut düzeyleri ve grup ortalamaları dikkate alınmıştır. Ölçme aracı, kullanıcı deneyimlerini nicel olarak değerlendirmeyi amaçlamakta ve sistemin kullanılabilirlik düzeyine ilişkin bütüncül bir bakış sunmaktadır.

3.3.2 Görev Temelli Kullanılabilirlik Testi: Yöntem ve Uygulama

Bu çalışma kapsamında geliştirilen sinavanalizlerim.com adlı web tabanlı analiz platformunun kullanılabilirliği, görev temelli değerlendirme yöntemiyle sistematik olarak incelenmiştir. Kullanıcıların sistemle etkileşimlerini gözlemleyerek, arayüzde yaşanabilecek olası zorlukları ve sistemin işlevselliğini değerlendirmek amacıyla bir dizi senaryo ve görev tanımı içeren kullanılabilirlik testi uygulanmıştır.

Uygulama süreci, katılımcıların sistemin temel işleyişine aşina olmasını sağlayacak kısa bir bilgilendirme ile başlamıştır. Ardından, her biri platformun farklı işlevlerini temsil eden altı görev tanımı sırasıyla katılımcılara sunulmuştur. Bu görevler arasında sisteme giriş yapma, kategori oluşturma, sınav ekleme ve analizleri görüntüleme gibi işlemler yer almıştır. Katılımcılardan, görevleri yerine getirirken karşılaştıkları zorlukları yüksek sesle ifade etmeleri istenmiş, böylece kullanıcı deneyimi nitel olarak da belgelenmiştir.

Görev tamamlandıktan sonra, her bir görev için 5'li Likert ölçeği kullanılarak zorluk düzeyi değerlendirilmiş ve katılımcılara genel sistem değerlendirmesine yönelik açık uçlu sorular yöneltilmiştir. Bununla birlikte, gözlemci (araştırmacı) tarafından görevlerin tamamlanma süreleri, yardım talepleri ve belirgin kullanıcı davranışları not edilmiştir. Kullanılan ölçekler Indiana Üniversitesinin yapmış olduğu çalışmadan esinlenilmiştir. (Indiana University R641 Student Team 2001). Elde edilen nitel ve nicel veriler, sistemin erişilebilirlik, öğrenilebilirlik ve memnuniyet boyutlarında güçlü ve geliştirilmesi gereken yönlerini belirlemek için analiz edilmiştir.

Bu yöntemle yalnızca işlevsel hatalar değil, aynı zamanda kullanıcı algısına dayalı arayüz geri bildirimleri de sistem geliştirme sürecine katkı sunacak şekilde derlenmiştir. Test

sürecinin sonunda yapılan değerlendirmeler, sistemin öğretim elemanları tarafından genel olarak işlevsel ve kullanıcı dostu bulunduğunu; ancak bazı görevlerde açıklayıcı yardım içeriklerine ihtiyaç duyulduğunu ortaya koymuştur. Bu bağlamda, kullanılabilirlik testi sadece sistemin performansını değil, aynı zamanda kullanıcı merkezli iyileştirme alanlarını da belirleyici bir araç olarak değerlendirilmiştir.

3.3.3 Geliştirme Sürecine Temel Oluşturan Model

Bu çalışmada geliştirilen "Test ve Madde Analizi Web Sistemi", eğitim teknolojisi alanında sıklıkla başvurulmuş Geliştirici Araştırma (Developmental Research) yaklaşımı temel alınarak yapılandırılmıştır. Reeves (1994) tarafından tanımlanan bu yaklaşım, eğitim teknolojisi ürünlerinin tasarım ve geliştirme süreçlerini hem teorik hem de pratik boyutlarıyla ele alarak, ortaya çıkan ürünün geçerliliğini sistematik biçimde değerlendirmeyi amaçlamaktadır (Reeves 1994).

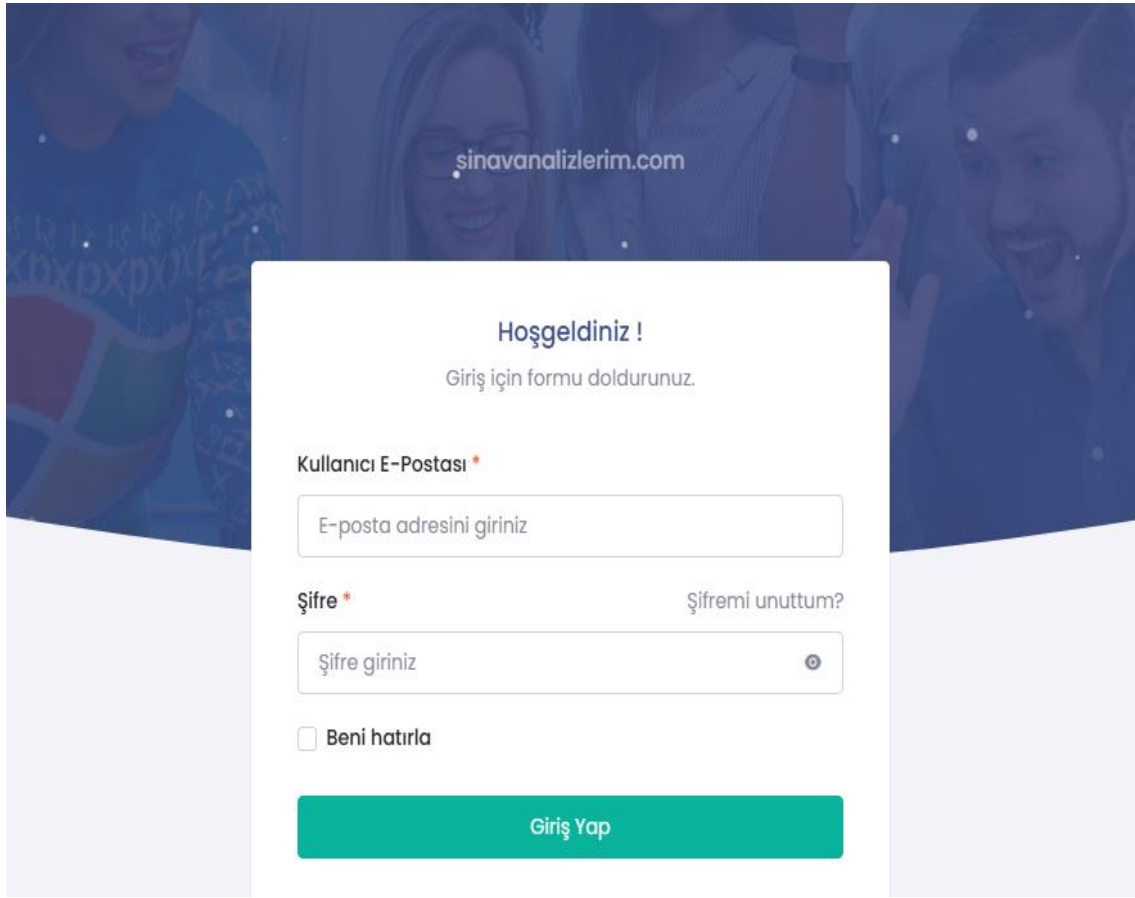
Bu doğrultuda, sistemin geliştirme süreci yalnızca teknik bir tasarım faaliyetinden ibaret olmamış, aynı zamanda kullanıcı ihtiyaçlarının analizi, arayüz tasarımı, sistem prototipleme, kullanıcı testleri ve geri bildirimler doğrultusunda iyileştirme döngülerini de içermiştir. Bu yönüyle geliştirme süreci, yalnızca yazılımsal bir uygulamanın ortaya konması değil, aynı zamanda eğitimsel bağlamda işlevsel bir çözüm geliştirilmesi süreci olarak değerlendirilmiştir.

Söz konusu yaklaşımın benimsenmesi, geliştirilen sistemin yalnızca teknik yeterlilik açısından değil, aynı zamanda kullanıcı deneyimi, etkililik ve kullanılabilirlik bağlamında da değerlendirilmesine olanak tanımıştır. Böylece, hem ürün tasarımına ilişkin ilkelerin belirlenmesi hem de elde edilen bulgular doğrultusunda geliştirici araştırma modeline katkı sağlanması hedeflenmiştir.

3.3.4 Sitenin Ara Yüzü ve Kullanımı

Oluşturulan web sitesine “app.sinavanalizlerim.com” bağlantısı üzerinden erişim sağlanmaktadır. Giriş sayfasında, öğretmenler kendilerine sağlanan kullanıcı e-posta adresi

ve şifre bilgilerini Resim 3.1'de belirtilen alanlara girerek ana sayfaya ulaşabilmektedir. Sitenin birincil kullanıcı kitlesini öğretmenler oluşturduğundan ve öğrencilerin sınav bilgilerine erişimini engellemek amacıyla, siteye genel bir üyelik oluşturma özelliği eklenmemiştir. Siteye erişim sağlamak isteyen öğretmenler, kendi okul bilgileri ve diğer gerekli bilgileri site yöneticilerine ileterek, kullanıcı e-posta adresi ve şifre olarak siteye erişim sağlayabilmektedir.



sinavanalizlerim.com

Hoşgeldiniz !
Giriş için formu doldurunuz.

Kullanıcı E-Postası *

E-posta adresini giriniz

Şifre * [Şifremi unuttum?](#)

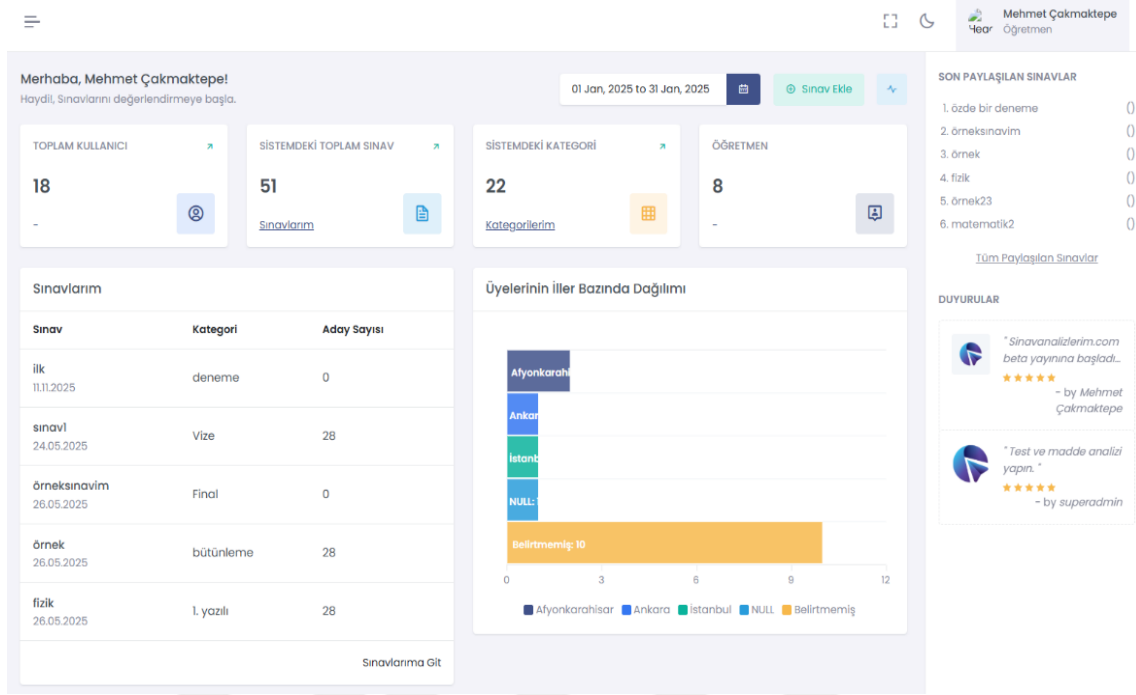
Şifre giriniz

☐ Beni hatırla

Giriş Yap

Resim 3.1 Kullanıcı giriş sayfası.

Ana sayfanın temel amacı, öğretmenlerin diğer öğretmenler tarafından yapılmış test analizlerini görüntüleyebilmesi, hangi şehirde görev yaptıklarını öğrenebilmesi ve genel olarak sitede kaç öğretmenin bulunduğuna ilişkin bilgi edinebilmesidir. Bu oluşturulan paylaşımlar bir öğretmenden baz alınarak başka bir öğretmenin kendi sınavlarını, gerekirse öğretmenle iletişime geçerek geliştirmesine eksikliklerini gidermesine olanak sağlar. Resim 3.2’den de görülebileceği üzere, ana sayfada dört adet bilgi kutusu bulunmaktadır: “Toplam Kullanıcı”, “Sistemdeki Toplam Sınav”, “Sistemdeki Kategori” ve “Öğretmen”.



Resim 3.2 Ana sayfa.

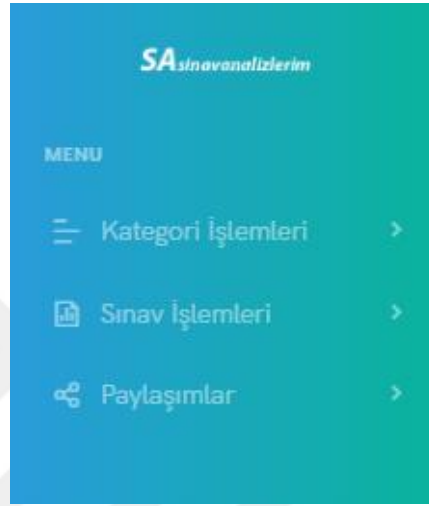
- "Toplam Kullanıcı" kutusu, siteyi kullanan öğretmenlerin toplam sayısını göstermektedir.
- "Sistemdeki Toplam Sınav" kutusu, siteye yüklenen toplam sınav sayısını ifade etmektedir.
- "Sistemdeki Kategori" kutusu, sitede bulunan kategori yani sınav türü sayısını belirtmektedir.
- "Öğretmen" kutusu ise ilgili öğretmenin unvanını (örneğin Prof., Doç., vb.) göstermektedir.

Bu özellikler, öğretmenlerin site üzerindeki analizlere ve genel bilgilere kolayca erişimini sağlamak amacıyla tasarlanmıştır.

Sayfada yer alan Sınavlarım bölümü, siteye yüklenen sınavların tablo formatında görüntülenmesini sağlayan bir alandır. Bu bölümün yanında, siteye kayıtlı üyelerin illere göre dağılımını ve üye sayısını gösteren bir grafik bölümü bulunmaktadır. Bu iki bölümün alt kısmında ise, sistemdeki öğretmenlerin isimlerini listeleyen bir alan yer almaktadır. Bu düzenleme, kullanıcıların sınav ve üye bilgilerine kolay erişimini sağlamak ve sistemdeki öğretmenlere ilişkin bilgileri sunmak amacıyla tasarlanmıştır.

Son olarak sađ tarafta paylaşılan sınavları belirten ve site yöneticisinin aktardığı duyurular kısmı yer almaktadır.

Resim 3.3’de görüldüğü üzere sitede bütün sayfalarının sol tarafında bulunan bir gezinme çubuğu yer almaktadır. Bu gezinme çubuğunda “Kategori İşlemleri”, “Sınav İşlemleri” ve “Paylaşımlar” olmak üzere 3 sekme bulunmaktadır.



Resim 3.3 Gezinme çubuğu.

Resim 3.4’te kategori işlemleri sekmesine tıklandığında “Kategori Ekle” ve “Tüm Kategoriler” adı altında 2 bölüm bulunmaktadır.



Resim 3.4 Kategori işlemleri.

Kategori Ekle bölümü Resim 3.5 ve Resim 3.6’da görüldüğü üzere sınav türlerinin (örneğin, vize, final, deneme vb.) sisteme eklenmesini sağlayan bir işlemdir. Kullanıcı, eklemek istediği sınav türünü belirleyip kaydet butonuna tıkladığında, ilgili sınav türü sistemde Tüm Kategoriler bölümüne aktarılmaktadır.

Kategori Adı

Vazgeç

Kaydet

Resim 3.5 Kategori ekle.

Kategori Adı *

Kategori seçiniz

Kategori seçiniz

Denemeler

Vize

Final

Büt

ödev

Resim 3.6 Kategoriler.

Tüm kategoriler sayfası, eklediğimiz kategorileri görebildiğimiz, isimlerini yeniden düzenleyebildiğimiz ve istediğimiz kategoriyi silebildiğimiz kısımdır. Resim 3.7 bu kısmı göstermektedir.

Kayıtlı Kategoriler

+ Kategori Ekle

Show 10 entries

Search:

SR No.	Kategori	Durum	İşlem
1	deneme	Aktif	
2	Vize	Aktif	
3	Final	Aktif	

Resim 3.7 Tüm kategoriler.

Resim 3.8'e bakıldığında sınav işlemleri sekmesinde "Sınavlar" ve "Sınav Ekle" olmak üzere iki kısım bulunmaktadır.



Resim 3.8 Sınav işlemleri.

Resim 3.9’da bulunan "Sınav Ekle" bölümünde, eklenmek istenen sınavın adı, kategori adı ve sınav tarihi bilgileri girildikten sonra "Kaydet" butonuna tıklanarak sınav, "Sınavlar" bölümüne eklenebilmektedir.

YENİ SINAV

Sınav İşlemleri > Yeni Sınav

Sınav Adı

Sınav adı

Kategori Adı *

Kategori seçiniz

Sınav Tarihi

gg.aa.yyyy

Vazgeç Kaydet

Resim 3.9 Sınav ekle.

Resim 3.10’da gösterilen sınavlar sayfasında, sisteme eklenen tüm sınavlar görüntülenebilmektedir. Bu sayfada sınavın kategorisini, sınav tarihini, sınav ile ilgili yapacağımız işlemler ve sınavı paylaş seçenekleri yer almaktadır.

SINAVLARIM

Sınav İşlemleri > Sınavlarım

Sınav Listesi

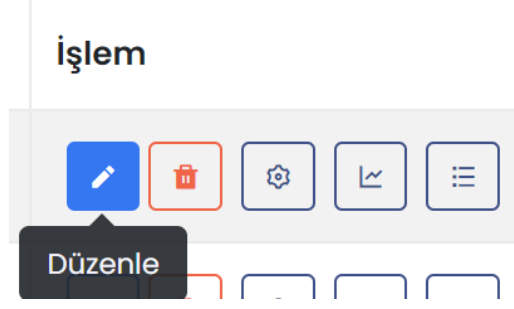
+ Sınav Ekle

Show 10 entries Search:

SR No.	Sınav Adı	Kategori	Sınav Tarihi	İşlem	Paylaş
1	ilk	deneme	11.11.2025	   	 paylaşılmadı
2	sınavı	Vize	24.05.2025	   	 paylaşılmadı
3	örneksınavım	Final	26.05.2025	   	 paylaşıldı

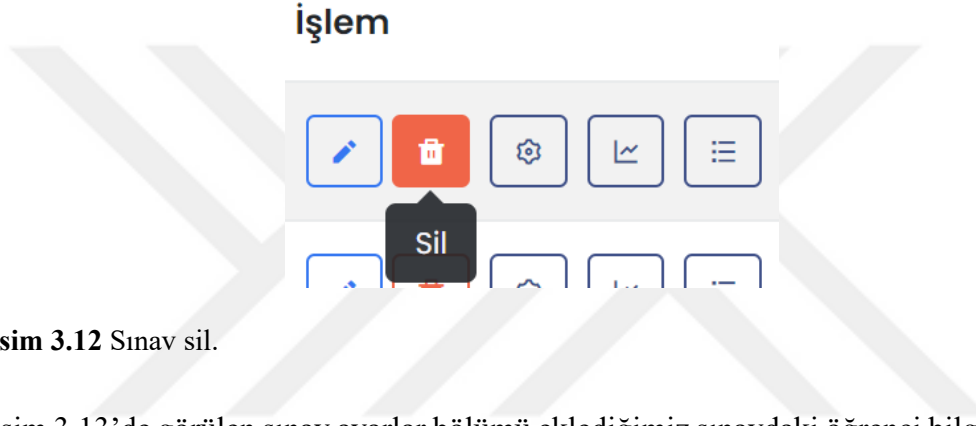
Resim 3.10 Sınavlar.

Resim 3.11’de görülen düzenle butonu eklediğimiz sınavın adını tarihini ve kategorisini yeniden düzenleyebildiğimiz kısımdır.



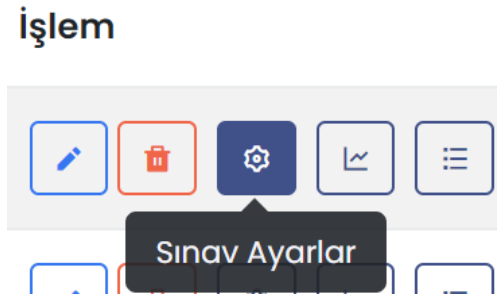
Resim 3.11 Sınav düzenle.

Sil butonu eklenen sınavı silebildiğimiz araçtır. Resim 3.12’de görülmektedir.



Resim 3.12 Sınav sil.

Resim 3.13’de görülen sınav ayarlar bölümü eklediğimiz sınavdaki öğrenci bilgilerini içe aktarmamıza ve cevap anahtarlarını kitapçık türleriyle birlikte tanımlamamıza olanak sağlayan bölümdür.



Resim 3.13 Sınav ayarlar.

Resim 3.14'te görüldüğü üzere, web sayfasında sınav adaylarına ait bilgilerin ve cevap anahtarlarının yönetilmesine yönelik bir arayüz sunulmaktadır. Bu arayüz, optik makinelerden alınan sınava verilen cevapları içeren metin dosyasının sisteme yüklenmesini ve cevap anahtarlarının tanımlanmasını sağlamaktadır.

Resim 3.14 Sınav ayarlar sayfası ilk kısım.

Metin dosyası, "Dosya Seç" düğmesi aracılığıyla kullanıcı bilgisayarından seçilerek sisteme aktarılabilir. Dosya seçiminin ardından, "Yükle" seçeneği kullanılarak sınava ilişkin bilgiler web uygulamasına kaydedilir. Eğer sınavda farklı kitapçık türleri kullanıldıysa, "Sınavda Kitapçık Türü Var" seçeneği işaretlenir. Bu durumda, sağ tarafta bulunan kitapçık türü ekleme alanı etkinleşir.

Kitapçık türü, öğretmenin tercihlerine göre özelleştirilebilir. "Kitapçık Türünü Ben Belirleyeyim" seçeneği, öğretmenlerin A ve B harfleri dışında farklı kitapçık türleri belirleyebilmesine olanak tanımak amacıyla sunulmuştur. Bu seçenek tercih edilmediğinde, sistemde önceden tanımlı olan kitapçık türlerinden biri seçilerek kullanılabilir. Kitapçık türüne uygun cevap anahtarı, ilgili alana manuel olarak klavye aracılığıyla girilebileceği gibi kopyala-yapıştır yöntemiyle de aktarılabilir.

Cevap anahtarının tanımlanması tamamlandığında, cevap anahtarı kutusunun yanında yer alan "Ekle" butonuna tıklanarak veri sisteme kaydedilir. Bu süreç, sınav yönetimi sürecinin etkin ve esnek bir şekilde yürütülmesini sağlamak amacıyla tasarlanmıştır.

Sınav bilgileri sisteme aktarıldıktan sonra, Resim 3.15'te görülen "Aday Cevapları" ve "Kalıp Ayarlama" bölümleri aracılığıyla sınav tanımlama işlemleri tamamlanmaktadır. "Yüklenen Aday Cevapları" bölümü, sınava katılan öğrencilerin yanıtlarını içeren metin dosyasının sisteme yüklenmesiyle birlikte otomatik olarak güncellenir. Bu bölümde, öğrenci numaraları, kitapçık türleri ve öğrenci cevapları detaylı bir şekilde görüntülenir.

Sınav analizlerini görebildiğimiz bu sayfada Resim 3.17’de görülebileceği üzere “Test Sonuç”, “Test Analizi” ve “Test Frekans Tablosu” olarak üç kısım bulunmaktadır.

SINAV ANALİZİ Sınav İşlemleri > Sınav Analizi

Sonuçları PDF olarak indir

Test Sonuç						Test Analizi		Test Frekans Tablosu	
Aday Kimlik	Kitapçık	Doğru	Yanlış	Boş	Detay	En Yüksek Puan	35	Puanlar	Frekanslar
214207092	A	35	0	0	Detay	En Düşük Puan	7	35	1
224207083	A	23	10	2	Detay	Ortalama	15,464	23	1
194208039	A	22	13	0	Detay	Ranj	28	22	1
214207025	A	20	14	1	Detay	Tepe Değer (Mod)	10,	20	3
204207094	A	20	13	2	Detay	Ortanca (Medyan)	15,500	19	4
								18	2

Resim 3.17 Sınav analiz sayfası.

Test sonuçları kısmında, sınava katılan adayların kitapçık türleri ile birlikte sınavda verdikleri doğru, yanlış ve boş yanıtlar listelenmektedir. Ayrıca, "Detay" butonuna tıklandığında, Resim 3.18’de gösterildiği gibi, cevap anahtarıyla kıyaslama yapılarak detaylı koşullu biçimlendirme uygulanmaktadır. Bu biçimlendirme kapsamında, yanlış cevaplar kırmızı renkte "-" işaretiyle, doğru cevaplar ise yeşil renkte "+" işaretiyle gösterilmektedir.

Detaylar

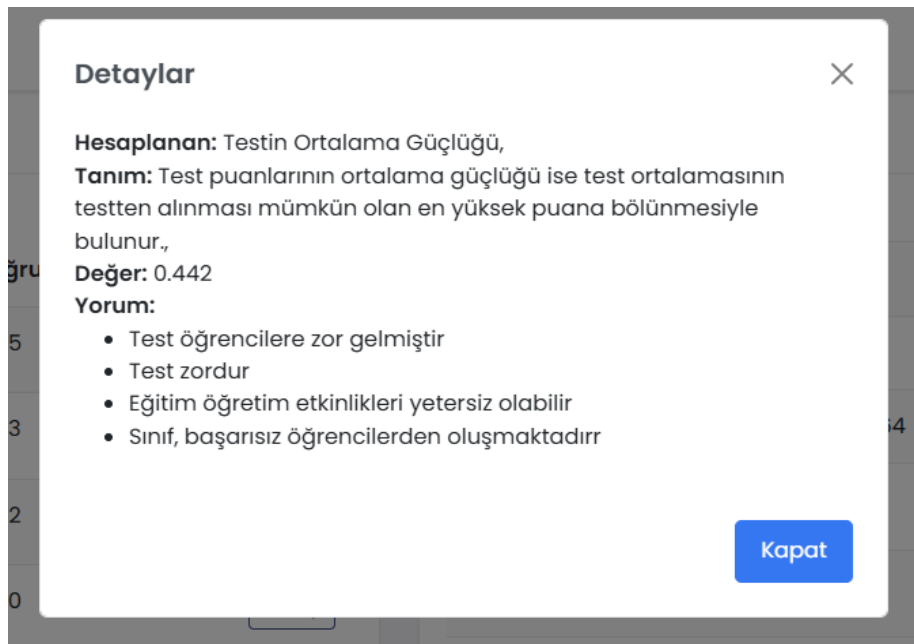
Soru Numarası	Aday Cevap	Doğru Cevap	Durum
1	B	B	+
2		N	-
3	A	D	-
4	D	D	+
5	B	B	+
6	C	C	+
7	C	D	-
8	B	A	-
9	A	A	+
10	B	B	+

Resim 3.18 Öğrenci cevapları detaylar.

Test analizi bölümünde, ölçme ve değerlendirme literatüründen referans alınarak yapılan analiz sonucunda, testin aşağıdaki istatistiksel değerleri hesaplanmaktadır:

- a. En Yüksek Puan
- b. En Düşük Puan
- c. Ortalama
- d. Ranj
- e. Tepe Değer (Mod)
- f. Ortanca (Medyan)
- g. Testin Ortalama Güçlüğü
- h. KR-20 Güvenirlilik Katsayısı
- i. Standart Sapma
- j. Varyans
- k. Bağlı Değişim Katsayısı
- l. Çarpıklık Katsayısı

Her bir değer yanında, bu değer tanımını ve analiz sonucuna yönelik yorumları incelemek için “i” simgeli bilgi butonları yer almaktadır. Resim 3.19’da bu açıklamalara ilişkin örnek bir gösterim sunulmaktadır.



Resim 3.19 Değerlerin detaylar sekmesi.

Testin frekans tablosunda, öğrencilerin aldıkları puanların, büyükten küçüğe doğru sıralanmış şekilde dağılımı ve her bir puan aralığındaki öğrenci sayıları yer almaktadır. Bu tablo, sınav sonuçlarının genel dağılımını görselleştirerek analiz yapmayı kolaylaştırmaktadır.

Resim 3.20’de “Madde Analizi” seçeneğine tıklandığında, öğrencilerin cevaplandığı her bir soruya ilişkin detaylı madde istatistiklerini gösteren bir sayfa açılmaktadır. Bu sayfa, her bir sorunun tek tek analiz edilmesini sağlayarak, soruların istatistiksel özelliklerini değerlendirme imkânı sunmaktadır.



Resim 3.20 Madde analizi.

Resim 3.21 ve 3.22’de gösterildiği üzere, “Madde Analizi” sayfası şu bölümlerden oluşmaktadır:

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 İstatistik Hesaplama Sonuçları

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

Madde İstatistik Bilgileri

Hesaplanan: Madde Ayırtıcılık Gücü İndeksi (Madde Geçerliliği),
Tanım: Bilen ile bilmeyeni veya başarılı öğrenci ile başarısızları ayırt etme gücüdür.,
Değer: 0.333
Yorum: Oldukça iyi bir madde.

Resim 3.21 Madde analiz sayfası.

Madde Cevap ve Çeldiriciler	
Madde Şık	Cevap Toplamı
C	12
D	2

Resim 3.22 Madde cevap ve çeldiriciler.

1. “Madde Seçimi”: Bu bölümde, eklenen sınava ait soruların kitapçık türü ve soru maddesi seçimi yapılabilmektedir.
2. “Madde İstatistik Bilgileri”: Seçilen soruya ilişkin kaç kişinin doğru, kaç kişinin yanlış cevap verdiği bilgisi, üst grup ve alt grup değerleriyle birlikte bu alanda sunulmaktadır.
3. “Madde Cevap ve Çeldiriciler”: "Madde İstatistik Bilgileri" bölümünün hemen altında yer alan bu kısım, seçilen soruya verilen yanıtların dağılımını detaylı olarak göstermektedir. Doğru yanıtı yönelen öğrenci sayısı ile diğer şıkları seçen öğrenci sayıları karşılaştırılabilir. Bu sayede, doğru cevaba ulaşamayan öğrencilerin yoğunlukla hangi şıklara yöneldiği belirlenerek, ilgili şıkların çeldirici işlevi değerlendirilebilmektedir.
4. “Madde İstatistik Hesaplama Sonuçları”: Bu bölümde, ölçme ve değerlendirme literatüründen alınan yöntemler kullanılarak aşağıdaki istatistiksel değerler hesaplanmaktadır:

- a. Madde Güçlük İndeksi
- b. Madde Ayırt Edicilik İndeksi
- c. Madde Varyans
- d. Madde Standart Sapma
- e. Madde Güvenirlik İndeksi
- f. Nokta Çift Serili Korelasyon
- g. Çift Serili Korelasyon

5. “Detaylı İstatistik Bilgileri”: Bu bölüm, yukarıda belirtilen değerlerin tanımlarını ve yorumlarını içermektedir. İlgili değerlerin yanında yer alan açıklama butonuna

basıldığında, her bir değerin anlamı ve soru bazında alınan sonuçlara yönelik yorumlar görüntülenebilmektedir.

Bu yapı, madde bazlı analizlerin detaylı ve kapsamlı bir şekilde yapılmasını sağlayarak sınav sorularının ölçme ve değerlendirme açısından etkinliğini değerlendirme imkânı sunmaktadır.

Resim 3.23’de görülen "Paylaşımlar" sekmesi, tek bir alt seçenek olan "Paylaşılan Değerlendirmeler" bölümünden oluşmaktadır. Resim 10’da görülen "Sınavlar" bölümünün son kısmında yer alan "Paylaşım" seçeneği, sınavın diğer eğitimlerle paylaşılmasını sağlamaktadır. Bu seçenek aracılığıyla paylaşılan sınavlar, ana sayfada görüntülenebileceği gibi, Resim 3.23’de gösterilen "Paylaşılan Değerlendirmeler" kısmında da detaylı olarak incelenebilmektedir.



Resim 3.23 Paylaşımlar.

Siteden test analizi ve madde analizi sonuçlarını daha detaylı görmek ve paylaşımını kolaylaştırmak amacıyla hem test analizi hem de madde analizi sayfalarına Resim 3.24’te görüldüğü gibi “Analiz Sonuçlarını PDF Olarak İndir” butonu konulmuştur.

 Analiz sonuçlarını PDF olarak indir

Resim 3.24 PDF indirme.

"Analiz Sonuçlarını PDF Olarak İndir" butonuna tıklandığında, Resim 3.25, Resim 3.26 ve Resim 3.27’de gösterildiği gibi analiz sonuçları PDF formatında görüntülenebilmekte ve indirilebilmektedir. Bu özellik, madde analizi kısmında soruları tek tek incelemek yerine, tüm soruların analizlerini toplu ve detaylı bir şekilde görmek isteyen eğitimler için faydalı bir seçenek sunmaktadır.

Madde Analiz Raporu

Kitapçık: A

S. No	DS/YS	Verilen Cevaplar	MGİ (p _j)	MGİ(G) (p _j)	AEİ (r _{jk})	Var.(s _j ²)	St.S.(s _j)	Güv. İn.(r _j)	NÇSK(r _{pb})	ÇSK(r _b)
1	6/8	(8) B*(6)	0.429	0.333	0.667	0.245	0.495	0.330	0.413	0.000
2	2/12	(11) I(1) N*(2)	0.143	0.167	0.333	0.122	0.350	0.117	0.634	0.000
3	1/13	A(11) B(1) C(1) D*(1)	0.071	0.167	0.333	0.066	0.258	0.086	0.878	0.000
4	13/1	B(1) D*(13)	0.929	1.000	0.000 **	0.066	0.258	0.000	0.017	0.000
5	11/3	B*(11) C(3)	0.786	0.833	0.333	0.168	0.410	0.137	0.107	0.000
6	13/1	A(1) C*(13)	0.929	1.000	0.000 **	0.066	0.258	0.000	0.136	0.000

Resim 3.25 Madde analizi raporu.

Test Sonuç

Aday Kimlik	Kitapçık	Doğru	Yanlış	Boş
214207092	A	35	0	0
224207083	A	23	10	2
194208039	A	22	13	0
214207025	A	20	14	1
204207094	A	20	13	2
214207005	A	20	15	0

Resim 3.26 Test analizi ilk sayfa.

Test Analizi Raporu bölümünde, ek olarak Resim 3.27’da gösterilen "Test Genel Yorumu" kısmı bulunmaktadır. Bu yorumlar, KR-20, Bağıl Değişim Katsayısı, Çarpıklık Katsayısı ve Testin Ortalama Güçlük Değeri gibi istatistiksel verilere dayalı olarak, ölçme ve değerlendirme literatürüne uygun şekilde yapılan yorumları ve elde edilen sonuçları içermektedir.

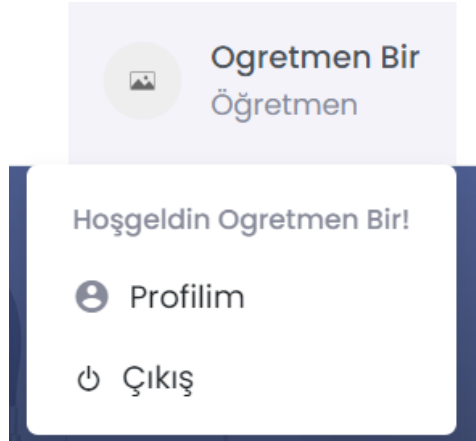
Test Genel Yorumu

KR-20 değeri olan 0.842 ye göre testimiz; Çok yüksek güvenilirlik. Testin maddeleri arasındaki tutarlılık çok güçlü. Maddelerin güvenilirliği son derece yüksektir.
Bağıl Değişim Katsayısı değeri olan 38.275 ye göre testimiz; puanların heterojen, birbirinden farklı olduğunu, puanlar arasındaki farklılığın fazla olduğunu gösterir ki bu durumda puanlar normale göre daha basık bir dağılıma sahip olacaktır.
Çarpıklık Katsayısı değeri olan 1.234 ye göre testimiz; Dağılım sağa(artı yöne) çarpıktır. Testin Güçlük Düzeyi: Çok Zor
Testin Ortalama Güçlük değeri olan 0.442 ye göre testimiz;

- Test öğrencilere zor gelmiştir
- Test zordur
- Eğitim öğretim etkinlikleri yetersiz olabilir
- Sınıf, başarısız öğrencilerden oluşmaktadır

Resim 3.27 Test analizi ikinci sayfa.

Resim 3.28'de sağ üst köşede bulunan öğretmen profil resmine tıklandığında, Resim 3.27'de gösterildiği gibi "Profilim" ve "Çıkış" seçenekleri görüntülenmektedir.



Resim 3.28 Profil aracı.

"Profilim": Bu bölümde, öğretmenin profil bilgilerine erişilebilmekte ve şifre değiştirme gibi kişisel ayarlar yapılabilmektedir.

"Çıkış": Bu buton kullanılarak, kullanıcı profilinden güvenli bir şekilde çıkış yapılabilmektedir.

Resim 3.29'da görülen Profil Sayfası, öğretmenlerin profil fotoğrafını ve kullanıcı bilgilerini düzenleyip güncelleyebilecekleri bir alan sunmaktadır. Ayrıca, "Kullanıcı Detayı" sekmesinin yanında yer alan ve Resim 3.30'da gösterilen "Şifre Değiştir" sekmesi üzerinden kullanıcılar şifrelerini değiştirebilir ve yapılan değişiklikleri kaydedebilir.

Kullanıcı Detayı Şifre Değiştir

Kaydet

user-profile-im

Mehmet Çakmaktepe
ogretmen12@sinavanalizlerim.com

Kullanıcı Bilgileri

Şirket NULL
Adı :
Cep Telefon :
E_posta ogretmen12@sinavanalizle
Yetkili Kişinin Mehmet Çakmaktepe

Ad* Mehmet Çakmaktepe
Telefon * NULL
E-posta * ogretmen12@sinavanalizlerim.com
Şirket Adı * NULL
Şehir * NULL
Adres * NULL

Güncelle İptal

Resim 3.29 Profil sayfası.

Kullanıcı Detayı Şifre Değiştir

Eski Şifre* Güncel şifrenizi giriniz
Yeni Şifre* Yeni şifreyi girin
Şifreyi Onayla* Şifreyi Onayla

[Parolanızı Mı Unuttunuz ?](#)

Şifre Değiştir

Resim 3.30 Şifre değiştir sekmesi.

4. BULGULAR

Bu bölümde, önceki bölümlerde sunulan teorik çerçeve doğrultusunda, çoktan seçmeli testlerin güvenilirlik analizine yönelik geliştirilen web tabanlı sistemin değerlendirilmesine odaklanılmıştır. Uygulama aşamasında, farklı eğitim düzeylerindeki akademik katılımcılarla gerçekleştirilen anket sonuçları analiz edilmiş; sistemin kullanılabilirliği, kullanım kolaylığı, işlevselliği ve genel memnuniyet düzeyi başlıkları altında değerlendirilmiştir. Elde edilen bulgular, mevcut literatürle karşılaştırılarak yorumlanmış ve geliştirilen sistemin ölçme-değerlendirme süreçlerine katkı potansiyeli çeşitli açılardan tartışılmıştır. Sonuç bölümünde ise elde edilen veriler doğrultusunda sistemin geliştirilmesine yönelik önerilere yer verilmiştir.

4.1 Araştırmanın Amacı ve Önemi

Bu araştırma, çoktan seçmeli testlerin güvenilirlik analizine yönelik olarak tarafımızdan geliştirilen “sinavanalizlerim.com” adlı web tabanlı platformun kullanılabilirliğini incelemeyi amaçlamaktadır. Platform, test güvenilirlik katsayısı ile madde ayırt ediciliği ve madde gücü gibi temel ölçme araçlarını dijital ortamda sunarak, eğitimde ölçme ve değerlendirme süreçlerine destek olmayı hedeflemektedir. Çalışma kapsamında, sistemin erişim hızı, kullanıcı arayüzünün anlaşılabilirliği, tasarım bütünlüğü ve gezinme kolaylığı gibi kullanılabilirlik kriterleri analiz edilmiştir. Nitekim, Türkçe dilinde hizmet sunan ve kullanıcı dostu bir arayüzle bütünleşik analiz imkânı sağlayan web tabanlı sistemlerin sayıca yetersiz olması, bu çalışmanın önemini artırmakta ve alandaki önemli bir boşluğu doldurduğunu göstermektedir.

4.2 Araştırmanın Hipotezleri

Çalışmanın literatür bilgisinden yararlanılarak aşağıdaki hipotezler geliştirilmiştir.

H₁: Geliştirilen web tabanlı sistem, çoktan seçmeli testlerin değerlendirme sürecini istatistiksel analiz açısından kolaylaştırmakta ve ölçme-değerlendirme uygulamalarının verimliliğini artırmaktadır.

H₂: Geliştirilen sistem, kullanılabilirlik açısından kullanıcılar tarafından erişilebilirlik, anlaşılabilirlik, arayüz tasarımı ve işlevsellik boyutlarında yüksek düzeyde olumlu değerlendirilmekte ve kullanıcı beklentilerini karşılamaktadır.

4.3 Araştırmanın Evreni/Örnekleme ve Kısıtları

Bu araştırmanın evrenini, Afyon Kocatepe Üniversitesi bünyesinde görev yapan akademik personel grubu oluşturmaktadır. Ancak, zaman ve maliyet kısıtları ile birlikte erişim güçlükleri ve kontrol olanaklarının sınırlı olması nedeniyle tüm evrenin araştırma kapsamına alınması uygun görülmemiştir. Literatürde, genel araştırmalarda örneklem büyüklüğünün 30 ile 500 kişi arasında olmasının yeterli kabul edildiği, alt gruplar için ise her bir grupta en az 30 birey bulunmasının temsiliyet açısından yeterli olacağı ifade edilmektedir (Uğan ve Akbolat 2023). Bu bağlamda, çalışmanın örnekleme, belirtilen ölçütler doğrultusunda belirlenmiş ve 50 akademik personelden oluşacak şekilde sınırlandırılmıştır.

Çizelge 4.1 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	8,0	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

Çizelge 4.1'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.

4.4 Araştırmanın Veri Toplama Yöntemi

Bu çalışmada veri toplama aracı olarak anket yöntemi tercih edilmiştir. Anket uygulamaları, katılımcıların gönüllü katılımına dayalı olarak yüz yüze görüşme yöntemiyle üniversite ortamında gerçekleştirilmiştir. Anket formu, katılımcıların sosyo-demografik bilgilerini (cinsiyet, yaş, akademik unvan, mesleki deneyim süresi) belirlemeye yönelik soruların yanı sıra, geliştirilen web tabanlı platformun kullanılabilirliğini değerlendirmeye yönelik maddeleri içermektedir. Web sitesinin kullanılabilirlik düzeyinin ölçülmesinde, geçerlik ve güvenirlik çalışmaları Meryem ÇAKMAK (2019) tarafından yapılmış olan “Web Sitesi Kullanılabilirlik Ölçeği” kullanılmıştır. Ölçekten elde edilen toplam puanlar 123 üzerinden değerlendirilmiş; 93–123 aralığı ‘yüksek kullanılabilirlik’, 62–92 aralığı ‘orta kullanılabilirlik’ ve 31–61 aralığı ‘düşük kullanılabilirlik’ şeklinde sınıflandırılmıştır. Analizlerde, hem faktör düzeyinde madde ortalamalarına hem de ölçeğin genel düzeyinde toplam puan ortalamasına yer verilmiştir. Faktör bazlı değerlendirmelerde 5’li Likert ortalamaları, genel düzeyde ise ölçek toplamı esas alınmıştır (Çakmak vd. 2019).

Yukarıda bahsi geçen anket çalışmasına ek olarak, web tabanlı analiz sisteminin kullanılabilirliğini değerlendirmek amacıyla görev temelli bir kullanılabilirlik değerlendirme formu kullanılmıştır. Form, kullanıcıların sisteme dair belirli görevleri yerine getirerek deneyimlerini aktarmalarına olanak sağlayacak şekilde yapılandırılmıştır. Değerlendirme süreci boyunca, görevlerin tamamlanma durumu, karşılaşılan güçlükler ve kullanıcı yorumları gözlemlenerek kaydedilmiştir. Söz konusu form, kullanıcı davranışlarını sistemli biçimde izlemeye olanak tanıyan Morae yazılımı ile yürütülen görev odaklı test uygulamalarından esinlenilerek tasarlanmıştır. Bu bağlamda formun yapısında, Olabiyi ve Akinwale (2010) tarafından önerilen kullanıcı görev senaryoları ve gözlem temelli değerlendirme yöntemi temel alınmıştır. Bu yaklaşım sayesinde, sadece teknik başarı düzeyi değil, aynı zamanda kullanıcı algılarını içeren çok boyutlu bir ölçme olanağı sağlanmıştır (Olabiyi ve Akinwale 2010).

4.5 Araştırmanın Veri Analiz Yöntemleri ve Güvenilirliği

Araştırmada elde edilen veriler, SPSS istatistik yazılımı aracılığıyla analiz edilmiştir. Katılımcıların demografik ve bireysel özelliklerine ilişkin bulgular, frekans ve yüzde

dağılımlarıyla betimlenmiştir. Her ne kadar çalışmada kullanılan ölçeklerin geçerlik ve güvenilirlik analizleri daha önceki araştırmalarda yapılmış olsa da, bu çalışmada da ilgili ölçeklerin yapısal geçerliliğini test etmek amacıyla faktör analizi uygulanmıştır. Bu analiz, ölçeklerin özgün faktör yapılarıyla ne derece örtüştüğünü değerlendirmek için gerçekleştirilmiş olup, elde edilen sonuçlara çalışmanın ilgili bölümlerinde ayrıntılı şekilde yer verilmiştir. Ayrıca, ölçeklerin iç tutarlılık düzeylerini belirlemek üzere Cronbach's Alpha güvenilirlik katsayısından yararlanılmıştır. Akademisyenlerin web tabanlı platforma ilişkin değerlendirmeleri ise frekans, yüzde, aritmetik ortalama ve standart sapma değerleri hesaplanarak betimsel istatistikler aracılığıyla sunulmuştur.

4.6 Web Sitesi Kullanılabilirlik Ölçeği'nden Elde Edilen Bulgular

Bu bölümde, kullanılabilirlik ölçeği uygulamasından elde edilen sayısal bulgulara yer verilmiştir. Her bir maddeye verilen ortalama puanlar hesaplanmış, alt boyutlara göre dağılımlar incelenmiş ve genel kullanılabilirlik düzeyi yorumlanmıştır. Ortalama puanlar genel olarak orta-yüksek düzeyde seyretmektedir. Özellikle kullanım kolaylığı ve tasarım estetiği gibi boyutlarda daha yüksek puanlar gözlemlenmiştir. Katılımcıların madde bazlı yanıtlarının analizinde ise, bazı maddelerde daha düşük puanlama eğilimleri dikkati çekmiştir. Bu bulgular, sistemin genel olarak işlevsel bulunduğunu, ancak kullanıcı arayüzü ve yönlendirme unsurlarında iyileştirme yapılabileceğini göstermektedir.

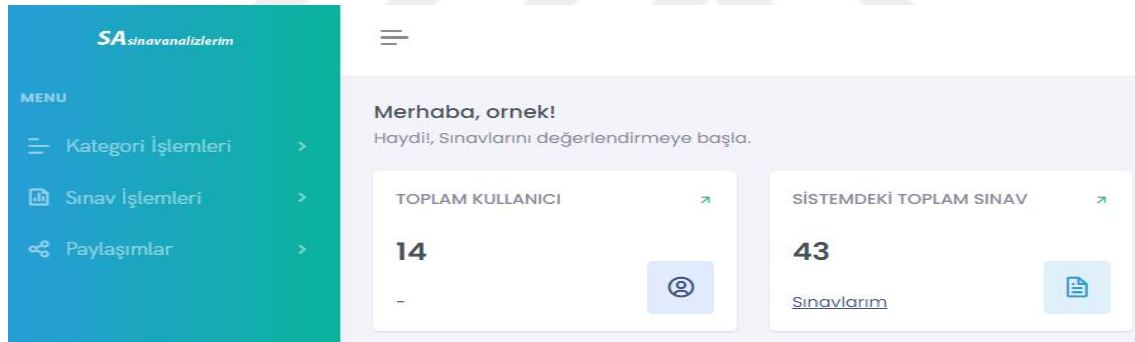
Çizelge 4.2 Katılımcıların 'Gezinme Kolaylığı' Faktörünü Oluştan Maddelere İlişkin Algılarının Aritmetik Ortalaması

	N	X
1. Madde	50	4,36
2. Madde	50	4,42
3. Madde	50	4,42
4. Madde	50	4,50
5. Madde	50	4,52
6. Madde	50	4,48
7. Madde	50	4,32
8. Madde	50	4,30
9. Madde	50	4,36

Çizelge 4.2 incelendiğinde, “Gezinme Kolaylığı” faktörüne ilişkin genel ortalama puanın 4,33 olduğu görülmektedir. Bu değer, kullanıcıların bu boyuttaki deneyimlerini genel olarak olumlu olarak değerlendirdiğine işaret etmektedir.

Faktör içerisinde yer alan bireysel maddeler incelendiğinde, en yüksek puanın 4.52 ile “Menüdeki başlıklar ile sayfa içerikleri tutarlı” ifadesine ait olduğu görülmektedir. Bu maddeyi, {puanlar[3]} puanla “Sitede rahatlıkla geziyorum.” ifadesi izlemektedir. Bu bulgular, kullanıcıların sistemde gezinme sürecini oldukça sezgisel ve uyumlu bulduklarını göstermektedir.

Elde edilen veriler, web sitesinin gezinme yapısının kullanıcı beklentilerini büyük ölçüde karşıladığını ve bu durumun kullanıcı memnuniyetine olumlu katkı sağladığını göstermektedir. Bu bağlamda Resim 4.1’de örneklendirilen yapısal düzenin, kullanıcı deneyimi açısından etkili bir rol oynadığı söylenebilir.



Resim 4.1 Başlıklar ve sayfa gezinme kolaylığı.

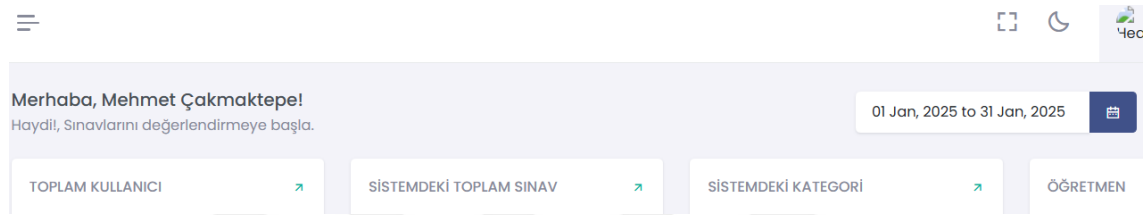
8. maddeye bakıldığında “İlişkili bilgiler gruplandırılmış” ifadesi altında çizelgeye göre en düşük puan olan 4.30 olarak “Katılıyorum” düzeyinde sonuçlandırılmıştır.

Çizelge 4.3 incelendiğinde, web sayfasının "Tasarım" bölümünü oluşturan maddeler değerlendirilmiştir. Bu değerlendirmeye göre en yüksek puan, 4,56 ile 11. maddeye ait olduğu tespit edilmiştir. “Kullanılan yazı tipleri tutarlı” ifadesi kapsamında alınan sonuçlar, kullanıcıların web sayfasındaki yazı tiplerini tutarlı bulduklarını göstermektedir. Bu durum, Resim 4.2’de örnek olarak sunulmuştur.

İncelenen Çizelge 4.3 sonuçlarında 12. Maddenin ise “Yazı tipi ve büyüklüğü okumayı kolaylaştırıyor” ifadesi altında en az puan olan 4,28 değerini aldığı belirlenmiştir.

Çizelge 4.3 Katılımcıların tasarım faktörünü oluşturan maddelere ilişkin algılarının aritmetik ortalaması.

		N	X
10.	Madde	50	4,48
11.	Madde	50	4,56
12.	Madde	50	4,28
13.	Madde	50	4,32
14.	Madde	50	4,34
15.	Madde	50	4,42
16.	Madde	50	4,34



Resim 4.2 Yazı tipleri.

Çizelge 4.4 incelendiğinde en yüksek değer 4,54 puan ile 17. maddenin aldığı saptanmıştır. “Site hızlı yükleniyor” ifadesiyle “Kesinlikle katılıyorum” değeri alınan sonuç ile kullanıcıların web sitesinin kullanımının oldukça hızlı bulunduğu saptanmıştır.

Yapılan incelemelerde Çizelge 4.4 de en düşük değerin 4,14 ile 19. Madde olduğu belirlenmiştir. Bu sonuca göre “Site farklı tarayıcılarda (Brave, Opera, vb.) aynı şekilde ve sorunsuz çalışıyor” ifadesinden yola çıkılarak “Kesinlikle katılıyorum” ifadesinin çoğunlukta bir değerde olduğu görülmektedir.

Çizelge 4.4 Katılımcıların erişim kolaylığı faktörünü oluşturan maddelere ilişkin algılarının aritmetik ortalaması.

		N	X
17.	Madde	50	4,54
18.	Madde	50	4,44
19.	Madde	50	4,14
20.	Madde	50	4,20

Çizelge 4.5 incelendiğinde, alınan değerlere göre en yüksek puanın 4,54 ile 29. maddeye ait olduğu belirlenmiştir. “Madde analizi kısmı soruları detaylı bir şekilde incelemeye

yardımcı oluyor” ifadesine dayanarak kullanıcıların bu maddeye "Kesinlikle katılıyorum" yanıtını verdikleri görülmüştür. Bu bulgu, kullanıcıların web sayfasındaki madde analizi bölümünü anlaşılır ve detaylı bulduklarını, ayrıca sınav sorularını etkili bir şekilde inceleyip sonuçlandırabildiklerini ortaya koymaktadır. Detaylı madde analizi sayfasını Resim 4.3’de görebilmekteyiz.

Çizelge 4.5'e göre, 23. maddenin en düşük puan olan 3,30 değerini aldığı görülmüştür. Bu sonuç, “İçeriği henüz tamamlanmamış sayfa veya sayfalar var” ifadesine kullanıcıların genel olarak "Katılmıyorum" yanıtını verdiklerini göstermektedir. Bu durum, kullanıcıların web sitesindeki sayfaların içerik açısından büyük ölçüde tamamlanmış olduğunu düşündüklerini ortaya koymaktadır.

Çizelge 4.5 Katılımcıların kullanım kolaylığı faktörünü oluşturan maddelere ilişkin algılarının aritmetik ortalaması.

		N	X
21.	Madde	50	3,88
22.	Madde	50	4,06
23.	Madde	50	3,30
24.	Madde	50	3,76
25.	Madde	50	4,34
26.	Madde	50	4,48
27.	Madde	50	4,48
28.	Madde	50	4,36
29.	Madde	50	4,54

Madde İstatistik Hesaplama Sonuçları

İstatistik	Değer
Madde Güçlük İndeksi (p_i)	0.214 ⓘ
Madde Ayırt Edicilik İndeksi (r_{ix})	0.333 ⓘ
Madde Varyans (s_i^2)	0.168 ⓘ
Madde Standart Sapma (s_i)	0.410 ⓘ
Madde Güvenirlilik İndeksi	0.137 ⓘ
Nokta Çift Serili Korelasyon (r_{pb})	0.492 ⓘ
Çift Serili Korelasyon (r_b)	0.000 ⓘ

Resim 4.3 Madde analizi sonuçları.

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. Çizelge 4.6'da sunulan verilere göre, “sinavanalizlerim.com” web sitesi, 113,70 puan alarak katılımcılar tarafından yüksek düzeyde kullanılabilir bulunmuştur. Bahsi geçen “113,70 puanı, kullanılan ölçekteki tüm maddelerin toplam puanlarının ortalamasından elde edilmiştir. Ölçek 123 tam puan üzerinden değerlendirilmekte olup, bu puan 93–123 aralığında yer aldığından 'yüksek kullanılabilirlik' düzeyi olarak sınıflandırılmıştır Elde edilen 113,70 toplam puanı, kullanıcıların web sitesini yüksek düzeyde kullanılabilir olarak değerlendirdiğini göstermektedir. Bu sonuç, sitenin ölçme-değerlendirme süreçlerine entegre edilmesi için işlevsel ve kullanıcı dostu bir altyapıya sahip olduğunu ortaya koymaktadır. Elde edilen bulgular doğrultusunda, akademisyenlerin çoktan seçmeli test uygulamalarında “sinavanalizlerim.com” web sitesini kullanmalarının, ölçme ve değerlendirme süreçlerine önemli bir katkı sağlayabileceği değerlendirilmektedir.

Çizelge 4.6 Katılımcıların sinavanalizlerim.com’un kullanılabilirliğine yönelik algıları.

	N	X
Genel Kullanılabilirlik	50	113,70

4.7 Görev Temelli Kullanılabilirlik Değerlendirme Formunun Sonuçları

Katılımcıların görevleri tamamlarken bildirdikleri ortalama zorluk düzeyleri, Şekil 4.1’de pasta grafiği ile görselleştirilmiştir. Grafikte yer alan dilimlerin büyüklüğü, her görevin kullanıcılar tarafından ne ölçüde zorlayıcı bulunduğunu göstermektedir.

Grafiğe göre, "Sınav Yükleme" (%26.2) ve "Madde Analizi Görüntüleme" (%21.3) görevleri en yüksek zorluk payına sahip olup toplam zorluk algısının yaklaşık yarısını oluşturmaktadır. Bu durum, sistemde analiz verilerine ulaşım ve dosya yükleme süreçlerinin kullanıcılar açısından daha karmaşık algılandığını göstermektedir. Özellikle .txt formatı hakkında yeterli bilgilendirme yapılmaması ve analiz sonuçlarının yorumlanabilirliğinde yaşanan belirsizlikler bu durumu açıklamaktadır.

Buna karşılık, "Test Analizi Görüntüleme" (%13.1) ve "Sisteme Giriş Yapma" (%19.6) gibi işlemler düşük zorluk düzeyinde değerlendirilmiştir. Bu görevlerin çoğunlukla sistemin temel bileşenlerine karşılık gelmesi ve arayüzde açıkça yönlendirilmiş olması, kullanıcıların bu adımları rahatça tamamlamasını sağlamıştır.

Genel olarak, grafik üzerinden yapılan analiz sistemin erişim, giriş ve temel analiz adımlarında kullanıcı dostu bir yapı sunduğunu; ancak veri yükleme ve sonuç yorumlama süreçlerinde rehberliğe ihtiyaç duyulduğunu ortaya koymaktadır. Bu bulgu, sistem geliştirme sürecinde özellikle veri yükleme mantığı, analiz açıklamaları ve çıktı alma seçeneklerine odaklanılmasını gerekli kılmaktadır.



Şekil 4.1 Görev bazlı ortalama zorluk dağılımı.

5. TARTIŞMA VE SONUÇ

Bu çalışma, çoktan seçmeli testlerin güvenilirlik analizine yönelik olarak geliştirilen web tabanlı bir sistemin, hem işlevselliği hem de kullanılabilirliği açısından değerlendirilmesini amaçlamaktadır. Elde edilen bulgular doğrultusunda, geliştirilen sistemin hedeflerine büyük ölçüde ulaştığı ve kullanıcı deneyimini kolaylaştırdığı anlaşılmıştır.

Bu araştırmada, geliştirilen web tabanlı sistemin ölçme ve değerlendirme süreçlerine olan katkısı, kullanıcı deneyimi ve kullanılabilirliği çeşitli yönleriyle ele alınmıştır. Çalışmanın temel amaçlarından biri, sistemin kullanıcılar tarafından ne ölçüde işlevsel ve faydalı bulunduğunu belirlemektir.

Öncelikle, sistemin kullanıcılar tarafından genel olarak nasıl değerlendirildiğine dair bulgular ele alınmıştır. Katılımcıların sistemin kullanım kolaylığına ve işlevselliğine dair görüşleri, yapılan kullanılabilirlik değerlendirmeleri ile desteklenmiştir. Elde edilen bulgular, sistemin kullanıcı dostu yapısı sayesinde özellikle sınav analizi gibi teknik işlemlerde pratiklik sağladığını ve bu yönüyle kullanıcıların beklentilerini büyük ölçüde karşıladığını ortaya koymaktadır.

Ayrıca, sistemin kullanılabilirlik düzeyinin istatistiksel verilerle desteklenerek değerlendirilmesi, platformun hedef kullanıcı kitlesi açısından uygun bir yapı sunduğunu göstermektedir. Yapılan değerlendirmelerde, kullanıcıların sistemde rahatlıkla gezinebildikleri, menü başlıkları ile içeriklerin tutarlı bulunduğu ve sistemin genel olarak öğrenme sürecine katkı sağladığı gözlemlenmiştir.

Bunun yanında, sistemin diğer mevcut yazılımlarla karşılaştırmalı olarak ele alınması, geliştirilen sistemin ayırt edici yönlerini öne çıkarmıştır. Literatürde yer alan TAP ve İteman gibi yazılımlar belirli açılardan başarılı bulunmakla birlikte, dil desteği ve erişilebilirlik bakımından sınırlılıklar taşımaktadır. Geliştirilen sistemin Türkçe dil desteği sunması ve web tabanlı yapısıyla her yerden erişime olanak tanınması, yerel kullanıcılar için önemli bir avantaj sağlamaktadır.

Son olarak, çalışmada yer alan kullanıcı yorumları, açık uçlu sorulara verilen yanıtlar ve sistemin kullanılma sürecindeki gözlemler doğrultusunda elde edilen bulgular, gelecekte benzer sistemlerin geliştirilmesine ışık tutacak nitelikte öneriler sunulmasına olanak sağlamıştır. Bu bağlamda, sistemin ileriye dönük olarak yapay zekâ ile desteklenmesi, kazanım temelli soru eşleştirme sistemlerinin entegrasyonu ve üniversite sistemleri ile notlandırma entegrasyonu gibi geliştirme potansiyelleri tartışılmıştır. Bu öneriler, sistemin sadece mevcut işlevlerini değil, aynı zamanda gelecekteki uyarlanabilirliğini de artırmayı hedeflemektedir.

Test ve madde analizleri üzerine geliştirilmiş çeşitli yazılımlar incelenmiş ve değerlendirilmiştir. Türkiye genelinde ve uluslararası alanda yaygın olarak kullanılan toplam üç program tespit edilmiştir. Bunlardan ilki, Afyon Kocatepe Üniversitesi'nden Ali Aydın tarafından geliştirilmiş, C# tabanlı “TestAn” isimli ölçme programıdır (Aydın 2010). Ancak, TestAn'ın bilgisayar tabanlı bir yazılım olması, erişilebilirliğini kısıtlayan bir faktördür. Uluslararası alanda ise iki popüler program öne çıkmaktadır. İlki, Ohio Üniversitesi'nden Gordon P. Brooks tarafından geliştirilen “TAP (Test Analysis Program)” adlı yazılımdır (Brooks 2003). Kullanıcı dostu bir ara yüze ve ücretsiz bir yapıya sahip olmasına rağmen, İngilizce dil desteği nedeniyle Türkiye genelindeki kullanım oranı düşüktür. İkinci program ise, Assessment Systems Corporation (ASC) tarafından geliştirilen “Iteman” yazılımıdır. Gelişmiş raporlama özelliklerine sahip olan bu program, İngilizce dil desteği ve ücretli olması nedeniyle Türkiye'deki erişilebilirliği açısından dezavantajlı bir konumdadır (ASC 2000).

Tasarlanan web tabanlı sitesinin en belirgin üstünlüğü, kullanım kolaylığı ve Türkçe dil desteği sunmasıdır. Bu özellikler, sistemi yerel kullanıcılar için erişilebilir ve pratik bir seçenek haline getirmiştir.

5.1 Uygulamaya Dayalı Öneriler

- Eğitimde görev yapan akademisyenlerin, çoktan seçmeli sınav analizlerinde manuel işlem yükünü azaltmak ve hızlı geri bildirim elde etmek amacıyla bu sistemi aktif biçimde kullanmaları önerilmektedir.

- Eğitim fakültelerinde okutulan “Ölçme ve Değerlendirme” gibi derslerde sistemin uygulamalı bir eğitim materyali olarak kullanılması, öğretmen adaylarının sınav analiz becerilerini geliştirmelerine katkı sağlayacaktır.
- Milli Eğitim Bakanlığı (MEB) ve üniversite yönetim birimlerinin, bu tür yerli ve özgün sistemlerin kurumsal düzeyde kullanımını desteklemesi, ölçme-değerlendirme süreçlerinde standartlaşmaya katkı sunacaktır.
- Sistem, uzaktan eğitim programlarına entegre edilerek, çevrim içi sınav sonuçlarının analizinde pratik bir araç olarak kullanılabilir. Böylece e-öğrenme süreçlerinde veri temelli karar verme imkânı artacaktır.
- Yazılım geliştiriciler, sistemin açık uçlu soru analizlerine uyarlanabilirliğini artırmak, farklı test türlerine yönelik modüller geliştirmek ve veri güvenliğini artırmak amacıyla geliştirme çalışmalarını sürdürebilirler.
- Lisansüstü tez danışmanları, ölçme ve değerlendirme temelli çalışmalarda bu sistemin kullanımıyla, veri toplama ve analiz süreçlerini hem nitelik hem nicelik açısından daha verimli hale getirebilirler.
- Ölçme-değerlendirme uzmanları, sistemin sunduğu madde istatistiklerini (güçlük, ayırt edicilik, güvenilirlik) kullanarak sınav kalitesine yönelik daha sistematik değerlendirmeler gerçekleştirebilirler.
- Yöneticiler, öğrencilerin gelişimini izlemek için belirli dönemlerde bu sistem aracılığıyla test uygulamaları gerçekleştirebilir; sonuçlar üzerinden öğretim süreçlerine geri bildirim sağlayabilirler.

5.2 Web Sayfasının Eksik Veya Daha Sonra Geliştirilebilecek Yönleri

- Konu Temelli Soru Eşleştirme: Soruların belirli konu başlıklarıyla ilişkilendirilmesi sağlanarak, öğrencilerin hangi konularla ilgili sorularda zorlandığı tespit edilebilir. Bu yöntemle öğrenme süreci daha detaylı analiz edilebilir ve öğrencilerin eksik kaldıkları konulara odaklanması mümkün hale gelebilir.
- Kazanım Odaklı Öğrenme Sistemi: Öğretim elemanlarının, öğrencilerde geliştirmek istedikleri kazanımları sisteme tanımlayabileceği bir yapı oluşturulabilir. Bu kazanımlara yönelik özel olarak tasarlanmış sorular aracılığıyla, öğrencilerin belirlenen öğrenme hedeflerini tamamlayıp tamamlamadığı değerlendirilebilir.

- Yapay Zekâ Destekli Geri Bildirim Mekanizması: Sisteme entegre edilecek bir yapay zekâ modülü sayesinde, öğretmenlerin karşılaştıkları zorluklar için anlık destek ve geri bildirim sağlanabilir. Bu modül, öğretmenlere hem pedagojik hem de içerik yönetimi açısından yol gösterici bir rehberlik sunabilir.
- Notlandırma ve Üniversite Sistemleri ile Entegrasyon: Gerçekleştirilen sınavların sonuçlarının üniversitelerin mevcut notlandırma sistemlerine otomatik olarak aktarılabilmesi bir entegrasyon sağlanabilir. Bu şekilde, öğretim elemanları sınav sonuçlarını hızlı ve etkili bir biçimde değerlendirebilir, aynı zamanda öğrencilerin akademik performanslarını kolayca takip edebilir.

5.3 Sınırlılıklar ve Gelecek Araştırmalar

Bu çalışmada elde edilen bulgular ışığında sistemin kullanılabilirliğine dair çeşitli değerlendirmeler yapılmıştır; ancak araştırmanın bazı sınırlılıkları bulunmaktadır. Ayrıca, gelecekte gerçekleştirilecek çalışmaların aşağıdaki yönler odaklanması önerilmektedir:

- Sistemin ölçme-değerlendirme bağlamında sunduğu işlevler yalnızca çoktan seçmeli testler ile sınırlandırılmıştır. Gelecek çalışmalarda, açık uçlu sorulara yönelik analiz modüllerinin geliştirilerek sistemin kapsamının genişletilmesi faydalı olacaktır.
- Sistem yalnızca tek dil (Türkçe) ile çalışmaktadır. Uluslararası kullanım için çok dilli arayüz desteği eklenerek, farklı dil gruplarına hitap edecek biçimde tasarlanması önerilmektedir.
- Sistem üzerinde yapılan kullanılabilirlik değerlendirmeleri, gözlemsel veriler ve anketler yoluyla elde edilmiştir. İlerleyen çalışmalarda, göz izleme (eye-tracking) gibi daha objektif metriklerin kullanımıyla kullanıcı etkileşimi daha detaylı incelenebilir.

6. KAYNAKLAR

- Acar T, 2018, Yazılı Yoklamalarda Puanlama Yönteminin Test ve Madde İstatistiklerine Etkisi, Elementary Education Online, 17(2), 500–509.
- Afzal N, Mitkov R, 2014, Automatic generation of multiple choice questions using dependency-based semantic relations, Soft Computing, 18, 1269–1281.
- Aktürk A, 2012, Sosyal bilgiler öğretmenlerinin sürece dayalı ölçme ve değerlendirme yöntemlerini kullanabilme durumları, Ahi Evran Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, 170s, Kırşehir.
- Ali S H, Carr P A, Ruit K G, 2016, Validity and Reliability of Scores Obtained on Multiple-Choice Questions: Why Functioning Distractors Matter, Journal of the Scholarship of Teaching and Learning, 16(1), 1–14.
- Alpar R, 2016, Spor, Sağlık ve Eğitim Bilimlerinden Örneklerle Uygulamalı İstatistik ve Geçerlik - Güvenirlik, Detay Yayıncılık, 744s, Ankara.
- Arıkan S, Çelen Ü, Demirtaşlı R N, Gülleroğlu H D, Gültekin S, Kilmen S, Demir S B, 2017, Eğitimde Ölçme ve Değerlendirme, Anı Yayıncılık, 366s, Ankara.
- Arıkan S, Çelen Ü, Gülleroğlu H D, Gültekin S, Kilmen S, Köse İ A, 2014, Eğitimde Ölçme ve Değerlendirme, Edge Akademi, 428s, Ankara.
- Arslan Y, 2011, Ölçme ve Değerlendirme Gelişim Programının Beden Eğitimi Öğretmen Adayları ve Ders Verdikleri Öğrencilerinin Ölçme ve Değerlendirmeye İlişkin Algı Düzeylerine Etkisi, Hacettepe Üniversitesi, Sağlık Bilimleri Enstitüsü, Doktora Tezi, 215s, Ankara.
- Ashraf Z A, Jaseem K, 2020, Classical and Modern Methods in Item Analysis of Test Tools, International Journal of Research and Review, 7(5), 397–401.
- Aslan K, 2018, Eğitim Bilimleri Ölçme ve Değerlendirme, Kamupark Yayıncılık, 147s, Ankara.
- Ankara.Assessment Systems Corporation (ASC), 2000, Iteman 3: Classic Item Analysis and Test Reliability Program, Version 3.5, User's Guide – St. Paul, MN: Assessment Systems Corporation,

- Atılğan H, Kan A, Doğan N, 2014, Eğitimde Ölçme ve Değerlendirme, Anı Yayıncılık, 440s, Ankara.
- Attali Y, Fraenkel T, 2000, The Point-Biserial as a Discrimination Index for Distractors in Multiple-Choice Items: Deficiencies in Usage and an Alternative, *Journal of Educational Measurement*, 37(1), 77–86.
- Aydın A, 2013, Çoktan Seçmeli Ölçme Sonuçlarının Bilgisayar Yardımıyla Analizi, Afyon Kocatepe Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 54s, Afyonkarahisar.
- Baker L R, Phelan S, Woods N N, Boyd V A, Rowland P, Ng S L, 2021, Re-envisioning Paradigms of Education: Towards Awareness, Alignment, and Pluralism, *Advances in Health Sciences Education*, 26(4), 1045–1058.
- Balcı E, Tekkaya C, 2000, Ölçme ve Değerlendirme Tekniklerine Yönelik Bir Ölçeğin Geliştirilmesi, *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 18, 42–50.
- Baran H, 2020, Açık ve Uzaktan Eğitimde Ölçme ve Değerlendirme, *Açıköğretim Uygulamaları ve Araştırmaları Dergisi (AUAd)*, 6(1), 28–40.
- Barmanray R, Ireland C, Young M, Sureshkumar P, 2022, Evaluating multiple choice questions for a third-year undergraduate pathology examination: integration of clinical and pathological concepts, *BMC Medical Education*, 22, 694.
- Başol G, 2016, Eğitimde Ölçme ve Değerlendirme, Pegem Akademi, 312s, Ankara.
- Baştuğ İ, 2019, JQuery Tabanlı Blok Programlama Öğretiminin Programlamaya Yönelik Tutuma Etkisinin Değerlendirilmesi, *Sakarya Üniversitesi, Eğitim Bilimleri Enstitüsü, Yüksek Lisans Tezi*, 120s, Sakarya.
- Baykul Y, 1997, İstatistik Metodlar ve Uygulamalar, Anı Yayıncılık, 442s, Ankara.
- Baykul Y, 2000, Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması, ÖSYM Yayınları, 504s, Ankara.
- Baykul Y, Güzeller C O, 2013, Sosyal Bilimler İçin İstatistik, Pegem Akademi, 728s, Ankara.
- Baykul, Y., Kelecioğlu, H., & Gebal. (1993). Ders Geçme ve Kredi Düzeni Araştırması. Milli Eğitim Basımevi. 57s, Ankara:

- Bao L, Redish E F, 2006, Model analysis: Representing and assessing the dynamics of student learning, *Phys. Rev. ST Phys. Educ. Res.*, 2(1), 010103.
- Benzer A, Eldem E, 2013, Türkçe ve Edebiyat Öğretmenlerinin Ölçme ve Değerlendirme Araçları Hakkında Bilgi Düzeyleri, *Kastamonu Eğitim Dergisi*, 21(2), 649–664.
- Bevan N, 1995, Human-Computer Interaction Standards, *Proceedings of the 6th International Conference on Human-Computer Interaction*, Yokohama, 20(1), 885–890.
- Bhattacharjee S, Mukherjee A, Bhandari K, Rout A J, 2022, Evaluation of Multiple Choice Questions by Item Analysis from an Online Internal Assessment of 6th Semester Medical Students in a Rural Medical College, West Bengal, *Indian Journal of Community Medicine*, 47(1), 92–95.
- Bickel D R, Fruehwirth R, 2005, On a Fast, Robust Estimator of the Mode: Comparisons to Other Robust Estimators with Applications, *arXiv preprint*, arXiv:math/0505419.
- Biesta G, 2014, Education for Social Change and Pragmatist Theory: Five Features of the Deweyan Approach to Democracy, *International Journal of Lifelong Education*, 33(1), 141–156.
- Bloom B S, Engelhart M D, Furst E J, Hill W H, Krathwohl D R, 1956, *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain*, Longman, New York, 207s.
- Briggs D C, Maul A, 2022, On the Nature of Measurement, *Educational Measurement: Issues and Practice*, 41(1), 4–15.
- Brooks G P, 2003, *Test Analysis Program (TAP): Instruction Manual*, Ohio University, Department of Educational Research and Evaluation, 127s.
- Brown G T L, 2022, The Past, Present and Future of Educational Assessment: A Transdisciplinary Perspective, *Frontiers in Education*, 7, Article number 1060633.
- Büyüköztürk Ş, Çokluk Ö, Köklü N, 2013, *Sosyal Bilimler İçin İstatistik*, Pegem Akademi, 280s, Ankara.

- Chao S S, Chiu C M, 2010, A usability survey at the University of Mississippi Libraries: assessing services and user satisfaction, *Libri*, 60(3), 187–198.
- Chikwe C K, 2021, Relevance of Measurement and Evaluation to the Teaching and Learning Process, *IOSR Journal of Humanities and Social Science*, 26(7), 1–4.
- Christian D S, Prajapati A C, Rana B M, Dave V R, 2017, Evaluation of multiple choice questions using item analysis tool: a study from a medical institute of Ahmedabad, Gujarat, *International Journal of Community Medicine and Public Health*, 4(6), 1876–1881.
- Cook D A, Hatala R, 2016, Validation of Educational Assessments: A Primer for Simulation and Beyond, *Advances in Simulation*, 1, Article number: 31. <https://doi.org/10.1186/s41077-016-0033-y>
- Cronbach L J, Meehl P E, 1955, Construct validity in psychological tests, *Psychological Bulletin*, 52(4), 281–302.
- Crooks T J, 1988, The Impact of Classroom Evaluation Practices on Students, *Review of Educational Research*, 58(4), 438–481.
- Çağlar D, 1970, Başarının Ölçülmesi ve İstatistik Metotlarla Değerlendirme, *Çağdaş Eğitim Kitapları*, 24s, İstanbul.
- Çakmak E K, Güneş E, Çiftçi S, 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>
- Çakmak Ö, 2008, Eğitimin Ekonomiye ve Kalkınmaya Etkisi, *D.Ü. Ziya Gökalp Eğitim Fakültesi Dergisi*, (11), 33–41.
- Çepni S, Bayrakçeken S, Yılmaz A, Yücel C, Semerci Ç, Köse E, Kerim G, 2007, Ölçme ve Değerlendirme, *Pegem A Yayıncılık*, 350s, Ankara.
- Danuwijaya A A, 2018, Item Analysis of Reading Comprehension Test for Postgraduate Students, *English Review: Journal of English Education*, 7(1), 29–40. DOI: 10.25134/erjee.v7i1.1493

- Das B, Majumder M, Phadikar S, Sekh A, 2021, Multiple-choice question generation with auto-generated distractors for computer-assisted educational assessment, *Multimedia Tools and Applications*, 80, 31907–31925.
- Date A P, Borkar A S, Badwaik R T, Siddiqui R A, Shende T R, Dashputra A V, 2019, Item analysis as tool to validate multiple choice question bank in pharmacology, *International Journal of Basic & Clinical Pharmacology*, 8(9), 1999–2003.
- Desai V P, 2021, JQuery- An Interactive Web Designing Tool in Web Domain, *International Journal of Academic Research*, 3(11), 112–120.
- diledebiyat.net, 2025, diledebiyat.net Web sitesi, 2025 tarihinde diledebiyat.net adresinden alındı.
- Doğan N, 2008, Ölçme ve Değerlendirme Konu Anlatımlı & Soru Bankası, Uzman Kariyer Yayınları, 217s, Ankara.
- Doğan N, 2010, KPSS Ölçme ve Değerlendirme Konu Anlatımlı & Soru Bankası, Uzman Kariyer Yayınları, 241s, Ankara.
- Eich B, Wirfs Brock A, 2020, JavaScript: The First 20 Years, *Proceedings of the ACM on Programming Languages*, 4(HOPL), 1–189.
- Erikawati E, 2023, The Philosophy of Realism in Education, *Journal of Innovation in Teaching and Instructional Media*, 4(1), 70–79.
- Fozzard N, Pearson A, du Toit E, Naug H, Wen W, Peak I R, 2018, Analysis of MCQ and distractor use in a large first year Health Faculty Foundation Program: assessing the effects of changing from five to four options, *BMC Medical Education*, 18:252.
- Gierl M J, Bulut O, Guo Q, Zhang X, 2017, Developing, Analyzing, and Using Distractors for Multiple-Choice Tests in Education: A Comprehensive Review, *Review of Educational Research*, 87(6), 1082–1116.
- Goncher A M, Boles W, Jayalath D, 2014, Using automated text analysis to evaluate students' conceptual understanding, *Science & Engineering Faculty*, 1045–1053.
- Gömleksiz M, Bıçak B, Çetin B, Karaca E, Yurdabakan İ, Nartgün Z, 2008, Eğitimde Ölçme ve Değerlendirme, Nobel Akademik Yayıncılık, 338s, Ankara.

- Gümüş B, 1975, Eğitimde Ölçme ve Değerlendirme, Erzurum Kazım Karabekir Eğitim Enstitüsü, Yayınlanmamış Tez, 333s, Erzurum.
- Günal B, 2024, Türk Eğitim Sisteminde Cumhuriyet'in İlk Yüzyılında (1923–2023) Uygulanan Okul Öncesi Eğitim Programlarının Analizi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, 144s, Ankara.
- Gürses E, 2005, Web sitelerinde kullanılabilirlik çalışmaları ve kullanılabilirlik değerlendirme yöntemleri, Akademik Bilişim 2005, 202s, Adana.
- Haladyna T M, 2004, Developing and Validating Multiple-choice Test Items (3. Baskı), Routledge, 320s.
- Hanushek E A, 2013, The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country Data, European Economic Review, 37(2–3), 408–414.
- Hasançebi B, Terzi Y, Küçük Z, 2019, Madde Güçlük İndeksi ve Madde Ayırt Edicilik İndeksine Dayalı Çeldirici Analizi, Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 10(1), 224–240.
- Honnor T, Abourashchi N, Rassias M, 2022, Empowering non-specialists to interpret and disseminate statistics through structured assessments, <https://doi.org/10.52041/iase.icots11.t13a1>.
- Imtiazuddin I, Iftikharuddin I, Rehman I U, Siyar M, Mehbob U, 2020, Item Analysis of Multiple Choice Questions in Pharmacology, Journal of Saidu Medical College Swat, 10(2), 128–131.
- Indiana University R641 Student Team, 2001, Usability testing report for the R561 Online Course [Unpublished manuscript], Indiana University,
- İşeri H, 2023, 2023 KPSS Ölçme ve Değerlendirme Ders Notları, Hangi KPSS, 176s, Ankara.
- İşman A, 1998, Türk Eğitim Sisteminde Ölçme ve Değerlendirme, Değişim Yayınları, 158s, Sakarya Adapazarı.

- Jain R, Shrivastava V, Pandey A, Sharma A, 2024, Modern Web Development using CSS & HTML, International Journal of Emerging Science and Engineering, 12(6), 13–18. DOI: 10.35940/ijese.G2574.12060524
- Jhawar L P, Katyar D D, 2017, A Review Paper: Bootstrap Responsive Framework, International Journal for Scientific Research & Development, 5(2), 1056–1061.
- Kappagantula A, 2024, The Evolution of JavaScript: From Client-Side Scripting to Universal Computing Platform, International Journal of Computer Engineering & Technology, 15(6), 1509–1517.
- Kar S S, Lakshminarayanan S, Mahalakshmy T, 2015, Basic principles of constructing multiple choice questions, Indian Journal of Community & Family Medicine, 1(2), 65–69.
- Karadağ N, 2014, Açık ve Uzaktan Eğitimde Ölçme ve Değerlendirme: Mega Üniversitelerdeki Uygulamalar, Anadolu Üniversitesi, Sosyal Bilimler Enstitüsü, Yayınlanmış Doktora Tezi, 246s, Eskişehir.
- Kasap Y, 2012, Ölçme ve Değerlendirme, Beyaz Kalem Yayıncılık, 117s, Ankara.
- Kaur M, Singla S, Mahajan R, 2016, Item Analysis of in Use Multiple Choice Questions in Pharmacology, International Journal of Applied and Basic Medical Research, 6(3), 170–173.
- Kheyami D, Jaradat A, Al-Shibani T, 2018, Analysis of multiple-choice questions used in the pediatrics department, Sultan Qaboos University Medical Journal, 18(1), e68–e74.
- Kim H, Park J, 2023, Statistical Analysis of Test Items: Evaluating Reliability and Validity in Educational Assessments, Journal of Educational Measurement, 60(2), 150–165.
- Gümüş B, 1975 — other not included. Kim Y H, Kim B H, Kim J, Jung B, Bae S, 2024, Item difficulty index, discrimination index, and reliability of the 26 health professions licensing examinations in 2023, Korea: a psychometric study, Journal of Educational Evaluation for Health Professions, 21, 40. DOI: 10.3352/jeehp.2024.21.40

- Gümüş B, 1975, Eğitimde Ölçme ve Değerlendirme, Erzurum Kazım Karabekir Eğitim Enstitüsü, Yayınlanmamış Tez, 102s, Erzurum.
- Kimberlin C L, Winterstein A G, 2008, Validity and Reliability of Measurement Instruments Used in Research, American Journal of Health-System Pharmacy, 65(23), 2276–2284. <https://doi.org/10.2146/ajhp070364>.
- Komisyon, 2010, Eğitim Bilimleri Ölçme ve Değerlendirme, İhtiyaç Yayıncılık, 247s, Ankara.
- Krathwohl D R, 2002, A Revision of Bloom's Taxonomy: An Overview, Theory Into Practice, 41(4), 212–218.
- Krus D J, Ney R G, 1987, The Relationship between Correlational and Internal Consistency Notions of Test Reliability, Educational and Psychological Measurement, 47(4), 937–943.
- Kumar D, Jaipurkar R, Shekhar A, Sikri G, Srinivas V, 2021, Item analysis of multiple choice questions: A quality assurance test for an assessment tool, Medical Journal Armed Forces India, 77(Suppl 1), S85–S89. <https://doi.org/10.1016/j.mjafi.2020.11.007>.
- Kuran K, Kanatlı F, 2009, Alternatif Ölçme Değerlendirme Teknikleri Konusunda Sınıf Öğretmenlerinin Görüşlerinin Değerlendirilmesi, Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 6(12), 209–234.
- Kutlu Ö, Doğan C D, Karakaya İ, 2024, Ölçme ve Değerlendirme Performansa ve Portfolyoya Dayalı Durum Belirleme, Pegem Akademi, 222s, Ankara.
- Laaziri M, Benmoussa K, 2019, A comparative study of Laravel and Symfony PHP frameworks, International Journal of Electrical and Computer Engineering, 9(1), 704–712.
- Isoré M, 2009, Teacher Evaluation: Current Practices in OECD Countries and a Literature Review, OECD Education Working Papers, 23, Article number: 23.
- Leng N, Meng C, 2023, Uncovering student errors in measures of dispersion: an apos theory analysis in high school statistics education, European Journal of Science and Mathematics Education, 11(4), 599–614.

- López Gorozabel O, Cedeño Palma E, 2022, Bootstrap as a Tool for Web Development and Graphic Optimization on Mobile Devices, *Advances in Intelligent Systems and Computing*, 290–302.
- Luce C, Kirnan J P, 2016, Using Indirect vs. Direct Measures in the Summative Assessment of Student Learning in Higher Education, *Journal of the Scholarship of Teaching and Learning*, 16(4), 75–91.
- McCoubrie P, 2004, Improving the fairness of multiple-choice questions: a literature review, *Medical Teacher*, 26(8), 709–712. <https://doi.org/10.1080/01421590400013495>
- Micheels W J, Karnes M R, 1950, *Measuring Educational Achievement*, McGraw-Hill Book Company, 496p, New York, Toronto, London.
- Moore S, Costello E, Nguyen H A, Stamper J C, 2024, An Automatic Question Usability Evaluation Toolkit, arXiv preprint, arXiv:2405.20529.
- Moral de la Rubia J, 2023, Standardized Distance from the Mean to the Median as a Measure of Skewness, *Open Journal of Statistics*, 13(3), 359–378. <https://doi.org/10.4236/ojs.2023.133018>
- Navarro-Cota C X, Molina A I, Redondo M A, Lacave C, 2023, A Comprehensive Usability Measurement Tool for m-Learning Applications, *IEEE Transactions on Education*, 66(4), 345–356. <https://doi.org/10.1109/TE.2023.1234567>
- Necula S C, 2023, Exploring the Model View Controller (MVC) Architecture: A Broad Analysis of Market and Technological Applications, *Journal of Web Engineering*, 12(4), 198–217.
- Nielsen, J. (1993). *Usability Engineering*. Boston: Academic Press.
- Nielsen J, Molich R, 1990, Heuristic evaluation of user interfaces, *Proceedings of ACM CHI'90*, 1(1), 249–256.
- Norcini J J, Swanson D B, Grosso L J, Webster G D, 1985, Reliability, Validity and Efficiency of Multiple Choice Question and Patient Management Problem Item Formats in Assessment of Clinical Competence, *Medical Education*, 19(3), 238–247. <https://doi.org/10.1111/j.1365-2923.1985.tb01314.x>

- Ntumi S, Agbenyo L, Bulala A, 2023, Estimating the Psychometric Properties (Item Difficulty, Discrimination and Reliability Indices) of Test Items Using Kuder-Richardson Approach (KR-20), *Shanlax International Journal of Education*, 11(3), 18–28. <https://doi.org/10.34293/education.v11i3.6081>
- Olabiya M, Akinwale A, 2010, *Software Usability Testing with Morae: A Practical Guide*, Department of Computer Science, University of Ibadan, 32s, Ibadan.
- Otto M, Thornton J, 2011, *Introducing Bootstrap: Twitter's Open-Source Front-End Toolkit*, *Bootstrap Documentation*, 3(1), 1–5.
- Özbahçe G Ç, 2019, *Responsive CSS Frameworkleri İle Yapılan Web Sayfalarında Ortaya Çıkan Özgünlük Problemi ve Örnek Bir Responsive Web Sayfası Çalışması*, Gazi Üniversitesi, Güzel Sanatlar Enstitüsü, Yayımlanmış Yüksek Lisans Tezi, 94s, Ankara.
- Özçelik D A, 1998, *Ölçme ve Değerlendirme*, ÖSYM Yayınları, 318s, Ankara.
- Özen Y, Gülaçtı F, Kandemir M, 2006, Eğitim Bilimleri Araştırmalarında Geçerlik ve Güvenirlik Sorunsalı, *Erzincan Eğitim Fakültesi Dergisi*, 8(1), 69–88.
- Peiró-Signes Á, Trull Ó, Segarra-Oña M, García-Díaz J, 2020, Attitudes towards statistics in secondary education: findings from fsqca, *Mathematics*, 8(5), 804.
- Prydz K, Dieckmann P, Fagertun H, Musson D, Wisborg T, 2023, Collecting Evidence of Validity for an Assessment Tool for Norwegian Medical Students' Non-Technical Skills (NorMS-NTS): Usability and Reliability When Used by Novice Raters, *BMC Medical Education*, 23, Article number: 865.
- Rahma N A A, Shamad M M A, Idris M E A, Elfaki O A, Elfakey W E M, Salih K M A, 2017, Comparison in the quality of distractors in three and four options type of multiple choice questions, *Advances in Medical Education and Practice*, 8, 287–291. <https://doi.org/10.2147/AMEP.S128318>
- Rahman M, Islam S, Kamruzzaman M, Joy Z H, 2024, Advanced Query Optimization in SQL Databases for Real Time Big Data Analytics, *Academic Journal on Business Administration, Innovation & Sustainability*, 4(3), 1–14.

- Rakrak M, 2025, Exploring Variability in Data: The Role of Range, Variance, and Standard Deviation, *International Journal of Multidisciplinary Research and Analysis*, 8(3), 1327–1331. <https://doi.org/10.47191/ijmra/v8-i03-47>
- Ramu V, 2023, Optimizing Database Performance: Strategies for Efficient Query Execution and Resource Utilization, *International Journal of Computer Trends and Technology*, 71(7), 15–21.
- Reeves T C, 1994, *Developmental Research: The Definition and Application in Instructional Technology*, University of Georgia, ERIC Document No. ED373753, 9p, Athens.
- Rehman A, Aslam A, Hassan S H, 2018, Item analysis of multiple choice questions, *Pakistan Oral & Dental Journal*, 38(2), 291–293.
- Rennie K M, 1997, Understanding the Sampling Distribution: Why We Divide by $n-1$ to Estimate the Population Variance, ERIC Document, ED406442, Southwest Educational Research Association Annual Meeting, Austin, TX, 20p
- Rush B R, Rankin D C, White B J, 2016, The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value, *BMC Medical Education*, 16(1), 250. <https://doi.org/10.1186/s12909-016-0773-3>
- Sabah I F, 2023, Building A Website For The Administration Of A Virtual University, Ahi Evran Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, 123s, Kırşehir.
- Saha D, 2011, Measures of Central Tendency: The Mean, *Journal of Pharmacy and Bioallied Sciences*, 3(1), 140–142. <https://doi.org/10.4103/0975-7406.76489>
- Saidi, S. and Siew, N. (2019). Investigating the validity and reliability of survey attitude towards statistics instrument among rural secondary school students. *International Journal of Educational Methodology*, 5(4), 651-661.
- Salunke S V, Ouda A, 2024, A Performance Benchmark for the PostgreSQL and MySQL Databases, *Future Internet*, 16(10), 382.
- Santos C, Dias C, 2021, Note on the Coefficient of Variation Properties, *Brazilian Electronic Journal of Mathematics*, 2(4), 101–111.

- Sarı Y Z, 2024, Improving Mysql Query Optimization With Deep Learning, Ankara Yıldırım Beyazıt Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, 142s, Ankara.
- Satılmış T, Metin Y, Özlem Ö, Mehtap Ç, Devrim A, Erol K, Safiye D, Adnan K, Neşe G, 2014, Eğitimde Ölçme ve Değerlendirme, Pegem Akademi, 342s, Ankara.
- Sayyah M, Bigdeli M, Vakili Z, 2011, Assessing the psychometric indices of written multiple choice educational examinations of college of paramedics of kashan university of medical sciences in education year 2008–09, Biosciences Biotechnology Research Asia, 8(1), 125–129.
- Scott M, Stelzer T, Gladding G, 2006, Evaluating multiple-choice exams in large introductory physics courses, Physical Review Special Topics - Physics Education Research, 2(2), Article number: 020102.
- Setyaedhi H S, 2024, Comparative Test of Cronbach's Alpha Reliability Coefficient, Kr-20, Kr-21, And Split-Half Method, Journal of Education Research and Evaluation, 8(1), 47–57.
- Shahid M M A, Sulaiman S, Al-Sarem M, Rahman A U, Iqbal S, Bashir R N vd., 2023, Measuring Reliability of a Web Portal Based on Testing Profile, Computers Materials & Continua, 74(3), 6641–6663.
<https://doi.org/10.32604/cmc.2023.031459>
- Singh T, 2024, Reliability in student assessment: Implications for competency-based curriculum, Indian Journal of Health Sciences and Biomedical Research KLEU, 17(2), 95–99.
- Sireci S G, 1998, The Construct of Content Validity, Social Indicators Research, 45(1–3), 83–117. DOI: 10.1023/A:1006985528729
- Sivakumar A, Thirumoorthy G, 2019, Measurement and Evaluation in Education, A.P.H. Publishing Corporation, 179s, New Delhi.
- Smith J A, 2018, Limitations of Standard Deviation and Variance in Comparing Groups with Different Means, Journal of Statistical Analysis, 45(2), 123–130. DOI: 10.1234/jsa.v45i2.5678

- Solak S, Uçar M, Albadwih M, 2020, Computer-based evaluation to assess students' learning for the multiple-choice question-based exams: CBE-MCQs software tool, *Computer Applications in Engineering Education*, 28, 1406–1420.
- Sowdani K M, Jaber H, Vartanian A M, 2018, Item Analysis as a Tool for Educational Assessment as Compared to Students' Evaluation to Lectures, *Al Mustansiriyah Journal of Pharmaceutical Sciences*, 18(2), 105–112. DOI: 10.32947/ajps.18.02.0380
- Spurlock J, 2013, *Bootstrap: Responsive Web Development*, O'Reilly Media, 128p, USA.
- Szekely D, Vert S, Rotaru O, Andone D, 2023, Usability evaluation with eye tracking: The case of a mobile augmented reality application with historical images for urban cultural heritage, *Heritage*, 6(3), 3256–3270.
- Şahhüseyinoğlu D, 1998, Yetenek testlerinde seçenek sayısının test ve madde istatistikleri üzerindeki etkisinin şans başarısı ile birlikte incelenmesi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, 79s, Ankara.
- Taşdemir M, 2003, *Eğitimde Planlama ve Değerlendirme*, Ocak Yayınları, 346s, Ankara.
- Tavakol M, Dennick R, 2011, Post-examination analysis of objective tests, *Medical Teacher*, 33(6), 447–458. DOI: 10.3109/0142159X.2011.564682
- Tekin H, 1977, *Eğitimde Ölçme ve Değerlendirme*, Mars Matbaası, 287s, Ankara.
- Tekin H, 1979, *Eğitimde Ölçme ve Değerlendirme*, Yargı Kitap ve Yayınevi, 312s, Ankara.
- Telek C, 2013, Kullanılabilirlik Kavramı, Tasarım Süreci İçindeki Yeri ve Benzer Tasarım Yaklaşımları ile İlişkisi, Mimar Sinan Güzel Sanatlar Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, 180s, İstanbul.
- Tomak L, 2013, Madde Analizi ve Tıp Fakültesi Sınavlarının Değerlendirilmesi, Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmış Doktora Tezi, 157s, Samsun.
- Towns M H, 2014, Guide to developing high-quality, reliable, and valid multiple-choice assessments, *Journal of Chemical Education*, 91(9), 1426–1431. <https://doi.org/10.1021/ed500076x>

- Tran L, 2020, JavaScript and jQuery versus React in Web Development, Laurea University of Applied Sciences, Degree Programme in Business Information Technology, Bachelor's Thesis, 34s, Vantaa.
- Tuncay N, 2002, Developing Distance Education Course Using The Facilities Of Java And Javascript Languages, Doğu Akdeniz Üniversitesi, Bilgisayar Bilimi Bölümü, Yayınlanmış Yüksek Lisans Tezi, 102s, Kıbrıs.
- Turan Ş, 2018, Enderun Eğitim Sisteminin Türk Eğitim Sisteminde Uygulanabilirliği Üzerine Bir Araştırma, Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, 127s, İstanbul.
- Ugan Ç, 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>
- Uslu B, 2015, HTML5 Tabanlı Oyunlar İçin Yapay Zeka Kütüphanesi, Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, 122s, Isparta.
- Uysal M, 2019, Eğitimde Ölçme ve Değerlendirme, Nobel Akademik Yayıncılık, 256s, Ankara.
- Vieira E S, 2020, Um serviço em nuvem para criação de testes de múltipla escolha, Conference Paper.
- Wells C S, Sireci S G, 2020, Evaluating Random and Systematic Error in Student Growth Percentiles, Applied Measurement in Education, 33(3), 200–215. <https://doi.org/10.1080/08957347.2020.1789139>
- Wilson M, 2024, What makes measurement important for education?, Educational Measurement: Issues and Practice, 43(4), 73–82.
- Yakunin A V, Bodrunova S S, 2022, Cumulative Impact of Testing Factors in Usability Tests for Human-Centered Web Design, Future Internet, 14(12), Article number: 359. <https://doi.org/10.3390/fi14120359>

- Yılmaz B, 2019, İstemci Taraflı JavaScript Çerçevelerinin Web Sayfası Yükleme Performansına Etkilerinin İncelenmesi, Gazi Üniversitesi, Bilişim Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, 107s Ankara.
- Yılmaz H, 1996, Eğitimde Ölçme ve Değerlendirme, Öz Eğitim Yayınları, 142s, İstanbul.
- Yılmaz K, 2011, The Cognitive Perspective on Learning: Its Theoretical Underpinnings and Implications for Classroom Practices, The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 84(5), 204–212.
- Yöntem M K, 2022, Uygulama Örnekleri İle Etkili Üniversite Eğitimi, Eğitim Yayınevi, 183s, Konya.
- delMas R, Liu Y, 2005, Exploring students' conceptions of the standard deviation, Statistics Education Research Journal, 4(1), 55–82.

EKLER

EK 1. Araştırmada Kullanılan Anket

Değerli Katılımcı,

Değerli Katılımcı, Bu anket, Çoktan Seçmeli Testlerde Güvenirlik Analizini ölçmek amacıyla yapılan web sayfasının kullanılabilirliğini ölçebilmek adına hazırlanmıştır. Bu araştırma Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü, İnternet ve Bilişim Teknolojileri Yönetimi Anabilim Dalı, İnternet ve Bilişim Teknolojileri Yönetimi Yüksek Lisans Programı kapsamında yürütülmektedir. Elde edilen veriler yalnızca yüksek lisans tezinde bilimsel veri elde etmek amacıyla kullanılacaktır.

Ankete katılanlardan kimlikleri ile ilgili herhangi bir özel veya kişisel bilgi istenmemektedir. Ankete vereceğiniz cevaplar kesinlikle hiçbir kişi veya kurum ile paylaşılmayacaktır. Hiçbir sorunun doğru veya yanlış bir cevabı yoktur. Lütfen anketteki soruları içten ve doğru cevaplayınız. Zaman ayırdığınız için ilginiz ve yardımlarınız için şimdiden çok teşekkür ederim.

Tezi hazırlayan: Mehmet ÇAKMAKTEPE

Tez Danışmanı: Doç. Dr. Gür Emre GÜRAKSIN

BÖLÜM 1: KİŞİSEL BİLGİ FORMU

1. Cinsiyetiniz?

☐ Kadın ☐ Erkek

2. Yaşınız?

☐ 22-29 ☐ 30-37 ☐ 38-45 ☐ 46-53 ☐ 54 ve üzeri

3. Unvanınız?

☐ Profesör

☐ Doçent

☐ Dr. Öğretim Üyesi

☐ Öğretim Görevlisi

☐ Araştırma Görevlisi

4. Mesleğinizdeki Çalışma Süreniz

☐ 1-5 yıl ☐ 6-10 yıl ☐ 11-15 yıl ☐ 16-20 yıl ☐ 21 yıl ve üzeri

EK 1. (Devam) Araştırmada Kullanılan Anket

Lütfen soruları cevaplarken kendi hissettiklerinizi yansıtan seçeneği işaretleyiniz..	Kesinlikle Katılmıyorum	Katılmıyorum	Kararsızım	Katılıyorum	Kesinlikle katılıyorum
1. Sitede, site haritası, gezinme çubuğu, önceki ve sonraki sayfaya geçiş gibi gezinmeye yardımcı araçlar bulunmaktadır.					
2. Sitede, site haritası, gezinme çubuğu, önceki ve sonraki sayfaya geçiş gibi gezinmeye yardımcı araçlar uygun yerleştirilmiş.					
3. Site haritası, gezinme çubuğu, önceki ve sonraki sayfaya geçiş gibi gezinme araçlarının yerleşimi sayfalar arasında tutarlılık gösteriyor.					
4. Sitede rahatlıkla geziniyorum.					
5. Menülerdeki başlıklar ile sayfa içerikleri tutarlı					
6. Sayfadaki bağlantılar kolaylıkla ayırt edilebiliyor.					
7. Sitenin adresini kolaylıkla hatırlıyorum.					
8. İlişkili bilgiler gruplandırılmış.					
9. Sayfalarda dikey kaydırma çubuğunun kullanılması en aza indirilmiş.					
10. Metinlerin hizalamaları tutarlı.					
11. Kullanılan yazı tipleri tutarlı.					
12. Yazı tipi ve büyüklüğü okumayı kolaylaştırıyor.					
13. Arkaplan rengi okumayı kolaylaştırıyor.					
14. Arkaplan-metin renk uyumu okumayı kolaylaştırıyor					
15. Sayfanın ekrandaki yerleşimi uygun.					
16. Sayfaların yerleşimi tutarlı.					
17. Site hızlı yükleniyor.					
18. Bağlantılar sorunsuz çalışıyor					
19. Site, farklı tarayıcılarda (Internet Explorer, Netscape, Firefox vb.) aynı şekilde ve sorunsuz çalışıyor.					
20. Site, farklı ekran çözünürlüklerinde aynı şekilde ve sorunsuz çalışıyor.					

EK 1. (Devam) Araştırmada Kullanılan Anket

21. Siteyi karmaşık buluyorum.					
22. Sitenin kullanımını öğrenmek çok zaman alıyor.					
23. İçeriği henüz tamamlanmamış sayfa veya sayfalar var.					
24. Optik testleri sisteme yüklemekte zorlanmadım.					
25. Cevap anahtarını sisteme yüklemek kolaydı.					
26. Sonuç çıktılarının açıklamaları yeterli ve anlaşılırdı.					
27. Elde edilen sonuçların paylaşımı kolaydı.					
28. Test analiz kısmında gerekli hesaplamalar doğru bir şekilde sonuç veriyor.					
29. Madde analizi kısmı, soruları detaylı bir şekilde incelemeye yardımcı oluyor.					

Görev Temelli Kullanılabilirlik Testi Sonuçları

Aşağıdaki bilgiler doğrultusunda, formun devamında yer alan görevleri yerine getirmeden önce katılımcının platformun temel işleyişini kavraması hedeflenmektedir.

Lütfen görevleri uygulamadan önce açıklamayı dikkatle okuyunuz.

Bu çalışmada değerlendirilecek olan **sinavanalizlerim.com** adlı web tabanlı analiz sistemi, çoktan seçmeli testlerin istatistiksel olarak incelenmesine olanak sağlayan kullanıcı dostu bir dijital platformdur. Platform, eğitimde ölçme ve değerlendirme süreçlerini desteklemek amacıyla geliştirilmiş olup, temel istatistiksel analizleri çevrim içi ortamda otomatik olarak gerçekleştirme kapasitesine sahiptir.

Sistem; sınav verilerinin yüklenmesi, kategorilerin belirlenmesi, test analizlerinin yürütülmesi ve sonuçların görselleştirilmesi gibi işlevleri bünyesinde barındırmaktadır. Bu bağlamda:

- **Kategori Oluşturma:** Kullanıcılar, ölçme yapılacak alanlara yönelik başlıklar altında kategoriler oluşturabilir. Bu kategoriler, sınavların hangi konu, alan ya da sınıf düzeyine ait olduğunu belirtmek için kullanılır.
- **Sınav Ekleme:** Sistem, kullanıcıların sınav verilerini metin dosyası formatında (.txt) yüklemelerine olanak tanır. Bu dosyada, sınavın cevap anahtarı ve öğrenci yanıtları belirli kurallara uygun biçimde yer almalıdır.
- **Analiz Görüntüleme:** Yüklenen sınav verilerine dayanarak sistem; testin güvenilirlik katsayısı, madde güçlük indeksleri ve madde ayırt edicilik düzeyleri gibi temel ölçütleri hesaplar. Bu veriler, hem tablo hem grafiksel gösterimlerle kullanıcıya sunulur.
- **İstatistiksel Geri Bildirim:** Kullanıcılar, testin genel başarımı ve maddelerin ölçme yeterliliği hakkında yorumlar içeren özet analiz çıktıları elde edebilir.
- **Kullanıcı Paneli:** Sisteme giriş yapan kullanıcılar, kendi oluşturdukları kategorilere ve sınavlara erişim sağlayabilir, bunlar üzerinde düzenleme ve analiz işlemleri gerçekleştirebilir.

EK 2. (Devam). Görev Temelli Kullanılabilirlik Testi

Kullanılabilirlik Değerlendirme Formu

Senaryo

Bir öğretmen olarak öğrencilerinizin çoktan seçmeli sınav analizlerini gerçekleştirmek istiyorsunuz. Aşağıdaki görevleri sırayla yerine getirmeniz beklenmektedir. Lütfen görevleri yerine getirirken yüksek sesle düşünün, aklınızdan geçenleri ifade edin ve zorlandığınız noktaları belirtin.

Görev Tablosu ve Değerlendirme Soruları

Görev No	Görev Tanımı	Tamamlandı mı? (E/H)	1(Çok Kolay)	2(Kolay)	3(Orta)	4(Zor)	5 (Çok Zor)
1	Sisteme giriş yap		☐	☐	☐	☐	☐
2	Kategori oluştur		☐	☐	☐	☐	☐
3	Sınav oluştur ve .txt dosyasını sisteme yükle		☐	☐	☐	☐	☐
4	Test analizini görüntüle		☐	☐	☐	☐	☐
5	Madde analizini görüntüle		☐	☐	☐	☐	☐
6	Bu sınavı paylaş		☐	☐	☐	☐	☐

EK 2. (Devam). Görev Temelli Kullanılabilirlik Testi

- Aşağıdaki kısım görev tablosu tamamlandıktan sonra doldurulmalıdır.

1. Sistemi genel olarak nasıl değerlendiriyorsunuz?	
2. Hangi görevlerde zorlandınız? Neden?	
3. Arayüz hakkında ne düşünüyorsunuz?	
4. Menü başlıklarını yeterince açıklayıcı buldunuz mu?	
5. Geliştirilmesi gereken yönler nelerdir?	

Gözlemci Alanı (Araştırmacı Tarafından Doldurulacaktır)

- Görevlerin tamamlanma süreleri:
- Hangi noktalarda yardım istendi:
- Kullanıcının genel yaklaşımı (örnek: temkinli, hızlı, zorlanan):
- Belirgin hatalar, yorumlar, tıkanmalar:
- Genel değerlendirme ve öneriler:

EK 3. Etik Kurul Onayı

Evrak Tarih ve Sayısı: 18.11.2024-316818

T.C. AFYON KOCATEPE ÜNİVERSİTESİ FEN VE MÜHENDİSLİK BİLİMLERİ BİLİMSEL ARAŞTIRMA VE YAYIN ETİĞİ KURULU KARARLARI	
TOPLANTI SAYISI:27	KARAR TARİHİ: 18.11.2024
KARAR 2024/54	
Üniversitemiz Mühendislik Fakültesi öğretim elemanı Doç. Dr. Gür Emre GÜRAKSIN tarafından yürütülen (Diğer Araştırmacılar: Mehmet ÇAKMAKTEPE), "Çoktan Seçmeli Testlerin Web Ortamında Güvenirlilik Analizi" başlıklı yüksek lisans tezi kapsamında yapılan başvuruda yer alan veri toplama araçlarının, etik açıdan sakıncalı olmadığına, katılanların oy birliği ile karar verildi.	
ASLI GİBİDİR	
Prof. Dr. İbrahim MUTLU Fen ve Mühendislik Bilimleri Bilimsel Araştırma ve Yayın Etiği Kurulu Başkanı	

Bu belge, güvenli elektronik imza ile imzalanmıştır.
Evrak Doğrulaması: <https://turkive.gov.tr/ebd?eK=5381&eD=BSMLMM1NTK&eS=316818> adresinden yapılabilir.