



T.C.
AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANABİLİM DALI

DOKTORA
TEZİ

BOBUT UYGULAMASINDA KULLANICI
DENEYİMİNE ETKİ EDEN DEĞİŞKENLERİN VE
OLASI ETKİLERİNİN BELİRLENMESİ

ALPER TOSUN

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
DOKTORA PROGRAMI

ANTALYA, 2025

T.C.
AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
DOKTORA PROGRAMI

**BOBUT UYGULAMASINDA KULLANICI DENEYİMİNE ETKİ EDEN
DEĞİŞKENLERİN VE OLASI ETKİLERİNİN BELİRLENMESİ**

DOKTORA TEZİ

Alper TOSUN

Danışman: Doç. Dr. Alper SİNAN

Antalya, 2025

T.C.

AKDENİZ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Alper TOSUN'un bu çalışması 17.06.2025 tarihinde jürimiz tarafından **Eğitim Bilimleri** Ana Bilim Dalı **Eğitimde Ölçme ve Değerlendirme** Doktora Programında **Doktora Tezi** olarak **oy birliği/oy çokluğu** ile kabul edilmiştir.

Başkan: Prof. Dr. Bayram BIÇAK

Akdeniz Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü

Üye: Prof. Dr. Hakan KOĞAR

Akdeniz Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü

Üye: Doç. Dr. Eren Can AYBEK

Pamukkale Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü

Üye: Doç. Dr. Murat Doğan ŞAHİN

Anadolu Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü

Üye (Danışman): Doç. Dr. Alper SİNAN

Akdeniz Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü

DOKTORA TEZİNİN ADI: BOBUT Uygulamasında Kullanıcı Deneyimine Etki Eden Değişkenlerin ve Olası Etkilerinin Belirlenmesi

ONAY: Bu tez, Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulunun tarihli ve sayılı kararıyla kabul edilmiştir.

Doç. Dr. Güçlü ŞEKERCİOĞLU

Enstitü Müdürü

DOĞRULUK BEYANI

Yüksek Lisans tezim olarak sunduğum “BOBUT Uygulamasında Kullanıcı Deneyimine Etki Eden Değişkenlerin ve Olası Etkilerinin Belirlenmesi” adlı bu çalışmayı bilimsel, ahlak ve geleneklere aykırı düşecek bir yol ve yardıma başvurmaksızın yazdığımı, yararlandığım eserlerin kaynakçalardan gösterilenlerden oluştuğunu ve bu eserleri her kullanışmada alıntı yaparak yararlandığımı belirtir; bunu onurumla doğrularım. Enstitü tarafından belli bir zamana bağlı olmaksızın, tezimle ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara katlanacağımı bildiririm.

Alper TOSUN

TEŞEKKÜR

İlk olarak Türkiye Cumhuriyeti'nin kurucusu Gazi Mustafa Kemal Atatürk'ü saygı ve minnetle andığımı; onun “Hayatta en hakiki mürşit ilimdir” sözü doğrultusunda yürüdüğüm bu akademik yolculukta rehberliğini her zaman hissettiğimi belirtmek isterim.

Akademik yolculuğumda en büyük rehberim, bilgeliğiyle bana yön veren, sabrıyla destek olan ve her zaman yanımda hissettiren değerli tez danışmanım Doç. Dr. Alper SİNAN'a sonsuz teşekkür ederim. Varlığı bu süreci yalnızca bir akademik yolculuk olmaktan çıkarıp, benim için gerçek bir öğrenme serüvenine dönüştürmüştür. Bu alana gönül vermemde büyük etkisi olan, her zaman yolumu aydınlatan ve içten rehberliğiyle ilham kaynağım olan Doç. Dr. Eren Can AYBEK'e derin bir teşekkür borçluyum. İyi ki yollarımız kesişti.

Tez izleme sürecindeki desteği ve değerli katkılarıyla bana akademik yönden güç katan, aynı zamanda jüri başkanlığını da yürüten Prof. Dr. Bayram BIÇAK'a; kıymetli bilgi ve önerileriyle tezimde iz bırakan değerli jüri üyeleri Prof. Dr. Hakan KOĞAR ve Doç. Dr. Murat Doğan ŞAHİN'e gönülden teşekkür ederim. Lisansüstü eğitim yolculuğumda bana her zaman destek olan, bilgi ve deneyimleriyle yolumu aydınlatan Doç. Dr. Bilal Barış ALKAN ve Doç. Dr. Osman AKHAN'a teşekkür ederim.

Hayatımın her döneminde yanımda olan, iyi günde olduğu kadar en zor zamanlarımda da inancını bir an olsun yitirmeyen canım annem Meryem TOSUN'a, her daim desteğiyle arkamda dimdik duran sevgili babam Rafet TOSUN'a ve her zaman koruyup kollayan kıymetli ağabeyim Ahmet Oğuzhan TOSUN'a minnettarım. Bu başarıda sizin sevginizin ve emeğinizin izi sonsuza dek var olacak.

Bu sürecin en güzel yanı, kalbimde hep özel bir yerde olan sevgili nişanlım Merve ÇÖRTÜK'le aynı hayalleri paylaşmaktı. Varlığı, sevgisi, sabrı ve desteğiyle yolumu aydınlatan; en çaresiz hissettiğim anlarda bana yeniden umut aşılayan bir yoldaş oldu. İyi ki varsın, iyi ki birlikteyiz.

Ve son olarak; bu süreçte yanımda olan, beni sabırla dinleyen, yeri geldiğinde motive eden yeri geldiğinde sadece sessizce yanımda duran tüm dostlarıma... Dinar'daki çocukluk arkadaşlarıma, Denizli'de bu yola çıkarken beni cesaretlendiren dostlarıma ve Antalya'daki yol arkadaşlarıma gönülden teşekkür ederim. Sizlerin varlığı bu yolculuğu daha anlamlı kıldı.

ÖZET

BOBUT UYGULAMASINDA KULLANICI DENEYİMİNE ETKİ EDEN DEĞİŞKENLERİN VE OLASI ETKİLERİNİN BELİRLENMESİ

TOSUN, Alper

Doktora Tezi, Eğitim Bilimleri Ana Bilim Dalı

Tez Danışmanı: Doç. Dr. Alper SİNAN

Haziran 2025, xi + 138 sayfa

Bu araştırma, Bilgisayar Ortamında Bireye Uyarlanmış Testlerde kullanıcı deneyimine etki eden değişkenlerin ve olası etkilerinin belirlenmesi amacıyla yapılmıştır. Bu bağlamda yürütülen çalışma yakınsayan paralel karma yöntem deseni çerçevesinde gerçekleştirilmiştir. Araştırmanın nitel boyutu öğrenci görüşlerinin içerik analizi yöntemiyle incelenmesini içerirken, nicel boyutu tarama araştırması olarak gerçekleştirilmiştir. Ayrıca araştırma BOBUT tasarımlarına ilişkin katkılarıyla Ar-Ge çalışması özelliği göstermektedir. Bu doğrultuda madde havuzunun oluşturulması için eğitimde ölçme ve değerlendirme dersine ilişkin 130 test maddesi yazılmıştır. Alan ve dil uzmanlarının görüşleri doğrultusunda 40 madde havuzdan çıkarılarak bireylere 90 test maddesi uygulanmıştır. Uygulamalar eksik test deseni ve tüm maddelerin uygulanması olmak üzere iki farklı şekilde yapılmıştır. Madde Tepki Kuramı varsayımlarından yerel bağımsızlık ilkesini ihlal eden ve madde uyum indeksi düşük olan 15 madde daha testten çıkarılmıştır. Nihai havuz olarak belirlenen 75 madde için 293 eğitim fakültesi öğrencinden toplanan verilere dayalı post-hoc simülasyonlar yürütülmüştür. Simülasyon sonuçlarına göre MTK modeli olarak Rasch model; durdurma kuralı olarak 10 maddelik standart test uzunluğu; yetenek kestirimi için En Yüksek Sonsal ve başlangıç düzeyi için θ 'nın -0.5 ile +0.5 aralığından rastgele seçilen bir madde ile uygulanmasının en uygun ölçüt yöntem ve değerler olduğuna karar verilmiştir. BOBUT kullanıcı deneyimine etki edeceği düşünülen geribildirim kapsamı ve süre bilgisi değişkenlerine göre dört BOBUT tasarlanmıştır. BOBUT uygulamasına katılan bireyler bir web sitesi aracılığıyla farklı testlere atanmıştır. Uygulamaya 131 eğitim fakültesi öğrencisi katılmış ve uygulama sonrasında BOBUT deneyimine ilişkin görüşleri toplanmıştır. BOBUT uygulaması sonucunda geniş kapsamlı geri bildirimde sahip testleri alan bireylerin teste olan güvenlerinin ve motivasyonlarının daha yüksek olduğu görülmüştür. Süre bilgisini görmeyen

bireylerin görüşleri farklılık gösterse de bireyler genellikle stres düzeyinin daha düşük olduğunu söylemişlerdir. Ayrıca bireylerin BOBUT uygulamasına ve sistemine yönelik görüşleri incelenmiştir. Sonuç olarak geniş kapsamlı ve anlaşılır bir şekilde verilen geribildirim kullanıcının deneyimini iyileştireceği düşünülmektedir.

Anahtar Kelimeler: BOBUT, Kullanıcı Deneyimi, Geribildirim, Süre



ABSTRACT

DETERMINATION OF VARIABLES AFFECTING USER EXPERIENCE IN CAT APPLICATION AND THEIR POSSIBLE EFFECTS

Tosun, Alper

Ph.D. Thesis, Department of Educational Sciences

Supervisor: Assoc. Prof. Alper SİNAN

June 2025, xi + 138 pages

This study was conducted to determine the variables affecting user experience in computer adaptive tests and their possible effects. The study was conducted within the framework of a convergent parallel mixed methods design. The qualitative dimension of the study involved examining student opinions using content analysis methods, while the quantitative dimension was conducted as a survey study. In addition, the study demonstrates its R&D characteristics with its contributions to CAT designs. In this direction, 130 test items related to the measurement and evaluation course in education were written to create the item pool. Based on the opinions of subject and language experts, 40 items were removed from the pool, and 90 items were administered to individuals. The applications were conducted in two different ways: an incomplete test pattern and the application of all items. Fifteen items that violated the principle of local independence, one of the assumptions of Item Response Theory, and had a low item fit index were also removed from the test. Post-hoc simulations were conducted based on data collected from 293 education faculty students for the final pool of 75 items. According to the simulation results, the Rasch model was determined to be the most appropriate measurement method and value for the IRT model, with a standard test length of 10 items as the stopping rule; Maximum a Posteriori for ability estimation and a randomly selected item from the range of -0.5 to +0.5 for the initial level. Four CAT's were designed based on the scope and duration of feedback variables that were thought to affect the CAT user experience. Individuals participating in the CAT application were assigned to different tests via a website. A total of 131 education faculty students participated in the application, and their opinions regarding the CAT experience were collected after the application. As a result of the CAT application, it was observed that individuals who took tests with extensive feedback had higher confidence and motivation in the test. Although the opinions of individuals who did not see the time information varied, individuals generally

reported lower stress levels. In addition, individuals' opinions about the CAT application and system were examined. In conclusion, it was observed that comprehensive and clear feedback improves the user experience.

Keywords: CAT, User Experience, Feedback, Time



İÇİNDEKİLER

ÖZET.....	i
ABSTRACT	iii
İÇİNDEKİLER	v
TABLolar LİSTESİ.....	viii
ŞEKİLLER LİSTESİ.....	ix
KISALTMALAR LİSTESİ	x

BÖLÜM I

GİRİŞ

1.1. Problem Durumu	2
1.2.Araştırmanın Amacı ve Alt Problemler	3
1.3. Araştırmanın Önemi	4
1.4. Varsayımlar	5
1.5. Sınırlılıklar	5
1.6. Tanımlar	6

BÖLÜM II

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

2.1. Bilgisayar Ortamında Bireye Uyarlanmış Testler	7
2.1.1. Bilgisayar Ortamında Testlerin Gelişimi.....	7
2.1.2. BOBUT Uygulamasının İşleyişi ve Avantajları	8
2.1.1.1. Ölçme Kesinliği	10
2.1.1.2. Zaman ve Madde Tasarrufu	12
2.1.1.3. Kullanıcı Deneyimi	14
2.1.1.4. Test Güvenilirliği ve Geçerliliği	15
2.1.3. BOBUT'un Eğitimde Kullanım Alanları	17
2.1.4. BOBUT'un Uygulanabilirliği ve Gelecek Yönelimler	18
2.1.4.1. BOBUT Uygulamalarında Karşılaşılan Sorunlar	19
2.1.4.2. BOBUT Uygulamalarında Güncel Araştırma Alanları.....	20
2.1.4.2.1. Yapay Zeka Destekli Bireyselleştirilmiş Sistemler	21
2.1.4.2.2. Kullanıcı Deneyimi Odaklı Test Yenilikleri.....	22

2.2. Madde Tepki Kuramı (MTK).....	22
2.3. BOBUT'un Bileşenleri	33
2.3.1. İlk Madde Seçimi	34
2.3.2. Yetenek Kestirimi	34
2.3.2.1. En Çok Olabilirlik Yöntemi(Maximum Likelihood Estimation- MLE)	35
2.3.2.2. Bayesci Yöntemler.....	35
2.3.2.2.1. Maximum a Priori (MAP).....	36
2.3.2.2.2. Expected a Posteriori (EAP)	36
2.3.3. Madde Seçme Yöntemleri	36
2.3.3.1. Maksimum Fisher Bilgi(MFB).....	37
2.3.3.2. Bayesci Yöntemler.....	37
2.3.3.3. Aralık Bilgi Ölçütü(ABÖ).....	37
2.3.4. Test Sonlandırma Kuralı.....	38
2.3.4.1. Sabit Test Uzunluğu Kuralı	38
2.3.4.2. Değişken Test Uzunluğu Yöntemleri.....	39
2.3.4.2.1. Standart Hata Temelli Test Sonlandırma	39
2.3.4.2.2. Kesme Puanı Temelli Test Sonlandırma	40
2.4. Kullanıcı Deneyimi ve Eğitim Teknolojisi	40
2.4.1. Kullanıcı Deneyimi Kavramı.....	40
2.4.2. Eğitim Teknolojilerinde Kullanıcı Deneyiminin Önemi	42
2.5. Geribildirim ve Test Tasarımı	44
2.5.1. Geribildirim Türleri.....	45
2.6. İlgili Araştırmalar	47

BÖLÜM III

YÖNTEM

3.1. Araştırmanın Modeli.....	50
3.2. Çalışma Grubu.....	51
3.3. Veri Toplama Araçları	52
3.3.1. Madde Havuzuna İlişkin Uzman Görüşü	52
3.3.2. Eğitimde Ölçme ve Değerlendirme Madde Havuzunun Oluşturulması ve Veri Toplama Aracına İlişkin Eksik Test Deseni.....	53
3.3.3. Concerto ve BOBUT Uygulama Formlarının Özellikleri	54

3.3.4. Kullanıcı Deneyimine İlişkin Görüş Formu	58
3.3.5. Verilerin Toplanması Sürecinde Rastgelelik	59
3.4. Verilerin Toplanması	60
3.4.1. Kâğıt Kalem Uygulaması	61
3.4.2. Canlı BOBUT Uygulaması.....	61
3.5. Verilerin Analizi.....	62
3.5.1. Model Varsayımlarının Sınanması.....	62
3.5.2. Post-hoc Simülasyon.....	64
3.5.3. BOBUT Kullanıcı Deneyimine İlişkin Görüşlerin İncelenmesi	65

BÖLÜM IV

BULGULAR

4.1. Madde Parametreleri.....	70
4.2. MKE ve Madde Bilgi Fonksiyonu	74
4.3. Madde Havuzu Güvenirliği ve Hatası	76
4.4. Post-hoc Sonuçları.....	78
4.5. BOBUT Uygulamasına İlişkin Betimsel İstatistikler	79
4.6. BOBUT Kullanıcı Deneyimi Görüşme Formuna İlişkin Bulgular	80
4.7. BOBUT Kullanıcı Deneyimi Anketine İlişkin Bulgular	92

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

5.1. Sonuç ve Tartışma	96
5.2. Öneriler	108
5.2.1. Araştırmacılara Öneriler	108
5.2.2. Uygulayıcılara Öneriler	109

KAYNAKÇA.....	111
EK-1. ETİK KURUL VE MEB ARAŞTIRMA İZİN BELGELERİ.....	134
ÖZ GEÇMİŞ	135
BENZERLİK RAPORU	137
BİLDİRİM.....	138

TABLÖLAR LİSTESİ

Tablo 3.1. Madde Havuzunun Oluşturulmasındaki Çalışma Grubuna İlişkin Demografik Özellikler	51
Tablo 3.2. BOBUT Uygulamasında Kullanıcı Deneyiminin Araştırılmasına Yönelik Çalışma Grubuna İlişkin Demografik Özellikler	52
Tablo 3.3. Test Kitapçıklarına İlişkin Kapsam Geçerliliği.....	53
Tablo 3.4. BOBUT Uygulamalarına İlişkin Değişkenler	58
Tablo 3.5. Yerel Bağımsızlık Varsayımını Karşılamayan Maddelere İlişkin Yen'in Q3'ü	62
Tablo 3.6. Maddelere İlişkin Uyum İndeksleri.....	63
Tablo 3.7. Açık Uçlu Sorulara Yönelik Cevap Temaları	66
Tablo 3.8. Açık Uçlu Soruların Fleiss Kappa Katsayısına İlişkin Bulguları.....	67
Tablo 3.9. BOBUT Kullanıcı Deneyimi Anketine İlişkin Betimsel İstatistikler	68
Tablo 4.1. Temel Kavramlar Konusuna İlişkin Maddelere Ait Madde Parametreleri.....	71
Tablo 4.2. Ölçme Araçlarının Nitelikleri Konusuna İlişkin Maddelere Ait Madde Parametreleri.....	72
Tablo 4.3. Ölçme Araçları Konusuna İlişkin Maddelere Ait Madde Parametreleri	73
Tablo 4.4. Madde ve Test İstatistiği Konusuna İlişkin Maddelere Ait Madde Parametreleri .	73
Tablo 4.5. Post-Hoc Simülasyon Sonuçları.....	78
Tablo 4.6. Canlı BOBUT Sonuçlarına İlişkin Betimsel İstatistikler.....	79
Tablo 4.7. BOBUT Kullanıcı Deneyimi Motivasyonu Düzeyine İlişkin Ki Kare.....	80
Tablo 4.8. BOBUT Kullanıcı Deneyimi Stres Düzeyine İlişkin Ki Kare	83
Tablo 4.9. BOBUT Kullanıcı Deneyimi Soru Ekranı Öğelerine İlişkin Ki Kare	85
Tablo 4.10. BOBUT Kullanıcı Deneyimi Sistem Önerilerine İlişkin Ki Kare.....	88
Tablo 4.11. BOBUT Kullanıcı Deneyimi Geribildirim Ekranına İlişkin Ki Kare.....	89
Tablo 4.12. BOBUT Kullanıcı Deneyimi Memnuniyet Düzeyine İlişkin Kruskal Wallis-H Bulguları.....	93
Tablo 4.13. BOBUT Kullanıcı Deneyimi Geribildirim Doğruluğuna İlişkin Kruskal Wallis-H Bulguları.....	93
Tablo 4.14. BOBUT Kullanıcı Deneyimi Geribildirim Güvenine İlişkin Kruskal Wallis-H Bulguları.....	94
Tablo 4.15. BOBUT Kullanıcı Deneyimi Geribildirim Çaba ve Motivasyona Etkisine İlişkin Kruskal Wallis-H Bulguları.....	94
Tablo 4.16. Testleri BOBUT Olarak Alma İsteğine İlişkin Kruskal Wallis-H Bulguları.....	95

ŞEKİLLER LİSTESİ

Şekil 2.1. BOBUT Akış Şeması(Davey, 2011).	9
Şekil 2.2. 1PLM İçin MKE'ler	30
Şekil 2.3. 2PLM İçin MKE'ler	31
Şekil 2.4. 3PLM İçin MKE'ler	32
Şekil 2.5. 4PLM İçin MKE'ler	33
Şekil 3.1. BOBUT Karşılama Ekranı	55
Şekil 3.2. Sürenin Gösterildiği Yanıt Verilmiş Madde.....	55
Şekil 3.3. Sürenin Gösterilmediği Yanıt Verilmiş Madde	56
Şekil 3.4. Geniş Geribildirim Sayfası	56
Şekil 3.5. Dar Geribildirim Sayfası	57
Şekil 3.6. Rastgeleliğin Sağlanması İçin Oluşturulan Web Sitesi Karşılama Ekranı.....	60
Şekil 4.1. Maddelere Ait Güçlük Parametrelerinin Dağılımı.....	70
Şekil 4.2. Madde Havuzundaki Maddelere İlişkin Madde Karakteristik Eğrileri.....	74
Şekil 4.3. Madde Havuzundaki Maddelere İlişkin Madde Bilgi Fonksiyonları.....	75
Şekil 4.4. Madde Havuzuna İlişkin Test Bilgi Fonksiyonu ve Standart Hata.....	76
Şekil 4.5. Madde Havuzuna İlişkin Güvenirlilik.....	77

KISALTMALAR LİSTESİ

1PLM	: 1 Parametrelı Lojistik Model
2PLM	: 2 Parametrelı Lojistik Model
3PLM	: 3 Parametrelı Lojistik Model
4PLM	: 4 Parametrelı Lojistik Model
ABÖ	: Aralık Bilgi Ölçütü
ALES	: Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı
ASVAB	: Armed Services Vocational Aptitude Battery
BDT	: Bilgisayar Destekli Test
BOBUT	: Bilgisayar Ortamında Bireye Uyarlanmış Test
BTT	: Bilgisayar Tabanlı Test
CITO	: Centraal Instituut voor Toetsontwikkeling
EAP	: Beklenen A Posteriori Tahmini (Expected A Posteriori Estimation)
GMAT	: Graduate Management Admission Test
GRE	: Graduate Record Examination
IELTS	: International English Language Testing System
KL	: Kullback-Leibler (Kullback-Leibler Divergence)
KTK	: Klasik Test Kuramı
MAIS	: Minenesota Adaptive Instruuctional System
MAP (Growth)	: Measures of Academic Progress
MAP	: Maksimum A Posteriori Tahmini (Maximum A Posteriori Estimation)
MFB	: Maksimum Fisher Bilgisi (Maximum Fisher Information)
MKE	: Madde Karakteristik Eğrisi
MLE	: Maksimum Olabilirlik Tahmini (Maximum Likelihood Estimation)
MST	: Çok Aşamalı Test (Multi Stage Testing)
MTK	: Madde Tepki Kuramı
NCLEX	: National Council Licensure Examination
PISA	: Uluslararası Öğrenci Değerlendirme Programı (Programme for International Student Assessment)
RIT	: Rasch Unit
RMSE	: Kök Ortalama Kare Hatası (Root Mean Square Error)

TIMMS	: Uluslararası Matematik ve Fen Eğilimleri Araştırması (Trends in International Mathematics and Science Study)
TOEFL	: Test of English as a Foreign Language
UX	: User Experience
YDS	: Yabancı Dil Bilgisi Seviye Tespit Sınavı



BÖLÜM I

GİRİŞ

Test uygulama yaklaşımlarının bilgisayar tabanlı olarak uygulanması çağdaş ölçme ve değerlendirme süreçleri açısından önemli bir gelişmedir. Bilgisayar biliminin gelişmesi ve eğitim sürecine dahil edilmesinden önce ortaya atılan ve hala kullanılmakta olan klasik test kuramına (KTK) dayalı ölçme süreçleri, tüm bireylerin aynı test maddelerini alması ile yürütülmektedir. Ancak bu yöntem ölçme güvenirliği ile ilgili sınırlılıklar oluşturmakta ve bireysel farklılıkları göz ardı etmektedir (Çelen & Aybek, 2013). 1970'lerde madde tepki kuramının (MTK) ortaya çıkışı ile bireylerin yetenek düzeylerine göre uygulanacak maddelerin seçilebilmesine olanak sağlayan bireyselleştirilmiş testler birçok bilim dalı için ölçme değerlendirme yaklaşımlarında köklü bir dönüşüm başlatmıştır (Hambleton vd., 1991). Bireye uyarlanmış testlerde MTK uygulamalarının yaygınlaşmasıyla birlikte testlerin hem geçerlilik hem de güvenirlik açısından daha etkin bir hale getirilebilmesi mümkün olmuştur. Bu gelişmelerin bilgisayar ortamında uygulanması, bilgisayar tabanlı test (BTT) ve daha sonraki süreçte bilgisayar ortamında bireye uyarlanmış testler (BOBUT) olarak karşımıza çıkmıştır (van der Linden & Hambleton, 1997).

BOBUT uygulamalarının yaygınlaşması test uygulamalarında geleneksel yaklaşımlardan uyarlanabilir testlere geçilebilmesine imkân sağlayan önemli bir adımdır (Kezer & Koç, 2009). Uygulama, maddelerin güçlük parametreleri ile bireylerin yetenek düzeyleri arasındaki dinamik uyum ile çalışmaktadır. Böylece yetenek düzeyi bireye uyarlanmış bir test yardımıyla ölçülebilmektedir. Ayrıca bu durum bireye uygun olmayan maddelerin uygulanmasını önleyerek testin süresini kısaltmaktadır. BOBUT, bireye belirlenen ilk maddenin uygulanmasıyla başlar ve yanıtı göre yetenek kestirimi yapılır. Daha sonra kestirilen yetenek düzeyine göre bireye en uygun maddenin uygulanması sağlanarak süreç devam eder. Bu iteratif süreç testin durdurma kriterleri sağlanana kadar devam eder. Belirlenen kriter sağlandığında test sonlandırılır ve değerlendirme adımına geçilir (Chang & Ying, 1996; Kingsbury & Weiss, 1983; Van Der Linden & Glas, 2010).

Bireylerin yetenek düzeyinin bilgisayar temelli bir şekilde bireyselleştirilmiş olarak ölçülmesinin test geliştirme sürecinde birçok avantajı vardır. BOBUT'un MTK temelli bir uygulama olması, bireylerin yetenek düzeyine en uygun maddelerin seçilerek verilmesine imkân tanır ve bu durum ölçme hatalarını minimuma indirerek daha hassas sonuçlar elde

edilmesini sağlar. Bu durum testten elde edilen puanların geçerlilik ve güvenilirliğinin artmasını sağlamaktadır (Hambleton vd., 1991; van der Linden, 2016). Öte yandan sabit uzunluğa sahip testlerde bireylerin tüm test maddelerine cevap vermesi beklenirken, BOBUT uygulamasında bireylerin kendi yetenek düzeyine uygun olarak seçilmiş maddeleri cevaplaması beklenmektedir. BOBUT'un bireylerin test performansına göre şekillenen ve sonlandırılan bir yapıya sahip olması test süresinde %30 ila %50 oranında bir kısaltma yapılmasına imkân sağlamaktadır (Eroğlu & Kelecioğlu, 2013). Bu yapı sayesinde hem test süresi kısaltılır hem de kullanılan madde sayısı bireye göre belirlenmiş olur (Kezer & Koç, 2009). Diğer yandan testin BOBUT olarak uygulanması kullanıcı deneyimini iyileştirmektedir. BOBUT uygulamalarının başarısı büyük ölçüde kullanıcı deneyimine bağlıdır. Bireylerin yetenek düzeyine göre seçilen maddelerin uygulanmasıyla bireylerin motivasyonunun korunması amaçlanır ve daha adil bir deneyim sunulmuş olur (Gasperson, 2014; Kezer & Koç, 2009). Ayrıca BOBUT uygulamaları test deneyiminin iyileştirilebilmesi adına çeşitli düzenlemeler yapılabilmesine imkân sağlamaktadır. Bu bağlamda bireylere anında geribildirimlerin verilebilmesi ve sürenin gösterilmesine ilişkin tercih sunabilmesi gibi düzenlemelerle kullanıcı deneyimi iyileştirebilir.

Geribildirim uygulamaları, bireylere sadece test sonuçlarının verilmesini değil aynı zamanda bireyin test performansı üzerinden değerlendirme yapılmasını ve öğrenme sürecini planlamasını sağlayan çalışmalardır (Hattie & Timperley, 2007). Ayrıca bireylerin aldıkları pozitif geribildirimlerin motivasyonu arttırdığı, öğrenme kaygısını düşürdüğü ve bireyin performansına ilişkin biçimlendirici bir rolü olduğu bilinmektedir (Burgers vd., 2015; Guo & Wei, 2019; Turan vd., 2011). Bunun yanında bireylerin bilgisayar ortamında uygulanan testlerdeki süre bilgisine ilişkin duygu ve düşünceleri kullanıcı deneyimi açısından son derece önemlidir (Hott vd., 2023; Truslow vd., 2022).

1.1. Problem Durumu

Bilgisayar Ortamında Bireye Uyarlanmış Testler (BOBUT), bireylerin yetenek düzeyine göre dinamik biçimde madde seçimi yapan ve ölçme süreçlerini bireyselleştiren modern bir değerlendirme yaklaşımıdır. BOBUT geleneksel sabit testlerle karşılaştırıldığında test süresini kısaltmakta, ölçme kesinliğini artırmakta ve testin psikometrik niteliklerini iyileştirmektedir (Akhtar & Kovács, 2024). Bu özellikleri sayesinde BOBUT, özellikle geniş

ölçekli test uygulamaları ve çevrim içi öğrenme ortamlarında yaygın biçimde kullanılmaya başlanmıştır (van Wijk vd., 2024).

Ancak testlerin teknik avantajlarına rağmen BOBUT uygulamalarının başarısı yalnızca ölçme doğruluğu ile sınırlı değildir; bireylerin test deneyimi süresince yaşadığı psikolojik ve bilişsel durumlar da bu sistemlerin etkililiğini doğrudan etkilemektedir (Barrett vd., 2021). Özellikle kullanıcıların motivasyonu, memnuniyeti, stres düzeyi ve testin adil olduğu yönündeki algıları test sürecine yönelik tutumlarını şekillendirmekte, bu durum ise testin geçerliği üzerinde dolaylı etkiler yaratabilmektedir (Germanakos vd., 2020).

Literatürde BOBUT uygulamalarının psikometrik üstünlüklerine ilişkin çok sayıda çalışma bulunmasına rağmen kullanıcı deneyimi boyutunu derinlemesine ele alan araştırmalar oldukça sınırlıdır. Örneğin; test sırasında verilen geri bildirimlerin kapsamı, test süresinin kullanıcıya gösterilip gösterilmemesi ve arayüz tasarımı gibi unsurların stres, dikkat, öğrenme tutumu ve memnuniyet üzerindeki etkileri yeterince sistematik biçimde incelenmemiştir (Cheng, 2023). Test tasarımında bu tür faktörler çoğu zaman sezgisel kararlarla belirlenmekte ve bilimsel verilere dayalı olarak yapılandırılmamaktadır.

Ayrıca kullanıcı deneyimine dair bireysel farklılıkların dikkate alınmaması özellikle düşük yetenek düzeyine sahip bireylerde motivasyon kaybı ve kaygı artışı, yüksek yetenek düzeyindekilerde ise sıkılma ve dikkatsizlik gibi olumsuz sonuçlara yol açabilmektedir (Akhtar & Kovács, 2024). Bu durum, sistemin bireyselleştirilmiş doğasıyla çelişmekte ve öğrenme süreçleriyle bütünleşmesini zorlaştırmaktadır.

Dolayısıyla bu araştırma, literatürdeki şu temel boşluğa odaklanmaktadır: BOBUT sistemlerinin kullanıcı deneyimine ilişkin tasarım bileşenlerinin (geri bildirim yapısı, zaman göstergesi, arayüz düzeni vb.) bireylerin psikolojik tepkileri, motivasyonları ve memnuniyetleri üzerindeki etkilerini ampirik olarak değerlendiren kapsamlı çalışmaların eksikliği. Bu eksiklik, hem yeni nesil test sistemlerinin pedagojik etkililiği hem de kullanıcı merkezli tasarım ilkelerinin geliştirilmesi açısından önemli bir araştırma alanını temsil etmektedir.

1.2.Araştırmanın Amacı ve Alt Problemler

Bu araştırmanın amacı BOBUT uygulamalarında kullanıcı deneyimine etki edebilecek değişkenlerden geri bildirim ve süre gösterimi gibi öğelerin olası etkilerini belirlemektir. Bu amaç doğrultusunda araştırma sürecinde aşağıdaki sorulara yanıt aranmıştır:

1. BOBUT uygulamalarında geri bildirimin kapsamı, öğrenci memnuniyeti, motivasyonu ve öğrenme kazanımları üzerinde nasıl bir etki yaratmaktadır?
2. BOBUT uygulamalarında soru ekranında yer alan süre göstergesinin öğrencilerin zaman yönetimi davranışları ile olumsuz duygu ve deneyimleri üzerindeki etkileri nelerdir?
3. Geri bildirimin kapsamı ile süre bilgisinin etkileşimi BOBUT kullanıcı deneyimini nasıl etkilemektedir?
4. Öğrencilerin geri bildirim sayfasına yönelik beklentileri ve görmek istedikleri içerikler nelerdir?

1.3. Araştırmanın Önemi

Ölçme ve değerlendirme uygulamaları, eğitim sistemlerinin temel yapı taşlarından biridir. Özellikle dijitalleşmenin ivme kazandığı 21. yüzyılda bu uygulamalar hızla dönüşmekte ve teknolojik olanaklarla yeniden şekillenmektedir. Bu bağlamda Bilgisayar Ortamında Bireye Uyarlanmış Testler (BOBUT), bireyselleştirilmiş değerlendirme fırsatları sunarak ölçme alanında önemli bir paradigma değişimini temsil etmektedir. BOBUT'un psikometrik üstünlükleri arasında test süresinin kısalması, madde seçiminde verimliliğin artması ve yetenek düzeyine uygun madde sunumu yer almaktadır (van der Linden & Glas, 2000; Weiss & Şahin, 2024). Bu yönüyle test uygulamaları daha etkili, adil ve birey odaklı hâle gelmektedir.

Ancak literatürde BOBUT'un teknik altyapısı ve madde seçme algoritmalarına yönelik çalışmalar geniş yer tutmasına rağmen bu sistemlerin kullanıcı deneyimi üzerindeki etkileri yeterince incelenmemiştir (Wainer, 2000; Sands vd., 1997). Oysa bir testin başarısı yalnızca ölçme doğruluğu ile değil bireylerin test sürecine ilişkin duygusal, bilişsel ve davranışsal tepkileriyle de yakından ilişkilidir. Bireylerin psikolojik kabulü, motivasyonu ve memnuniyeti gibi değişkenler testin güvenilirliği ve geçerliğini dolaylı olarak etkileyebilmektedir. Özellikle pandemi sonrası yaygınlaşan çevrim içi test uygulamaları kullanıcı deneyiminin test performansı üzerindeki etkisini daha görünür kılmıştır (Cheng, 2023).

Bu noktada test sürecinde kullanıcıya sunulan geri bildirimlerin kapsamı, arayüzde süre göstergesinin yer alıp almaması ve genel kullanıcı arayüzünün tasarımı gibi unsurlar; bireylerin stres düzeyi, dikkat yönetimi ve öğrenme tutumları üzerinde belirleyici etkilere sahiptir (Alipour vd., 2023; Miraz vd., 2021; Wang vd., 2023). Ne var ki BOBUT

sistemlerinde bu tür kullanıcı etkileşimi değişkenlerine ilişkin ampirik araştırmalar oldukça sınırlıdır. Mevcut literatürde bu faktörlerin çoğu test geliştiricilerin sezgisel tercihlerine dayalı olarak belirlenmekte; bilimsel temelli yapılandırılmış tasarım modellerinden yoksun kalmaktadır (van der Linden vd., 2010; Weiss & Şahin, 2024).

Bu araştırma söz konusu literatür boşluğunu doldurarak BOBUT sistemlerinde kullanıcı deneyimini etkileyen arayüz tasarımı, geri bildirim yapısı ve süre göstergesi gibi değişkenlerin birey üzerindeki etkilerini sistematik biçimde ortaya koymayı amaçlamaktadır. Bu kapsamda elde edilecek bulgular, yalnızca psikometrik değil pedagojik açıdan da etkili test tasarımlarının oluşturulmasına katkı sağlayacaktır. Aynı zamanda kullanıcı merkezli tasarım anlayışının BOBUT sistemlerine entegrasyonu için uygulayıcılara ve politika yapıcılara bilimsel temelli öneriler sunarak bireylerin test sürecine yönelik olumlu deneyimler geliştirmelerine ve sistemin sürdürülebilirliğine katkı sağlayacaktır.

1.4. Varsayımlar

- Katılımcıların, BOBUT kullanıcı deneyimine ilişkin anket formunu samimi ve dikkatli bir şekilde yanıtladıkları varsayılmıştır.
- Her bir öğrencinin birbirinden bağımsız olarak ve eşit olasılıkla farklı koşullara sahip dört testten birini aldığı; buna bağlı olarak gruplara yansız atama yapıldığı ve grupların homojen olduğu varsayılmıştır.
- Geri bildirim ve süre göstergesi dışındaki arayüz bileşenlerinin her koşulda sabit tutulduğu ve ortaya çıkan deneyim farklılıklarının yalnızca manipüle edilen iki değişkene (geri bildirim ve süre göstergesi) bağlı olduğu varsayılmıştır.
- Araştırma örneklemini oluşturan bireylerin testi bilgisayar veya telefon gibi teknolojik araçlar üzerinden alırken herhangi bir kaygı ya da stres yaşamadıkları varsayılmıştır.

1.5. Sınırlılıklar

- Araştırmanın çalışma grubu uygulanan testin ölçme ve değerlendirme dersine ilişkin başarıyı ölçmesi nedeniyle belirli bir yaş grubu ve eğitim düzeyindeki bireylerle sınırlandırılmıştır.

- BOBUT uygulamasına katılan bireylerin birçoğunun sisteme yönelik aşinalığının yetersiz olması değerlendirmelerinde algısal yanlılıklara yol açma potansiyeline sahiptir.
- Test, BOBUT'un kullanıcı deneyimini araştırmak amacıyla oluşturulduğu için daha dar kapsamlı tutulmuş ve madde sayısı sınırlı olarak belirlenmiştir.
- Geri bildirimler, testin geneline yönelik verilmiş; konuya özel ayrıntılı geri bildirim sağlanmamıştır.

1.6. Tanımlar

BOBUT (Bilgisayar Ortamında Bireye Uyarlanmış Testler): Test maddelerinin bireyin önceki yanıtlarına göre gerçek zamanlı olarak seçildiği ve her adımda yetenek düzeyi kestiriminin güncellendiği bireyselleştirilmiş bir değerlendirme sistemidir.

Kullanıcı Deneyimi (UX): Bireylerin bir sistemle etkileşimleri sırasında yaşadıkları genel memnuniyet, kolaylık, verimlilik ve duygusal tepkilerin bütünüdür.

Geri bildirim: Kullanıcının yaptığı işlem veya verdiği yanıt karşılık sistemin sunduğu bilgilendirici tepkidir ve öğrenme ile etkileşimi desteklemeyi amaçlar.

BÖLÜM II

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

2.1. Bilgisayar Ortamında Bireye Uyarlanmış Testler

2.1.1. Bilgisayar Ortamında Testlerin Gelişimi

Bilgisayar teknolojilerinin eğitimde ölçme ve değerlendirme süreçlerine entegrasyonu, 1950'li yıllarda radyasyon, mühendislik ve psikoloji alanlarında geliştirilen ilk bilgisayar destekli öğretim ve test uygulamalarıyla başlamıştır. 1954 yılında IBM 701 gibi erken dönem bilgisayarlar basit zekâ testleri ve sezgisel adaptasyon görevlerinde kullanılmış; bu deneyler, testlerin bilgisayar ortamına taşınabilirliğine dair ilk verileri sunmuştur (Weiss & Şahin, 2024).

1960'larda optik okuyucuların yaygınlaşması kâğıt-kalem sınavların otomatik olarak puanlanmasını mümkün kılmış ve doğruluk ile hız açısından önemli ilerlemeler sağlamıştır (Landahl, 2024). Aynı dönemde PLATO gibi ders odaklı bilgisayar sistemleri sınav modülleri sunmaya başlamış; bu modüller hâlen sabit formda olup tüm öğrencilere aynı içeriği sunan bilgisayar tabanlı test (BTT) formatında çalışmıştır (van der Linden & Glas, 2000).

1970'lerin başında Cronbach ve Gleser'in geliştirdiği iki aşamalı test formatı, sabit form BTT'nin sınırlılıklarını aşmak amacıyla ilk uyarlanabilir yapıları ortaya koymuştur (Cronbach & Gleser, 1957). Betz ve Weiss (1973) bu yaklaşımları bilgisayar ortamında simüle ederek madde kümeleri arasında geçiş stratejilerinin test verimliliğini nasıl artırabileceğini göstermiştir (Sands vd., 1997).

Madde Tepki Kuramı'nın (MTK) 1960 ve 1970'lerdeki matematiksel gelişimi, BOBUT uygulamalarının psikometrik güvenilirliğini sağlamıştır (Birnbaum, 1968; Weiss, 2024). Lord'un yukarı-aşağı (up-and-down) yöntemleri ve Owen'ın Bayes-îçi kestirim prosedürleri gerçek zamanlı parametre güncellemeleri için önemli katkılar sunmuştur (Lord, 1968; Owen, 1975).

1975–1985 yılları arasında Minnesota Üniversitesi'nde Larkin ve Weiss'in yürüttüğü fleksilevel BOBUT simülasyonları, farklı zorluk düzeylerine sahip madde bloklarının çapraz kullanımını test ederek klasik sabit form testlere kıyasla %30'a varan test süresi kısalması ve ölçme duyarlılığı artışı sağlamıştır (Larkin & Weiss, 1975; Weiss & Şahin, 2024).

1980’lerde ABD Deniz Kuvvetleri’ne (Marine Corps) bağı ASVAB projesi, BOBUT uygulamalarını ilk kez operasyonel düzeye taşımış; geniş madde havuzları, gerçek zamanlı IRT parametre kestirimi ve maksimum bilgiye dayalı madde seçimi algoritmaları, saha uygulamalarında başarılı sonuçlar vermiştir (Sands vd., 1997).

1994 yılında NCLEX sınavında bilgisayarlı uyarlanabilir testin (CAT) kullanılmaya başlanması BOBUT sistemlerinin yaygınlaşmasında bir dönüm noktası olmuştur. Bu uygulama ile test sonlandırma kuralları ve dinamik madde seçimi yaklaşımları, bireylerin yetenek düzeyine hızlı biçimde uyum sağlayan testlerin gelişimini hızlandırmıştır (van der Linden & Glas, 2010).

1990’ların ortasında ETS tarafından GRE ve GMAT sınavlarında geliştirilen çok aşamalı uyarlanabilir test (MST) formatları ise içerik dengeleme ve katsayı tabanlı madde seçimi prosedürleriyle geçerlik ve güvenirlik düzeyini daha da artırmıştır (Thissen & Mislevy, 2000; Weiss & Kingsbury, 1984).

2000’li yıllarda MAP Growth gibi K–12 düzeyine yönelik ara (interim) değerlendirme sistemleri, Rasch temelli RIT ölçeğini kullanarak milyonlarca öğrenci için yıllık izleme testleri sunmuştur. Bu platformlar, hızlı sonuç üretimi ve veri görselleştirme özellikleri sayesinde öğretimsel karar destek sistemlerinin temelini oluşturmuştur (Stocking & Lewis, 1995; Oladele & Ndlovu, 2021).

Günümüzde BOBUT sistemleri, internet ve bulut tabanlı altyapılar aracılığıyla çok boyutlu adaptasyon modellerini desteklemekte; öğrenme analitiği ve bireysel geri bildirim entegrasyonlarıyla, eğitimde kişiselleştirilmiş öğrenme yol haritalarının oluşturulmasına olanak tanımaktadır (Chang & Ying, 1996; van der Linden & Glas, 2000). Bu tarihsel evrim test süreçlerinde hız, geçerlik ve güvenirlik kazanımlarını optimize ederken, eğitimdeki ölçme kültürünü de kökten dönüştürmüştür.

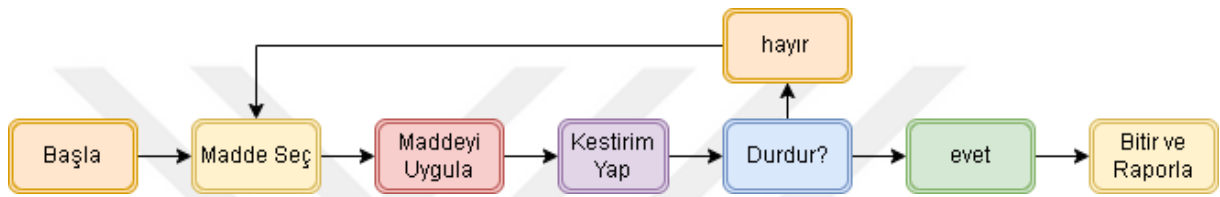
2.1.2. BOBUT Uygulamasının İşleyişi ve Avantajları

BOBUT sisteminin işleyişi, bireyin her maddeye verdiği cevaptan sonra yetenek düzeyinin yeniden kestirilmesi ve bu kestirime göre en fazla bilgi sağlayacak yeni maddenin seçilmesi esasına dayanır (Hambleton vd., 1991; Parshall vd., 2002). Seçilen her madde, birey hakkında maksimum bilgi sağlayacak şekilde belirlendiği için ölçme duyarlılığı artmakta ve test bireye özgü bir şekilde ilerlemektedir (Thissen & Mislevy, 2000; Wainer vd., 2000). Testin başlangıcından bitimine kadar süren bu yinelenen (iteratif) süreç, birey hakkında

yeterli bilgi elde edilene kadar devam eder (Luecht & Nungester, 1998; van der Linden & Pashley, 2009).

Bu sürecin ne zaman sonlanacağı, önceden belirlenmiş durdurma kurallarına göre belirlenir. Yaygın kullanılan durdurma kurallarından biri bireyin yetenek düzeyine ilişkin kestirimin belirli bir standart hata eşiğinin altına düşmesidir. Böylece bireyin yetenek düzeyine dair yeterli bilgi sağlanmış olur (Lewis & Sheehan, 1990; van der Linden & Veldkamp, 2004). Alternatif olarak sabit madde sayısı, zaman sınırı ya da başarı temelli durdurma kuralları gibi yapısal seçenekler de kullanılabilir (Chang & Zhang, 2002).

BOBUT'un işleyiş dinamikleri Şekil 2.1'de görsel olarak sunulmuştur (Davey, 2011).



Şekil 2.1. BOBUT Akış Şeması

Kaynak: Davey, 2011

Şekil 2.1 incelendiğinde testin ilk aşamasında bireye başlangıç maddesinin uygulandığı ardından bireyin yanıtına göre yetenek düzeyinin kestirildiği görülmektedir. Bu kestirime göre en fazla bilgi sağlayacak bir sonraki madde seçilmekte ve süreç bireyin her cevabından sonra yeniden yetenek düzeyi kestirimi ile devam etmektedir. Bu döngü, testin önceden belirlenen durdurma kriterlerinden biri sağlanana kadar sürmektedir. Test sonunda, bireyin yetenek düzeyine ilişkin nihai kestirim yapılmakta ve ölçme süreci tamamlanmaktadır (Linacre, 2000; Makransky & Glas, 2010).

Bu yinelenen süreç, BOBUT sistemlerinin bireyselleştirilmiş ölçme yaklaşımının nasıl işlediğini açıkça ortaya koymakta; madde seçimi, yetenek kestirimi ve test güvenirliliği gibi temel bileşenlerin anlaşılmasına zemin hazırlamaktadır (van der Linden & Glas, 2000; Weiss & Kingsbury, 1984).

BOBUT'un temel felsefesi, bireylerin yetenek düzeyine en uygun test maddelerinin sunulması yoluyla ölçme sürecinin hem geçerliliğini hem de kullanıcı odaklı verimliliğini artırmaktır. Bu doğrultuda geliştirilen BOBUT, yalnızca bireylerin yeteneklerini daha hassas biçimde ölçmekle kalmamakta aynı zamanda testin zamansal, bilişsel ve işlem yükünü de azaltmaktadır (van der Linden & Glas, 2000; Weiss & Şahin, 2024). BOBUT'un avantajları

dört temel başlık altında değerlendirilebilir. Birincisi bireyin yetenek düzeyine en uygun maddeleri alması sayesinde ölçme kesinliği artmakta, böylece bireyin gerçek yetenek düzeyine daha doğru bir şekilde ulaşılabilir (Lord, 2012; Thissen & Mislevy, 2000). İkinci olarak yalnızca bilgi değeri yüksek maddelerin uygulanması, test süresi ve madde tüketimini azaltmakta bu da zaman ve madde tasarrufu yapılmasına olanak tanımaktadır (Kalender, 2011; Sands vd., 1997). Üçüncü olarak testin bireyin performansına göre uyarlanması, kullanıcı deneyimini olumlu yönde etkilemektedir. Yetenek düzeyinin çok üzerinde ya da çok altında madde almayan bireyler, kaygı ve motivasyon açısından olumsuz etkilenmemektedir (Gasperson, 2014). Son olarak ölçme süreçlerinin bireyselleştirilmiş bir biçimde yapılandırılması, testlerin geçerlik ve güvenirliğini artırmakta; bu durum elde edilen sonuçların yorumlanabilirliğini de güçlendirmektedir (Hambleton vd., 1991; van der Linden & Pashley, 2010). BOBUT uygulamalarının sağladığı ölçme kesinliği, zaman ve madde tasarrufu, kullanıcı deneyimi ile testin geçerlik ve güvenirliği gibi boyutlar, ilerleyen bölümlerde ayrıntılı biçimde ele alınacaktır.

2.1.1.1. Ölçme Kesinliği

Ölçme kesinliği, bireyin sahip olduğu gerçek yetenek düzeyinin bir test aracılığıyla ne ölçüde doğru tespit edildiğini ifade etmektedir. Bireye uyarlanmış test sistemleri, bu kesinliği bireyin performansına dayalı olarak belirlenmiş ve en fazla bilgi sağlayan maddelerin uygulanmasıyla önemli ölçüde artırmaktadır (Bergstrom, 1992; Veldkamp & van der Linden, 2002). Geleneksel test uygulamalarında sabit uzunlukta testler kullanılmakta; bu durum, bireylerin yetenek düzeylerini dikkate almadan hem yüksek hem de düşük düzeydeki bireylerin performanslarının ölçülmesinde hatalara yol açabilmektedir. BOBUT uygulamaları ise her bireyin yetenek düzeyini bireyselleştirilmiş biçimde ölçerek bu probleme çözüm getirmektedir (Green vd., 1984).

BOBUT sistemlerinde ölçme kesinliğini artıran temel unsur, birey hakkında en fazla bilgi sağlayacak maddelerin seçilmesidir. Özellikle madde bilgi fonksiyonunun kullanımı sayesinde bireyin tahmini yetenek düzeyine en uygun madde seçilebilir (Chang & Ying, 1996; Kingsbury & Zara, 1991). Bu sayede testin farklı aşamalarında ölçme hatası azaltılmakta ve yetenek kestirimi daha güvenilir hâle gelmektedir. Maksimum Fisher bilgisi ya da a-katmanlı madde seçimi gibi algoritmalar bu amaçla sıkça tercih edilmektedir (Chang & Ying, 1999).

Test süresi boyunca birey hakkında en çok bilgiyi sağlayan maddelerin verilmesi, yalnızca test sonunda değil test boyunca da güvenilir yetenek kestirimleri yapılmasını sağlamaktadır. Brown ve Weiss (1977) tarafından yürütülen çalışmada bu yöntemle klasik testlere göre %50 daha az maddeyle benzer güvenilirlik düzeyine ulaşılabildiği gösterilmiştir. Sonraki araştırmalar, sabit standart hata ya da bilgi düzeyine dayalı test sonlandırma kurallarının ölçme kesinliğini artırdığını ortaya koymuştur (van der Linden & Boekkooi-Timminga, 1989).

Güncel BOBUT uygulamaları, ölçme kesinliğinin yalnızca test uzunluğu ile değil aynı zamanda madde havuzunun niteliğiyle de yakından ilişkili olduğunu ortaya koymuştur. Kaliteli bir madde havuzu, farklı yetenek düzeylerindeki bireylerin ölçülmesinde duyarlılığı ve ayırt ediciliği yüksek maddeler içermelidir (Veldkamp & van der Linden, 2002). Madde güçlük parametrelerinin geniş bir aralıkta dağılması, hem düşük hem de yüksek yetenek düzeyine sahip bireyler için daha fazla bilgi sağlamaktadır. Eggen (1999) madde havuzunda bilgi dengesini koruyan kısıtlamaların ölçme kesinliğini tüm yetenek aralıklarında sabit tutmada etkili olduğunu vurgulamaktadır. Ayrıca Chang ve Ying (1999) yüksek bilgi sağlayan maddelerin yalnızca bilgi düzeyiyle değil içerik kapsamı ve hedef yetenek aralığıyla da uyumlu şekilde seçilmesi gerektiğini belirtmektedir. Bu nedenle ölçme kesinliği yalnızca testin uzunluğuna değil aynı zamanda dengeli ve nitelikli bir madde havuzunun oluşturulmasına da bağlıdır.

Ölçme kesinliğini etkileyen bir diğer kritik nokta kullanılan yetenek kestirim yöntemleridir. BOBUT uygulamalarında yetenek kestiriminde yaygın olarak kullanılan Maximum Likelihood (MLE) ve Bayesci yöntemler bireyin yetenek kestiriminin doğru bir şekilde yapılabilmesi için çeşitli avantajlara sahiptir (Baker, 2001; Wang & Vispoel, 1998). Testin ilk aşamalarında bireyin yetenek düzeyi hakkında yeterli bilgi bulunmadığında Expected a Posteriori (EAP) gibi Bayesci yöntemler bu belirsizliği azaltarak daha güvenilir ölçümler yapılmasına olanak tanır (van der Linden, 1998). Ayrıca test süresince bu yöntemlerin kendini güncellemesi testin tüm aşamalarında daha güvenilir sonuçlar elde edilmesini mümkün kılar (Thissen & Mislevy, 2000). Tüm bu değişkenler düşünüldüğünde ölçme kesinliği BOBUT uygulamalarının yalnızca temel bir özelliği değil aynı zamanda testin işlevselliğini, geçerliliğini ve bireye uyarlanma düzeyini doğrudan etkileyen çok faktörlü bir kalite göstergesidir.

Ölçme kesinliğini etkileyen bir diğer kritik unsur kullanılan yetenek kestirim yöntemleridir. BOBUT uygulamalarında yaygın olarak kullanılan Maksimum Olabilirlik (Maximum Likelihood – MLE) ve Bayesci yöntemler bireyin yetenek düzeyinin doğru şekilde kestirilmesine olanak tanıyan çeşitli avantajlar sunmaktadır (Baker, 2001; Wang & Vispoel, 1998). Testin ilk aşamalarında birey hakkında yeterli bilgi bulunmadığında Expected a Posteriori – EAP gibi Bayesci yaklaşımlar bu belirsizliği azaltarak daha güvenilir ölçümler yapılmasına imkân verir (van der Linden, 1998). Ayrıca test süresi boyunca bu yöntemlerin kendini güncelleyebilmesi, tüm test sürecinde daha güvenilir sonuçlar elde edilmesini mümkün kılmaktadır (Thissen & Mislevy, 2000). Tüm bu değişkenler göz önünde bulundurulduğunda ölçme kesinliği yalnızca BOBUT’un temel bir özelliği değil aynı zamanda testin işlevselliğini, geçerliliğini ve bireye uyarlanma düzeyini doğrudan etkileyen çok boyutlu bir kalite göstergesi niteliğindedir.

2.1.1.2. Zaman ve Madde Tasarrufu

Geleneksel sabit uzunluklu testlerde tüm bireylerin aynı maddeleri alması, test süresinin uzamasına ve bilgi değeri düşük maddelerin gereksiz kullanımına neden olmaktadır. Bu durum yalnızca bireylerin yetenek düzeyinin çok üstünde ya da çok altında maddelerle karşılaşmasına yol açmakla kalmaz aynı zamanda test uygulamaları için ciddi bir zaman ve kaynak israfına da sebep olmaktadır (Lord, 2012; Parshall vd., 2002). Ayrıca orta güçlükteki maddeler yalnızca orta düzeyde yeteneğe sahip bireyler için nitelikli bilgi sağlamaktadır (van der Linden & Glas, 2000; Weiss & Kingsbury, 1984).

BOBUT, bu soruna çözüm olarak test süresince bireyin önceki maddelere verdiği yanıtlardan yola çıkarak yetenek kestirimi yapmakta ve en uygun maddeyi seçmektedir. Bu sayede, bireyin yetenek düzeyi hakkında bilgi vermeyen maddelerin uygulanması engellenmektedir (Chang & Ying, 1996; Stocking & Lewis, 1998). Bu yaklaşım sayesinde test süresi ortalama %30 ile %50 oranında azalmaktadır. Buna rağmen ölçme kalitesi ve güvenilirlik sabit uzunluktaki testlerle eşdeğer düzeyde kalmaktadır (Brown & Weiss, 1977; Eroğlu vd., 2015; van der Linden & Pashley, 2010).

BOBUT’un sağladığı zaman tasarrufu yalnızca test süresiyle sınırlı değildir. Test uygulamalarında gerekli olan bina, gözetmen, kâğıt, kalem, zaman ve güvenlik gibi fiziksel kaynakların kullanımında da önemli ölçüde azalma sağlanmaktadır (Wainer vd., 2000; Weiss & Şahin, 2024). Özellikle çevrimiçi test ortamlarında bu özellik testin her yerde ve her zaman

uygulanabilmesine olanak tanıyarak verimliliği artırmaktadır (Akhtar & Kovács, 2024; Lilley, 2007).

Test geliştirme süreci dikkate alındığında yazılan her bir maddenin belirli aşamalardan geçtiği ve madde yazarlığının uzmanlık gerektirdiği bilinmektedir. Bu nedenle sabit uzunluklu testlerde tüm maddelerin tüm bireylere sunulması hem madde israfına hem de güvenlik sorunlarına yol açmaktadır. Güvenlik nedeniyle bu maddelerin tekrar kullanılma olasılığı ortadan kalkmakta ve madde havuzları hızla tükenmektedir (Stocking & Lewis, 1998; Sympson & Hetter, 1985).

BOBUT uygulamaları ise madde seçimini yalnızca bireyin yetenek düzeyine göre değil aynı zamanda madde kullanım sıklığı ve güvenlik önceliklerine göre de gerçekleştirebilmektedir. Böylece hem bazı maddelerin tekrar tekrar sunulması engellenmekte hem de daha az kullanılan maddelerin ölçme katkısı artırılmaktadır (Davey & Parshall, 1995; Stocking & Lewis, 1998; van der Linden, 2005).

Öte yandan madde havuzunun sadece çok sayıda maddeden oluşması yeterli değildir. Madde havuzunun istatistiksel açıdan dengeli ve kapsamlı şekilde yapılandırılması gerekmektedir. Bu tür dengeli havuzlar, test sürecinde maddelerin daha adil ve eşit biçimde kullanılmasını sağlamak ve test algoritmalarının madde seçimini yalnızca bilgi düzeyine değil aynı zamanda içerik ve konu kapsamına göre de yönlendirmesine olanak tanımaktadır (Armstrong vd., 1992; Glas & van der Linden, 2001; Theunissen, 1985; Veldkamp, 2003).

BOBUT uygulamalarında kullanılan “gölge test” (shadow test) yaklaşımı gibi yöntemler, belirli içerik kısıtlarına uyan ve aynı zamanda bilgi değeri yüksek bireyselleştirilmiş maddeler arasından en uygun olanın seçilmesini mümkün kılmaktadır (van der Linden & Chang, 2003; van der Linden & Veldkamp, 2004). Böylece test sürecinde madde çeşitliliği korunurken aynı zamanda adaletli dağılım sağlanmakta bu da dolaylı olarak madde tasarrufu sağlamaktadır (Luecht & Nungester, 1998; Segall, 2005).

Sonuç olarak BOBUT sistemleri yalnızca test süresinin kısaltılması ve bireyselleştirme açısından değil aynı zamanda madde güvenliği, havuz dengesi ve madde kullanım sıklığı yönünden de sistemsel verimlilik sunmaktadır. Bu yönüyle yüksek güvenlik gerektiren geniş ölçekli testlerde BOBUT sistemleri yalnızca teknolojik değil aynı zamanda stratejik bir çözüm olarak öne çıkmaktadır. Bu durum hem psikometrik niteliklerin kalitesini artırmakta hem de uygulama maliyetlerinin etkinliğini yükseltmektedir (Patsula, 1999; Yen, 1993).

2.1.1.3. Kullanıcı Deneyimi

Kullanıcı deneyimi (User Experience – UX) bireyin bir ürün, hizmet ya da sistemle etkileşim kurduğu süreç boyunca edindiği bilişsel, duygusal ve davranışsal izlenimlerin bütünüdür (Lin, 2007). Eğitim teknolojileri bağlamında kullanıcı deneyimi yalnızca sistemin kullanılabilirliği ile değil aynı zamanda bireyin yaşadığı stres, kaygı, memnuniyet, motivasyon düzeyleri ve öğrenme sürecine olan etkileri ile de ilişkilidir (Akhtar & Kovács, 2024). Bu nedenle BOBUT tasarımında temel hedeflerden biri kullanıcı deneyimini iyileştirmektir. Çünkü bu sistemlerin başarısı yalnızca psikometrik niteliklerinin kalitesiyle değil bireyin test sürecine ilişkin algıları ve etkileşim düzeyiyle de doğrudan bağlantılıdır (van der Linden & Glas, 2000; Weiss & Şahin, 2024).

BOBUT uygulamalarında kullanıcı deneyimini olumlu etkileyen en önemli unsurlardan biri bireyin yetenek düzeyine göre madde seçimi yapılmasıdır. Bu strateji, bireyin yeteneği hakkında yüksek bilgi sağlayan maddelerin uygulanmasını mümkün kılmakta ve ölçmeye katkı sunmayan maddelerin sunulmasını engellemektedir (Chang & Ying, 1996; Stocking & Lewis, 1998). Böylece birey, çok kolay ya da çok zor maddelerle karşılaşmamaktadır. Bu durum, test motivasyonunun korunmasına ve deneyimin bireyselleşmesine olanak tanır (Kezer & Koç, 2009). Testi alan bireylerin yetenek düzeylerine uygun sorularla karşılaşmaları test süresince yaşanabilecek stres ve kaygı düzeyini düşürmektedir. Bu sayede bireyin teste yönelik güveni artmakta ve performansını daha rahat bir şekilde sergilemesi mümkün olmaktadır (Hambleton vd., 1991; Wainer vd., 2000). BOBUT, bireyin test boyunca başarısızlık hissi yaşamasını önleyerek daha olumlu bir test deneyimi sunmaktadır. Bu durum, sabit uzunlukta testlerde yaygın olarak karşılaşılan “çok zor sorularla erken karşılaşma” gibi olgular nedeniyle oluşan motivasyon kaybı ve testten kopmanın önüne geçmektedir (van der Linden & Glas, 2000; Weiss & Kingsbury, 1984).

Kullanıcı deneyimi açısından yalnızca maddelerin zorluk düzeyi değil aynı zamanda test süresi de belirleyici bir değişkendir. BOBUT’un doğası gereği yetenek kestirimine katkı sunmayan maddeler sorulmaz. Bu durum test süresini kısaltmakta böylece bireyin dikkat süresi korunurken test yorgunluğunun da önüne geçilmektedir (Kalender, 2011; van der Linden & Pashley, 2009). Araştırmalar uzun süreli testlerin bireylerde dikkat dağınıklığı, zihinsel yorgunluk ve algılanan stres düzeyi üzerinde olumsuz etkilere sahip olduğunu göstermektedir (Lilley vd., 2004; Martin & Lazendic, 2018). Bu nedenle BOBUT uygulamalarında kullanılan hata eşliğine veya bilgi yeterliliğine dayalı sonlandırma kuralları

yalnızca test verimliliğini artırmakla kalmaz aynı zamanda bireyin testi olumlu bir deneyim olarak algılamasını da sağlar (Weiss & Şahin, 2024).

Bu süreçte bireylerin testten aldıkları geribildirimler de kullanıcı deneyiminin önemli bir parçasıdır. BOBUT uygulamalarında sunulan geribildirimler, bireylerin güçlü ve gelişime açık yönlerini tanımalarına yardımcı olmaktadır. Özellikle biçimlendirici amaçla yapılan test uygulamalarında bireyselleştirilmiş geribirimler bireyin öğrenmeye olan ilgisini ve motivasyonunu olumlu yönde etkilemektedir (Economides, 2005; Lilley, 2007). Bu yönüyle BOBUT, yalnızca ölçme amacıyla değil aynı zamanda öğrenmeyi destekleyen bir yapı olarak da kullanıcı deneyimi açısından değerlidir.

2.1.1.4. Test Güvenilirliği ve Geçerliliği

Test güvenilirliği bir ölçme aracının duyarlı, tutarlı ve kararlı sonuçlar üretme kapasitesidir. Başka bir deyişle testin tekrarlandığında benzer sonuçlar vermesidir. Bu durum bireylerin testten elde ettikleri puanların büyük ölçüde gerçek yetenek düzeyine bağlı olarak kestirilmesi ve tesadüfi hatalardan arındırılmış olması anlamına gelir (Anastasi & Urbina, 1997; Hambleton vd., 1991). Güvenilir bir test bireyin yeteneğini sistematik biçimde ölçer ve bu ölçme farklı koşullarda tutarlılığını korur (Lord & Novick, 2008; Weiss & Şahin, 2024).

Geçerlilik ise bir testin amacına uygun olarak ölçmek istediği yapıyı ne derecede doğru biçimde ölçtüğü ile ilgilidir. Geçerliliği sağlanmış bir test elde edilen puanların doğru biçimde yorumlanmasına olanak tanır ve bu sayede daha sağlıklı kararlar alınabilir (APA, 1999; Messick, 1989). Geçerlilik; kapsam, yapı ve ölçüt geçerliliği gibi bileşenleri içerir (Messick & Linn, 1989). Özellikle eğitim ve psikolojik değerlendirme bağlamındaki geçerlilik, bireyin yalnızca mevcut düzeyini değil gelecekteki performansını da tahmin etme imkânı sunar (Crocker & Algina, 1986).

BOBUT, hem geçerlilik hem de güvenilirlik açısından geleneksel test uygulamalarına göre birçok avantaj sunar. Güvenirlik açısından en belirgin avantaj bireyin yetenek düzeyine uygun madde seçiminin yapılabilmesidir. MTK temelli olarak her bireye yeteneği hakkında en fazla bilgi sağlayacak madde yöneltilir. Bu durum testin standart hatasını azaltarak ölçmenin güvenilirliğini artırır (van der Linden & Hambleton, 1997; Weiss & Kingsbury, 1984). Örneğin Aybek'in (2016) yürüttüğü doktora çalışmasında yapılan simülasyonlar, BOBUT'un daha az sayıda maddeyle daha güvenilir ölçme sonuçları sağlayabildiğini ortaya koymuştur.

Geçerlilik açısından BOBUT uygulamaları bireyin yetenek düzeyine uygun maddelerin yöneltilmesi ile hem kapsam hem de yapı geçerliliğini artırır. Geniş ve dengeli madde havuzları, farklı yetenek düzeylerini kapsayarak testin hedeflediği özelliği daha eksiksiz biçimde temsil eder (Chang & Ying, 1996). Ayrıca bireylerin yetenek düzeyi hakkında en fazla bilgi sağlayan maddelerle yapılan ölçmelerin dışsal ölçütlerle yüksek korelasyon göstermesi BOBUT'un yüksek ölçüt geçerliliğine sahip olduğunu göstermektedir. Weiss ve Kingsbury (1984) BOBUT'tan elde edilen yetenek kestirimleri ile bireylerin okul başarıları arasında yüksek bir korelasyon olduğunu belirtmiştir. Benzer şekilde Bartham ve Hambleton'un (2006) çalışmaları da bireylerin BOBUT yetenek kestirimleri ile iş performansları arasında istatistiksel olarak anlamlı ilişkiler olduğunu ortaya koymuştur.

Tüm bu teorik çerçeve BOBUT uygulamalarının psikometrik açıdan neden tercih edilmesi gerektiğini göstermektedir. Öte yandan yapılan çalışmalar da BOBUT'un bu üstünlüğünü doğrulamaktadır. BOBUT'un bireyselleştirilmiş yapısı yalnızca doğru yetenek kestirimi yapmakla kalmaz aynı zamanda test süresini de kısaltır. Böylece bireyin dikkat seviyesi ve motivasyonu daha uzun süre korunabilir (Babcock & Weiss, 2012; Lilley vd., 2004). Bu özellik bireyler hakkında önemli kararların alındığı testlerde (örn., üniversiteye geçiş sınavları, dil yeterlilik testleri ya da mesleki sertifikasyon uygulamaları) kritik bir öneme sahiptir. Ayrıca öğrenme güçlüğü yaşayan bireylerin bulunduğu özel gruplar için daha adil ve bireyselleştirilmiş bir ölçme imkânı sunarak kaygı düzeyini azaltabilir (Wainer vd., 2000).

BOBUT'un bu avantajları yalnızca simülasyon çalışmalarıyla değil kurumsal düzeyde yapılan saha araştırmalarıyla da desteklenmektedir. Hart ve arkadaşlarının (2006) yürüttüğü çalışmada fiziksel işlevselliğe ilişkin geliştirilen BOBUT'un hem test süresini kısalttığı hem de geleneksel testlere göre daha güvenilir sonuçlar verdiği görülmüştür. Ayrıca Oswald ve diğerlerinin (2015) ABD Donanması için geliştirdikleri BOBUT uygulamasında adayların eğitim süreci performansları ile yetenek kestirimleri arasında güçlü ve anlamlı korelasyonlar bulunmuştur. Bu durum ölçüt geçerliliğinin yalnızca bireysel değil aynı zamanda kurumsal düzeydeki kararlar açısından da etkili olduğunu ortaya koymaktadır.

Tüm bu bulgular bir arada değerlendirildiğinde BOBUT uygulamalarının yalnızca teorik düzlemde değil uygulamada da geçerlilik ve güvenilirlik açısından güçlü bir yapı sunduğu anlaşılmaktadır. Gerçekleştirilen simülasyonlar ve uygulamalar BOBUT'un bireyselleştirilmiş ve etkili bir ölçme yaklaşımı olduğunu ortaya koymaktadır.

2.1.3. BOBUT'un Eğitimde Kullanım Alanları

BOBUT, eğitim arařtırmalarında geniş bir uygulama alanı bularak bireylerin öğrenme süreçlerine ve ölçme-değerlendirme yaklaşımlarına önemli katkılar sağlamaktadır. Özellikle bireylerin yetenek düzeyine göre test sürecinin uyarlanabilmesiyle ölçmeler daha duyarlı hâle gelmektedir. Bu durum, BOBUT'un farklı disiplinlerde de benimsenmesini mümkün kılmıştır.

Bireylerin akademik başarılarının değerlendirilmesi sürecinde yaygınlaşan BOBUT uygulamaları öğretmenlerin daha kısa sürede daha doğru sonuçlar elde edebilmesine imkân tanımaktadır (Eroğlu & Kelecioğlu, 2013). İlk BOBUT uygulamaları, 1970'li yıllarda bireyselleştirilmiş öğretimi desteklemek amacıyla ABD'de geliştirilen Minnesota Adaptive Instructional System (MAIS) ile başlamıştır (Sands vd., 1997). Günümüzde ise PISA (Programme for International Student Assessment) ve TIMSS (Trends in International Mathematics and Science Study) gibi uluslararası testlerde de bireye uyarlanmış test teknolojileri yaygın şekilde kullanılmaktadır (van der Linden & Glas, 2000).

BOBUT'un geniş örneklemeler üzerinde yapılan testlerde daha düşük standart hata ile ölçme yapılabilmesi bu yöntemin tercih edilmesinde belirleyici olmaktadır (Hambleton vd., 1991; van der Linden & Glas, 2000). Ayrıca sağlık, mühendislik ve bilişim gibi alanlarda gerçekleştirilen mesleki sertifikasyon testlerinde ve bireylerin iş performanslarını değerlendirmeye yönelik uygulamalarda da etkin şekilde kullanılmaktadır (van der Linden, 2016; Wainer vd., 2000).

Uluslararası ölçekte uygulanan GRE (Graduate Record Examination) ve GMAT (Graduate Management Admission Test) gibi testler BOBUT sistemlerinin kullanıldığı testlerdir (Weiss & Şahin, 2024). Benzer şekilde ABD'de temel akademik becerileri değerlendiren MAP Growth (Measures of Academic Progress) ve Hollanda'da uygulanan CITO sistemi ulusal düzeyde yürütölen BOBUT sistemlerine örnek teşkil etmektedir (Aybek & Çıkrıkçı, 2018; van der Linden & Glas, 2000).

BOBUT uygulamalarının kullanıldığı bir diğör alan ise yabancı dil testleridir. Bireylerin dört temel dil becerisini bireyselleştirilmiş şekilde ölçmesi bu yöntemin bu alanda yaygınlaşmasına olanak tanımıştır (Stocking & Lewis, 1995). TOEFL (Test of English as a Foreign Language) ve IELTS (International English Language Testing System) gibi uluslararası dil yeterliliği testlerinde BOBUT teknolojileri kullanılmaya başlanmıştır (Makransky vd., 2013).

BOBUT'un eğitim alanındaki kullanım alanlarından biri de özel eğitime ihtiyaç duyan bireylerin değerlendirilmesidir. Geleneksel testler bu bireylerin başarı düzeyini tam olarak yansıtamazken; BOBUT'un performansa dayalı madde seçme algoritmaları, bireylerin yetenek düzeyine uygun maddelerin yöneltmesine imkân tanır (Ebenbeck vd., 2024; Weiss & Şahin, 2024). Özellikle düşük yetenek düzeyine sahip bireyler için testin kolay maddelerle başlatılması ve cevaplara göre madde güçlüğü'nün güncellenmesi hem ölçmenin güvenilirliğini artırır hem de bireyin motivasyonunu destekler (Choi & McClenen, 2020; Wang & Vispoel, 1998).

Uzaktan eğitimin yaygınlaşmasıyla birlikte BOBUT uygulamalarının önemi de artmıştır. Bu sistemler, çevrimiçi platformlarda zaman ve madde tasarrufu sağlayarak değerlendirme süreçlerini daha verimli hâle getirir (Keskin & Günay, 2021; Oladele & Ndlovu, 2024). Ayrıca öğrencilere anlık geribildirim sunulması öğrenme süreçlerini desteklemektedir (Choi & McClenen, 2020). Sonuç olarak BOBUT'un sunduğu avantajlar sayesinde eğitim alanındaki kullanımı giderek yaygınlaşmaktadır. Bu sistemlerin çağdaş eğitim yaklaşımlarının gereksinimlerine yanıt verebilmesi ölçme ve değerlendirme alanında güçlü ve etkili bir araç hâline gelmesini sağlamaktadır (Eggen, 2018; Cheng, 2009).

2.1.4. BOBUT'un Uygulanabilirliği ve Gelecek Yönelimler

BOBUT, modern ölçme ve değerlendirme uygulamalarında yaygınlaşmış bir yaklaşımdır. Bireysel farklılıklara odaklanarak test deneyimini özelleştiren bir yapı sunar. Son yıllarda yapılan çalışmalar, BOBUT'un test süresinde sağladığı avantajın yanı sıra ölçme duyarlılığı ve test güvenirliliği gibi konulara da odaklanmaktadır (Weiss & Şahin, 2024). Ayrıca günümüz dünyasında hızla önem kazanan yapay zekâ ve makine öğrenmesi tabanlı algoritmaların BOBUT içinde kullanımı da araştırılan bir diğer konudur (Wang vd., 2021).

Test uygulamalarında BOBUT'un etkin şekilde kullanılabilmesi hem teknik hem de pedagojik anlamda ileri düzeyde bilgi birikimi gerektirmektedir. Özellikle madde havuzunun oluşturulması, algoritmaların belirlenmesi, yazılım ve ara yüzün kullanıcıya göre optimize edilmesi ile güvenlik gibi teknik konular, sistemin geçerliliği ve güvenirliliği üzerinde doğrudan etkilidir (Cheville vd., 2021). Bununla birlikte testin kapsama uygunluk çerçevesinde yapılandırılması test adaleti açısından kritik bir öneme sahiptir (Chapman vd., 2024).

Yeni nesil BOBUT uygulamaları, kullanıcı deneyimi odaklı tasarımlarla test arayüzlerini daha etkileşimli hâle getirmektedir. Bu bağlamda öğrencilere anlık ve açıklayıcı geribildirimler sunulmaktadır (Chen vd., 2022). Ayrıca test sürecinin bireyselleştirilmesiyle yalnızca bilgi düzeyi değil aynı zamanda bireysel öğrenme hızı ve tercihleri de ölçülebilir hâle gelmektedir (Weiss & Şahin, 2024). BOBUT araştırmalarında odaklanılan bir diğer yönelim ise dijital ortamlarda test güvenliğidir. Bu kapsamda sürekli test uygulamalarında madde kullanım sıklığının kontrolü test güvenliği açısından kritik öneme sahiptir (Wang vd., 2021). Güncel araştırmalar, psikometrik kaliteyi ve adil ölçme ilkelerini zedelemekten madde kullanım sıklığını azaltmayı amaçlamaktadır.

Sonuç olarak BOBUT sistemlerinin kullanımının sürdürülebilirliği için yazılım ve test teorisi temelli araştırmaların devam ettirilmesi gerekmektedir. Yeni teknolojilerle uyumlu, etik ve güvenilir ölçme sistemlerinin geliştirilmesi ölçme ve değerlendirme alanında dönüştürücü bir rol oynayacaktır.

2.1.4.1. BOBUT Uygulamalarında Karşılaşılan Sorunlar

BOBUT'un başarısı büyük oranda madde havuzunun kalitesine ve uygun şekilde kalibre edilmiş olmasına bağlıdır. Madde havuzunun nitelikleri testin güvenilirliğini, geçerliliğini ve ölçme doğruluğunu doğrudan etkilemektedir. BOBUT uygulamasına katılan her bireyin yetenek kestirimi doğru biçimde yapılmalıdır. Bunun için farklı güçlük düzeylerinde ve yüksek ayırt ediciliğe sahip maddelerden oluşan bir madde havuzuna ihtiyaç vardır (van der Linden & Glas, 2000). Geniş bir madde havuzu, her yetenek düzeyine uygun maddelerin seçilmesini mümkün kılarken aynı zamanda maddelerin tekrar edilme sıklığını azaltmaktadır. Dar bir madde havuzu ise bireylerin aynı maddelerle sık sık karşılaşmasına yol açabilir ve bu durum havuzun güvenliğini ve etkililiğini olumsuz etkiler (Gündeğer & Doğan, 2018; Stocking & Lewis, 1995). Ayrıca geniş kapsamlı bir madde havuzu test içeriğinin dengelenmesini sağlayarak test geliştiricilere kolaylık sunar.

Madde havuzunun önemli bir diğer özelliği de kapsam geçerliliğinin yüksek olmasıdır. Kalender (2011) çalışmasında kapsam geçerliliği daha yüksek olan bir havuzun testin genel geçerliliğini artırdığını ortaya koymuştur. Madde havuzunun kapsamını dengeli tutmak için gölge test algoritmalarından yararlanılabilir (Weiss & Şahin, 2024). Test geliştiricilerin dikkat etmesi gereken bir başka husus ise madde havuzunun güvenliğidir. Bu güvenlik testin güvenilirliği ve adaletinin korunması açısından son derece önemlidir. Madde kullanım

sıklığının kontrol edilmesi ya da havuzdaki maddelerin düzenli olarak güncellenmesi gibi yöntemlerle bu güvenlik sağlanabilir (Lilley, 2007; Stocking & Lewis, 1995). Ayrıca madde havuzunun değişen ihtiyaçlara ve test ortamlarına uyum sağlayacak şekilde belli aralıklarla güncellenmesi testin geçerliliğini ve güvenilirliğini artırmaktadır (Gündeğer & Doğan, 2018). Farklı kültürlere yönelik uygulamalarda ise madde havuzunun kültürel uygunluğu test edilerek gerekli uyarlamaların yapılması gereklidir (van der Linden & Glas, 2000).

BOBUT uygulamalarının verimli çalışabilmesi için hızlı işlemciler ve yeterli bellek kapasitesi gerekmektedir. Bu gereksinimler sağlanmadığında test süreci olumsuz etkilenebilir (Weiss & Şahin, 2024). Ayrıca, test sonuçlarının güvenli biçimde saklanması ve işlenmesi için gelişmiş sunucu teknolojilerinin kullanılması önerilmektedir (van der Linden & Glas, 2000). Günümüzde mobil cihazların yaygınlaşmasıyla, testlerin mobil uyumlu olması da önem kazanmıştır. Mobil cihazlarla uyumlu olarak geliştirilen kullanıcı dostu arayüzler, test deneyimini olumlu yönde etkilemektedir (Lilley, 2007).

BOBUT sistemlerinin başarısını etkileyen bir diğer unsur ise kullanıcıların sisteme dair bilgi düzeyi ve kabulüdür. Öğrenciler ve öğretmenler, geleneksel kâğıt-kalem testlerine daha alışkın olduklarından, BOBUT sistemlerine karşı psikolojik direnç gösterebilmektedir. Bu durum, özellikle bilgisayar okuryazarlığı düşük bireylerde daha sık gözlenmektedir (Lilley, 2007). Ayrıca test kullanıcıları algoritmaların kişilerarası adil bir değerlendirme yapıp yapmadığı konusunda şüphe duyabilir (Kalender, 2011). Bu şüphelerin giderilmesi için bireylere verilen geribildirimlerle sistemin işleyişi açıklanmalı ve kaygılar kontrol altına alınmalıdır. Algoritmanın nasıl çalıştığını bilen ve sonuçları doğru yorumlayabilen kullanıcılar sisteme karşı daha olumlu bir tutum geliştirebilir (Aybek, 2016). Öte yandan gerçek zamanlı bireyselleştirme özelliği, özellikle düşük yetenek düzeyindeki bireylerde kaygıyı artırabilecek bir diğer etmendir (van der Linden & Glas, 2000).

2.1.4.2. BOBUT Uygulamalarında Güncel Araştırma Alanları

Eğitimde ölçme ve değerlendirme sürecinde bilgisayar temelli uygulamaların yaygınlaşması, BOBUT araştırmalarının da derinleşmesine olanak sağlamıştır. Bu bağlamda araştırmacılar, BOBUT uygulamalarını daha etkili, adil ve kullanıcı dostu bir yapıya dönüştürmek amacıyla çalışmalar yürütmektedir. Bu doğrultuda, yapay zekâ algoritmalarının kullanımı ve kullanıcı deneyimi odaklı BOBUT çalışmaları giderek daha fazla önem kazanmaktadır.

2.1.4.2.1. Yapay Zeka Destekli Bireyselleştirilmiş Sistemler

BOBUT, bireyselleştirilmiş ölçme deneyimi sunmak amacıyla geliştirilen test sistemleridir. Uygulamalarda yaygın olarak Madde Tepki Kuramı (MTK) ve Bayesçi yöntemler kullanılmaktadır. Günümüzde BOBUT çalışmaları bu geleneksel yöntemlerle yapay zekâ ve makine öğrenmesi algoritmalarının birleşimine odaklanmaktadır. Bu bağlamda BOBUT uygulamalarının daha duyarlı bir ölçme süreci sunması ve kullanıcı odaklı, uyarlanabilir bir yapıya dönüştürülmesi hedeflenmektedir (Luo, 2020).

BOBUT uygulamalarında kullanılması hedeflenen derin öğrenme modelleri, testi alan öğrencinin önceki cevaplarından öğrenerek daha isabetli madde seçimi gerçekleştirebilmekte ve yetenek kestirimi sürecinde doğruluğu artırmaktadır (Pan vd., 2023). Bu yaklaşım bireylerin bilişsel düzeylerinin yanı sıra duyuşsal ve davranışsal özelliklerini de modele dâhil edebilmektedir. Örneğin yüz tanıma, duygu analizi ve dikkat düzeyi takibi gibi test sürecinde gerçek zamanlı toplanan veriler aracılığıyla testin duyuşsal ve davranışsal açıdan bireyselleştirilmesi sağlanabilir (Marfeo vd., 2022).

Yapay zekâ algoritmalarının BOBUT uygulamalarına bir diğer katkısı da otomatik madde üretimi çalışmalarıdır. Otomatik madde üretimiyle belirli konu alanlarında madde modelleri kullanılarak benzer biçimde binlerce eşdeğer madde üretilebilmektedir. Bu durum test maliyetlerini önemli ölçüde azaltmaktadır (Luo, 2020). Ayrıca chatbot destekli madde yazım sistemleri de geniş ölçekli test uygulamalarında ölçeklenebilir çözümler sunmaktadır.

Gelecekteki BOBUT uygulamalarına yönelik araştırmalarda kullanıcıların demografik özelliklerinin, öğrenme biçimlerinin ve performans geçmişlerinin testte dikkate alınması önemli bir odak noktasıdır. Bu değişkenlerin kullanılmasıyla oluşturulan testlerin daha bireyselleştirilmiş bir yapı sunacağı ve bu sayede bireylerin test kaygısının azaltılarak ölçme doğruluğunun artırılacağı öngörülmektedir (Postigo vd., 2020). Bu bağlamda geliştirilen testlerin eğitimde adil ölçme ilkeleri açısından da önemli katkılar sunacağı düşünülmektedir.

Son olarak BOBUT çalışmalarındaki yeni yönelimlerden biri de gerçek zamanlı öğrenen algoritmalarla ölçme yapılmasıdır. Bu tür modeller her cevaptan sonra yeni bilgilerle güncellenerek tahmin doğruluğunu artırmayı amaçlamaktadır (Pan vd., 2023). Tüm bu çalışmalar değerlendirildiğinde yapay zekâ ve makine öğrenmesi tabanlı bireyselleştirilmiş test sistemlerinin etik, teknik ve pedagojik açıdan geliştirilmesi gerekliliği öne çıkmaktadır.

2.1.4.2.2. Kullanıcı Deneyimi Odaklı Test Yenilikleri

BOBUT uygulamalarının başarısı yalnızca test algoritmalarına değil aynı zamanda kullanıcı dostu tasarımların oluşturulmasına da bağlıdır. Özellikle sade, etkileşimli ve kullanıcı dostu arayüz tasarımlarının kullanılması bireylerin teste yönelik algılarını olumlu yönde etkileyerek performansı artırabilmektedir (Wang vd., 2021). Günümüzde test uygulamalarının kullanıcı deneyimi açısından zenginleştirilmesini sağlayan bir diğer yenilik geribildirim sistemleridir. Gerçek zamanlı ve açıklayıcı geribildirimler sunan sistemler öğrenme sürecini desteklerken test kaygısını da azaltmaktadır (Cheville vd., 2021).

Kullanıcı deneyimini iyileştirmeye yönelik bir başka yaklaşım ise oyunlaştırma öğelerinin test yapısına entegre edilmesidir. Bu bağlamda test sürecine “rozet”, “puan” veya “seviye atlama” gibi motivasyonu artırıcı öğelerin eklenmesi özellikle küçük yaş gruplarında test katılımını ve ilgisini artırdığı gözlemlenmiştir (Vanduhe vd., 2020).

Çalışmalar, test uygulamalarının bireylerin bilgi işleme hızına göre bireyselleştirildiğinde zaman baskısının ortadan kalktığını göstermektedir. Bu sayede bireylerin yetenek düzeyleri daha güvenilir bir şekilde ölçülebilmektedir (Pan vd., 2023). Kullanıcı deneyiminin geliştirilmesine yönelik çalışılan bir diğer konu ise BOBUT uygulamalarının erişilebilirliğidir. Bu doğrultuda görme ve işitme engelli bireyler için sesli yönlendirmeler ve işaret dili modülleri geliştirilmiştir (Marfeo vd., 2022).

Çok dilli ve kültürlerarası uyarlamaların BOBUT uygulamalarına entegre edilmesi de kullanıcı deneyiminin geliştirilmesi açısından önem taşımaktadır. Farklı dil ve kültürlerdeki bireylerin testi yalnızca çevrilmiş bir içerik olarak değil bağlamsal uyarlama esasına dayalı olarak alması madde geçerliliği ve test adaletinin sağlanmasını kolaylaştırmaktadır (Xu vd., 2020).

2.2. Madde Tepki Kuramı (MTK)

Test kavramı, psikometrik ve ölçme kuramı bağlamında, bireylerin belirli bir özelliğe (örneğin yetenek, bilgi, tutum, kişilik) sahip olma düzeylerini belirlemek amacıyla kullanılan yapılandırılmış ve sistematik bir değerlendirme aracını ifade eder. Testler, bireylerin verdikleri tepkilerden yola çıkarak gizil (latent) özellikler hakkında çıkarım yapmayı amaçlayan araçlardır (Embretson & Reise, 2000). Test kavramının anlaşılabilmesi için ölçme mantığını, testlerin doğasını ve yapısını kavramak önemlidir. En kapsayıcı hâliyle test, belirli

bir çevreden ölçülmesi hedeflenen özelliğin belirlenebilmesi adına uygulanan bir dizi süreci ifade eder (Crocker & Algina, 1986).

Test geliştirme sürecinde öncelikle testten elde edilecek puanların ne amaçla kullanılacağı belirlenmelidir. Örneğin bireylerin belirli özelliklerinin tanınması üzerine kurgulanan tanımaya yönelik bir test için maddeler nispeten kolay olmalı, geniş yetenek aralığına sahip bireylerin katıldığı bir testte ise maddeler orta güçlükte, belirli bir hedef doğrultusunda yapılacak seçimi test etmek amacıyla daha zor düzeyde hazırlanmalıdır (Baykul, 2015).

Testin oluşturulması sürecinde amaçlar belirlendikten sonra bu amaçlara uygun bir çerçeve çizilmelidir. Bu çerçeve, testin kapsamı olarak adlandırılmakta olup ayrıntılı ve planlı bir şekilde oluşturulmalıdır. Seçilecek maddelerin testin kapsamını en iyi şekilde örneklendirmesi hedeflenir (Roid & Haladyna, 1982). Kapsamın uygunluğunu sağlamak amacıyla belirtke tablosu oluşturulması yaygın bir uygulamadır. Belirtke tablosu, testte yer alacak konuları ve bu konuların ağırlıklarını gösterir ve kapsam geçerliliğinin sağlanması açısından kritik öneme sahiptir (Atılğan vd., 2017).

Lindquist (1940) test geliştiricilerinin görevini "Ne ölçülecek?" ve "Nasıl ölçülecek?" soruları çerçevesinde tanımlamıştır. Testin kapsam geçerliliği sağlandıktan sonra bu kapsamın nasıl yoklanacağı yani madde havuzunun nasıl oluşturulacağı belirlenmelidir. Lindquist'e (1940) göre madde havuzu oluşturulurken önce uygun madde formatı seçilmeli, bu formatın testi alacak bireyler açısından anlaşılır ve uygulanabilir olup olmadığı doğrulanmalıdır. Ardından maddeleri yazacak uzmanlar belirlenmeli, gerekli eğitimleri aldıktan sonra maddeleri oluşturmalıdır. Maddeler yazıldıktan sonra dil uygunluğu, anlaşılabilirlik ve ölçme-değerlendirme açısından uygunluk kriterleri temel alınarak kalite kontrolü yapılmalı ve test uygulanabilir hâle getirilmelidir.

Maksimum performans testleri (örneğin yetenek ya da başarı testleri) farklı madde formatlarından oluşabilmektedir. Bugüne kadar kullanılan testler, bireyin cevabı kendi oluşturduğu sübjektif testler ve doğru cevabın seçenekler arasında verildiği objektif testler olmak üzere iki gruba ayrılmaktadır (Popham, 1990). Objektif testlerde bireyin ölçülmesi hedeflenen davranışların dışında şans yoluyla doğru cevaba ulaşması mümkündür. Örneğin çoktan seçmeli bir testte birey doğru yanıtı şansa seçebilir. Bu durumun önüne geçebilmek için çeldiriciler, madde kökü ve doğru yanıt kadar önemsenmeli; bilen ve bilmeyeni ayırt

edebilecek şekilde hazırlanmalıdır. Bu da test geliştirme uzmanının sadece teknik bilgiye değil aynı zamanda hedef grubu iyi tanımasına da bağlıdır.

Madde havuzu oluşturulduktan sonra test geliştirme uzmanlarının maddeleri çeşitli alan uzmanlarıyla birlikte incelemesi gerekir. Bu inceleme madde doğruluğu, kapsam uygunluğu, yazım tekniği, dil bilgisi, okunabilirlik düzeyi ve olası önyargılar açısından yapılmalıdır. Ardından maddeler küçük bir örneklem üzerinde denenmeli ve bireylerin maddelere verdikleri tepkiler gözlemlenerek gerekli revizyonlar yapılmalıdır. Bu sürecin sonunda testin nihai formu elde edilir (Crocker & Algina, 1986).

Test geliştirme uzmanları testin farklı koşullarda, farklı örneklemelerde ve farklı zamanlarda benzer sonuçlar üretmesini hedefler. Ancak bu sürece tesadüfi hatalar karışabilir. Güvenirlik, testin hatalardan ne ölçüde arınık olduğunu gösteren bir ölçüttür (Zeller & Carmines, 1980). Yani güvenirlik, ölçme işlemine karışan tesadüfi hataların etkisini ifade eder. Klasik test kuramı kapsamında güvenirlik formülünü ilk olarak Spearman (1910; 1961; 1987) geliştirmiştir. Spearman'ın formülüne göre;

$$X = T + E$$

X gözlenen puanı, T gerçek puanı ve E hata miktarını göstermektedir. Lord (1952) hata kavramına başka bir çözüm olarak, iyi tanımlanmış bir madde havuzundan rastgele seçilen maddeler oluşturulduğunda bu maddelerin binom standart hatası olarak ortaya çıktığını belirtmiştir.

Test geliştirme uzmanlarının bir diğer çalışması gereken konu ise testin amacına ne düzeyde uygun olduğudur. Testin amaca uygun olması geçerlilik olarak tanımlanmaktadır. Testin çalışılan amaca uygun olması ve uygulanabilecek diğer değerlendirme yöntemlerine göre tercih edilebiliyor olması gerekmektedir (Messick, 1980).

Test geliştirme sürecinde genel hedef geçerlilik ve güvenirlik durumları göz önünde bulundurularak minimum uzunluktaki testi ortaya koyabilmektir. Hedeflenen amaç doğrultusunda öncelikle geniş bir madde havuzu gereklidir. Daha sonra bu havuzdan geçerliliğe ve güvenirliğe en çok katkı verecek maddeler belirlenmelidir. Seçilen maddelerle nihai test formu oluşturulmalıdır. Ortaya konulacak nihai test formunun oluşturulabilmesi adına test geliştirme uzmanları madde analizlerini (madde güçlük parametresi, madde ayırt edicilik parametresi vb.) kullanmaktadır (Crocker & Algina, 1986). Madde analizleri sonucunda test geliştirme uzmanı geçerliliği ve güvenirliği kanıtlanmış nihai test formuna ulaşabilmektedir.

Eğitimde ve psikolojide testlerin oluşturulması sürecinde kullanılan modeller, test geliştirme uzmanlarına ve test kullanıcılarına uzun bir süredir hizmet etmektedir. Testlerin geliştirilmesinde kullanılan modeller incelendiğinde alan yazında uzun süredir var olan model klasik test kuramıdır (KTK). KTK ile oluşturulan modeller test geliştirme uzmanlarına uzun süre hizmet etmiştir (Hambleton, 2013). Ancak KTK'nın bazı sınırlılıkları farklı test geliştirme modellerinin ortaya çıkmasını sağlamıştır. Test modellerinin gelişiminde temel amaç bireyin yeteneğinin geçerli ve güvenilir bir şekilde ölçülmesidir. KTK'da yeteneğin ölçümü ile ilgili bulunan bazı sorunların çözümü adına test geliştiricileri madde tepki kuramını (MTK) geliştirmiştir. Bugün halen daha yaygın bir şekilde KTK kullanılmaya devam etmektedir (Hambleton, 1994; Reise vd., 2005). Madde tepki kuramının klasik test kuramı ile farklılıkları incelendiğinde (Koğar, 2021):

Ölçmenin standart hatası KTK'da tüm yetenek düzeyleri için aynıdır. Ancak MTK'da farklı yetenek düzeyleri için ölçmenin standart hatası değişiklik göstermektedir. KTK için testi alan bireylerin yetenek düzeyleri fark etmeksizin standart hatalarının aynı olması ve güvenilirlik düzeyinin her yetenek düzeyi için aynı kestirilmesi durumu çeşitli yanılgılara sebep olabilmektedir. MTK her yetenek düzeyi için farklı güvenilirlik kestirimi yaptığı için yetenek düzeyleri ile ilgili daha doğru bilgilere ulaşılmasına olanak sağlar. Yani test, farklı yetenek düzeylerindeki bireylere uyarlanabilir hale gelmiştir. Ölçüm hassasiyeti genel olarak algılanmak yerine bireylerin yetenek düzeyleri doğrultusunda özelleşmiştir (Embretson, 2000).

Güvenirliğin artırılması için KTK'da yapılan uygulamalardan birisi de test duyarlılığının artırılmasıdır. Testin duyarlılığının artırılması ile daha hassas ölçümler yapılabilmesine olanak sağlanmaktadır. KTK, madde sayısı ile güvenilirliğin doğrusal bir ilişki içerisinde olduğunu savunmaktadır. Yani madde sayısı arttıkça güvenilirlik de artmaktadır. Bu durum test için soru doyum noktasına ulaşana kadar devam eder. Ancak MTK için bireyin yetenek düzeyi ile ilgili en çok bilgiyi veren birkaç madde ile maksimum güvenilirlik sağlanabilmektedir. MTK'nın yapısı gereği doğru madde ile doğru bireyin eşleştirilmesi yapılabildiği için bireyin yetenek düzeyi hakkında en yüksek bilgi sağlayan maddelerin seçilmesi mümkündür. Böylece testin güvenilirliği, yetenek düzeyine bağlı olarak değişkenlik gösterebilmektedir (Baker, 2001; Hambleton, 1991). Bu sayede güvenilirlik daha az madde seçimi ile sağlanabilmektedir.

Bazı geniş ölçekli testlerde güvenilirliği sağlamak ve kopya riskinin önüne geçebilmek adına farklı test formları geliştirilmektedir (örneğin ALES, YDS). Bu şekilde geliştirilen formlardan elde edilen puanların karşılaştırılması gerekmektedir. KTK, farklı test formlarından elde edilen puanların karşılaştırılabilmesi için testlerin paralel olması varsayımını gerektirir. Ancak uygulamalar arasında testlerin paralel olmasını sağlamak oldukça güçtür. MTK'nın madde odaklı yaklaşımı ve farklı testlerde bulunan maddelerin aynı ölçek düzeyinde doğrudan karşılaştırılabilmesi testlerin doğrudan karşılaştırılmasına imkân tanır. MTK doğrusal olmayan yöntemler sayesinde testlerin paralellliği koşulunu gerektirmez (Baker & Kim, 2004; Kolen, 2004; Lord, 2012).

MTK'da madde parametreleri teorik olarak gruptan bağımsızdır ve değişmez kabul edilir. Yani madde gücüğü (b), madde ayırt ediciliği (a) ve şans (c) parametreleri farklı örneklemelerde uygulansa da gruba göre değişkenlik göstermez. Bu durum maddelerin farklı örneklemelerde ve koşullarda karşılaştırılmasına imkân sağlar. Ayrıca maddelerin farklı gruplarda herkes için aynı işlevi göstermesi testin adil davranmasına olanak tanır. Yeni geliştirilecek testlerde ve uyarlama çalışmalarında parametreler değişmez olduğu için geçerli ve güvenilir sonuçlar elde edilebilmektedir. MTK'nın bu avantajı sayesinde geçerliliği ve güvenilirliği kanıtlanmış geniş madde havuzları oluşturulabilir ve BOBUT uygulamalarında kullanılabilir. Farklı dil ve kültürlerde oluşturulmuş testlere ait madde parametrelerinin tutarlılığı değerlendirilebilmektedir (Baker, 2001; Camilli, 1994; Embretson & Reise, 2000; Hambleton, 1991).

Bireyin grup içerisindeki sıralamasını ve başarısını belirlemek için KTK'da dağılımın normallik göstermesi ve puanların standart puanlara dönüştürülerek bireyin grup içindeki yerinin belirlenmesi gerekmektedir. Bu uygulama, kriter referanslı ölçümler için anlamlı olmazken norm referanslı ölçümlerde kullanılabilir. Oysa MTK, kişi ve maddeyi aynı ölçek düzeyinde değerlendirdiği için yetenek düzeylerinin karşılaştırılması mümkündür. Normallik varsayımı sağlanmasa ve test puanları standart puanlara dönüştürülme dahi bireylerin yetenek düzeylerine göre başarı sıralamaları belirlenebilir (Cohen vd., 1996; De Ayala, 2013; Kleine, 2005).

KTK'ya göre normal dağılımın sağlanabilmesi için madde güçlüklerinin yaklaşık 0.50 civarında olması gerekir. Normal dağılım sağlandığında, gözlenen puan ile gerçek puan arasındaki ilişki daha anlamlı hale gelir. Ancak MTK, doğrusal olmayan lojistik ya da probit modeller kullandığı için veriler normal dağılım göstermese bile aralık veya oran düzeyinde

ölçekleme yapılabilir. Bu özellik, daha karmaşık analiz ve karşılaştırmalara olanak tanımaktadır (Allen & Yen, 2001; Crocker & Algina, 1986). Bununla birlikte, bazı araştırmalarda normallik varsayımı MTK için gerekli olmasa da, normallik sağlanmadığında yetenek kestirimlerinde sorunlar yaşanabileceği ortaya konmuştur (DeMars, 2010).

Karma madde formatında maddelerin ağırlıklandırılmasında ortaya çıkabilecek dengesizlikler, KTK'da toplam puanı olumsuz etkileyebilir. Buna karşılık MTK'da madde güçlüğü parametresine bağlı kalmaksızın kategori sayısı farklı olan maddeler bir arada kullanılabilir. Bu sayede kategori sayısından kaynaklanan etkiler bertaraf edilebilmektedir (Masters, 1982; van der Linden & Hambleton, 1997; Wilson, 2023).

KTK'da eşit aralık düzeyinde veri elde etmek zordur. Bu nedenle düşük yetenek düzeyine sahip bir bireyin puanını 1 artırması ile yüksek yetenek düzeyine sahip bir bireyin puanını 1 artırması aynı anlamı taşımamaktadır. Bu durum özellikle deneysel desenlerde farklı yetenek düzeylerinin etkilerini gözlemlemeyi zorlaştırmaktadır. Oysa MTK'da eşit aralık düzeyinde ölçek elde etmek daha kolaydır ve puan artışlarının anlamı daha açık bir şekilde yorumlanabilmektedir (Crocker & Algina, 1986; Embretson & Reise, 2000).

KTK ile MTK arasındaki çeşitli farklılıklara rağmen, her iki model de testi alan bireylerin verdiği yanıtlara odaklanmaktadır. Ayrıca, her iki modelde de testin yalnızca ölçülmesi hedeflenen özelliği değerlendirmesi bu özelliğin başka değişkenlerle karışmadan ölçülmesi ve bireylerin bir maddeye verdikleri yanıtların diğer maddelere verdikleri yanıtları etkilememesi gerekmektedir (Çıkrıkçı & Demirtaşlı, 1998).

MTK, KTK'nın sınırlılıkları nedeniyle 20. yüzyılın ortalarında geliştirilmiştir. KTK; test puanlarının hesaplanması, geçerlilik ve güvenirlik incelemeleri açısından önemli katkılar sunmuş olsa da bireylerin yeteneklerini ve test maddelerinin özelliklerini aynı ölçüm çerçevesinde ele alma konusunda yetersiz kalmıştır. MTK'nın temelleri Frederic Lord ve George Rasch tarafından 1950'li ve 1960'lı yıllarda atılmıştır. Lord, bireysel farklılıkların daha iyi anlaşılabilmesi amacıyla MTK'nın matematiksel yapısını oluştururken; Rasch, yalnızca madde güçlük parametresine dayanan ve 1 Parametrelili Lojistik Model (1PLM) olarak bilinen Rasch Modeli'ni ortaya koymuştur (Hambleton vd., 1991). Bu model, modern MTK'nın temel taşlarından biri haline gelmiştir.

1970'li yıllarda Birnbaum tarafından geliştirilen çok parametrelili modeller (2PLM ve 3PLM), MTK'nın teorik altyapısını genişletmiştir. Bu modeller, madde ayırt ediciliği ve şans parametresi gibi ek değişkenleri içermesi sayesinde bireylerin yetenek düzeyleri ile madde

özellikleri arasındaki ilişkiyi daha ayrıntılı biçimde açıklayabilmektedir (Embretson & Reise, 2000).

Bilgisayar teknolojisindeki gelişmelerle birlikte MTK'nın uygulamaları yaygınlaşmış ve daha uygulanabilir hale gelmiştir. Özellikle 1980'li yıllardan itibaren bilgisayarların yaygınlaşması karmaşık matematiksel hesaplamaların ve veri analizlerinin daha kısa sürede gerçekleştirilmesini mümkün kılmıştır. Bu gelişme, MTK'nın eğitim ve psikoloji gibi alanlarda geniş bir uygulama alanı bulmasını sağlamıştır (van der Linden, 2016). Bilgisayarlar sayesinde büyük ölçekli madde havuzları depolanabilmiş ve test geliştirme uzmanları önceden kalibrasyonu yapılmış maddelerden hızlı biçimde seçim yapabilmıştır (Hambleton, 1991).

MTK'nın istatistiksel olarak karmaşık tahminleme yöntemleri gerektirmesi bu hesaplamaların bilgisayar desteğiyle hızlı ve güvenilir şekilde yapılmasını önemli hale getirmiştir (De Ayala, 2013). Ayrıca veri analizinin görselleştirilmesi de kolaylaşmış; maddelerin yetenek düzeyine göre gösterdiği işlevlerin grafiksel olarak sunulduğu Madde Karakteristik Eğrileri (MKE) ve testin farklı yetenek düzeylerinde bilgi sağlama gücünü gösteren test bilgi fonksiyonu grafiklerinin kullanımı yaygınlaşmıştır (Reckase, 2009). Bu gelişmeler özellikle PISA ve TOEFL gibi geniş ölçekli ve standartlaştırılmış testlerde ya da psikolojik değerlendirmelerde, bireye uyarlanmış ve hassas ölçümler yapılmasına olanak tanıyan bilgisayar destekli testlerde MTK'nın tercih edilmesini sağlamıştır (Embretson & Reise, 2000).

MTK, ölçülmek istenen yetenek ile bir madde arasındaki ilişkiyi incelemeye yönelik geliştirilen matematiksel bir kuramdır. Bu kuram kapsamında oluşturulan modeller, bireyin yetenek düzeyine göre bir maddeyi doğru yanıtlayabilme olasılığını belirlemeyi amaçlayan lojistik temelli fonksiyonlara dayanmaktadır (Sünbül, 2020). MTK'nın temelinde yer alan Madde Karakteristik Eğrisi (MKE), bireyin yeteneği ile maddeye doğru yanıt verme olasılığı arasındaki ilişkiyi grafiksel olarak temsil eder. “S” şeklindeki bu eğri, MTK'nın anlaşılmasında ve yorumlanmasında kritik öneme sahiptir (Crocker & Algina, 2006b).

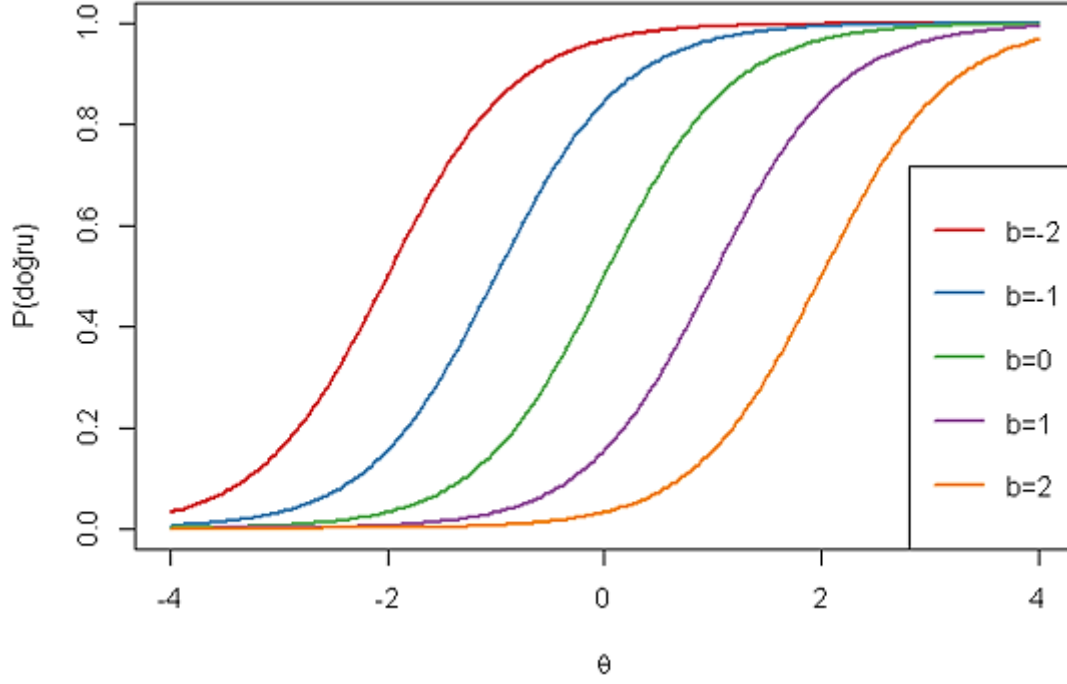
MTK'nın uygulanabilmesi için iki temel varsayım bulunmaktadır. Bunlardan ilki olan yerel bağımsızlık, bireylerin test maddelerine verdikleri yanıtların aynı yetenek düzeyinde istatistiksel olarak birbirinden bağımsız olması gerektiğini belirtir (Sünbül, 2020). İkinci varsayım olan tek boyutluluk ise, testin yalnızca tek bir yapıyı ölçmesi gerektiğini ifade eder. Bu koşul her ne kadar pratikte tam olarak sağlanmasa da başat bir faktörün ölçülmesi çoğunlukla yeterli kabul edilmektedir (Hambleton vd., 1991).

MTK modelleri, maddelerin puanlanma biçimine ve testin amacına bağlı olarak çeşitlenmektedir. Özellikle ikili puanlanan maddeler için geliştirilen modeller madde ayırt edicilik (a), madde güçlük (b), şans (c) ve üst asimptot (d) parametrelerine dayanmaktadır. Bu parametreler sırasıyla maddenin bireyler arasında ne derece ayırt edici olduğunu, hangi yetenek düzeyi için en bilgilendirici olduğunu, düşük yetenek düzeyindeki bireylerin maddeyi doğru yanıtlama olasılığını ve yüksek yetenek düzeyindeki bireylerin dikkatsizlik gibi nedenlerle maddeyi yanlış yanıtlama durumunu temsil eder (Baker, 2001; Weiss & Şahin, 2024).

Modelde hangi parametrelerin yer aldığına bağlı olarak MTK'da çeşitli modeller tanımlanmaktadır. Örneğin, yalnızca b parametresinin yer aldığı ve a parametresinin sabit kabul edildiği model 1PL modeli olarak adlandırılmaktadır. Bu modelde a parametresi 1'e sabitlenirse model Rasch Modeli olarak isimlendirilir. 1PL model ve Rasch model için kullanılan olasılık formülü;

$$P(\theta) = \frac{e^{\theta-b}}{1+e^{-1(\theta-b)}}$$

Bu formülde “b” madde güçlük parametresini, “ θ ” yeteneği, “e” matematiksel sabiti ifade eder ve değeri 2.718 olarak alınmaktadır. Model yalnızca madde güçlük parametresini içerdiği için, tüm maddelerin ayırt edicilik gücünün eşit olduğu varsayılır. Bu durum, özellikle maddeler arasında ayırt edicilik bakımından farklılıkların bulunduğu testlerde önemli bir sınırlılık oluşturmaktadır. MKE’ler incelendiğinde, 1PL modelde yer alan farklı maddelere ait eğrilerin birbirleriyle kesişmediği görülmektedir (Hambleton vd., 1991). Şekil 2.2’de görüldüğü üzere, $b = -2.0$ olan madde en kolay maddeyi, $b = 2.0$ olan madde ise en zor maddeyi temsil etmektedir.

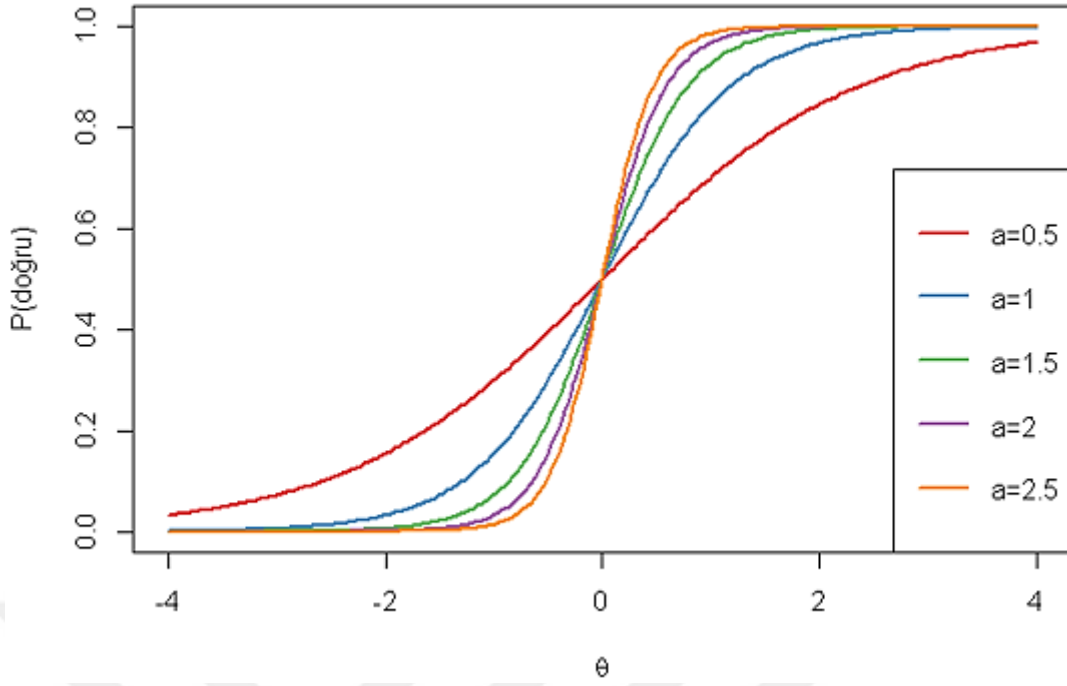


Şekil 2.2. 1PLM İçin MKE'ler

2PL modelde ise madde güçlük (b) ve madde ayırt edicilik (a) parametreleri modele dahil edilmekte, ancak şans parametresi (c) yer almamaktadır. Bu modele ait olasılık fonksiyonu:

$$P(\theta) = \frac{e^{a(\theta-b)}}{1+e^{-a(\theta-b)}}$$

Bu formül, 1PL modeline ek olarak madde ayırt edicilik parametresinin “a” dahil edilmesiyle oluşmaktadır. Ayırt edicilik parametresi modele eklendiğinde, maddelere ait karakteristik eğrilerin birbiriyle kesişmesi mümkün hale gelir. Ancak bu modelin parametre tahmini 1PL modele göre daha karmaşıktır ve daha büyük örneklemeler gerektirmektedir (Embretson & Reise, 2000). Şekil 2.3’te gösterildiği üzere, $a = .5$ olan madde en düşük ayırt ediciliğe, $a = 2.5$ olan madde ise en yüksek ayırt ediciliğe sahiptir. Şekilden de anlaşılabacağı gibi eğim parametresi olarak da bilinen ayırt edicilik değeri arttıkça MKE daha dik bir hal almakta ve madde bireyleri ayırt etmede daha hassas hale gelmektedir.

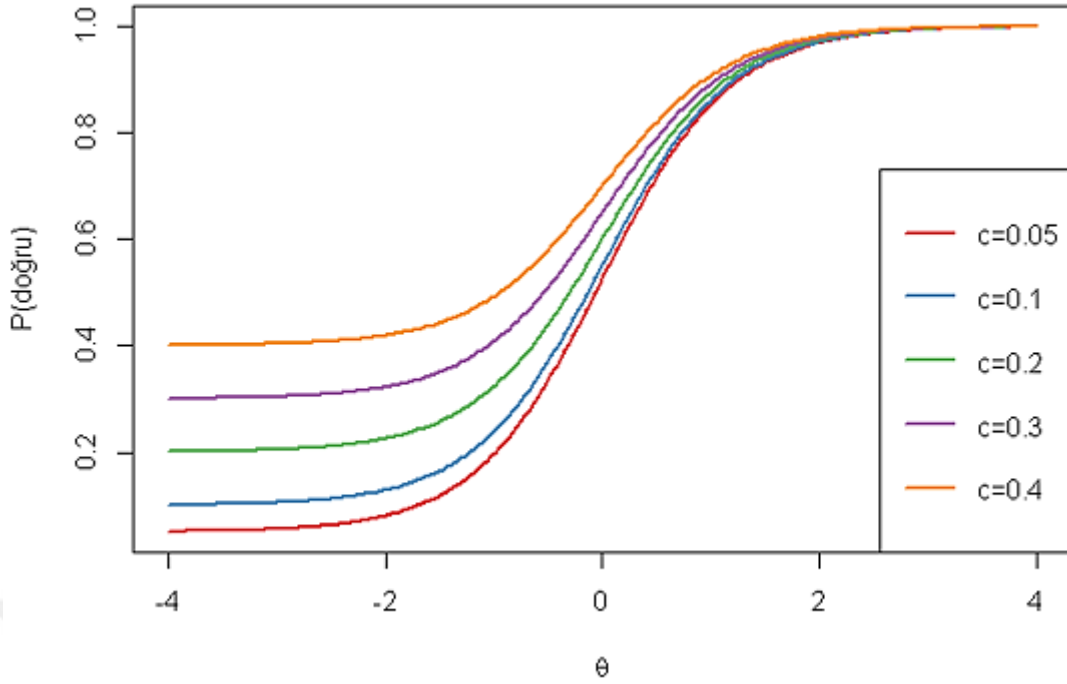


Şekil 2.3. 2PLM İçin MKE'ler

3PL modelinde madde güçlük parametresi (b), madde ayırt edicilik parametresi (a) ve düşük asimptot olarak adlandırılan şans parametreleri (c) birlikte kestirilmektedir. Bu modele ait olasılık fonksiyonu aşağıda verilmiştir:

$$P(\theta) = c + (1 - c) \frac{e^{a(\theta-b)}}{1 + e^{-a(\theta-b)}}$$

3PL model, 2PL modelin yapısına “c” parametresinin eklenmesiyle oluşturulmaktadır. Burada “c” parametresi, düşük yetenek düzeyine sahip bireylerin şans yolu ile doğru yanıt verme olasılığını temsil etmektedir. Bu parametrenin modele dahil edilmesiyle madde karakteristik eğrilerinin (MKE) başlangıç noktaları sıfırdan farklı seviyelerde olabilmektedir (Baker & Kim, 2017). Bu sayede özellikle düşük yetenek düzeyindeki bireyler için oluşabilecek yanlış modellemelerin önüne geçilmesi hedeflenmektedir. Ancak 3PL model, parametre tahmini açısından daha karmaşıktır ve daha büyük örneklem büyüklüklerine ihtiyaç duyar (Birnbaum, 1968). Şekil 2.4’te gösterildiği üzere, $c = .05$ olan madde daha düşük ve $c = .40$ olan madde ise daha yüksek bir başlangıç noktasından başlamaktadır. Bu durum şans parametresi arttıkça düşük yetenek düzeyindeki bireylerin doğru yanıt ulaşma olasılıklarının da arttığını ve dolayısıyla MKE’nin daha yüksek seviyelerden başladığını göstermektedir.



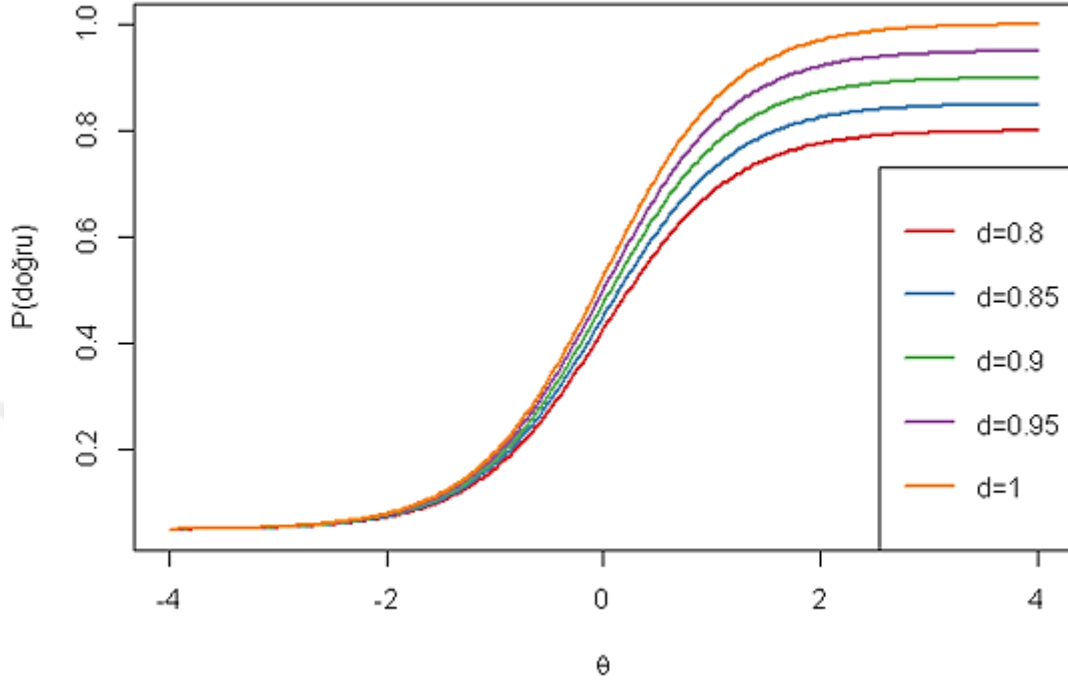
Şekil 2.4. 3PLM İin MKE'ler

4PL model ise MTK kapsamında geliştirilen daha yeni bir yaklaşımdır. Bu modelde, madde güçlük parametresi (b), madde ayırt edicilik parametresi (a), şans parametresi (c) ve üst asimptot parametresi (d) birlikte kestirilmektedir. 4PL model, yüksek yetenek düzeyine sahip bireylerin bazı maddeleri yanlış yanıtlayabileceği durumları da modelleyebilmesi bakımından önemlidir. Bu modele ait olasılık fonksiyonu aşağıda verilmiştir:

$$P(\theta) = c + (d - c) \frac{e^{a(\theta-b)}}{1 + e^{-a(\theta-b)}}$$

Bu formül 3PL modele üst asimptot parametresi “d” eklenerek elde edilmektedir. “d” parametresi, yüksek yetenek düzeyine sahip bireylerin maddeyi doğru yanıtlamada ulaşabilecekleri maksimum olasılığı ifade etmektedir (Veerkamp & Berger, 1997). Başka bir ifadeyle bu durum yetenek düzeyi yüksek bireylerin dikkatsizlik gibi dışsal nedenlerle maddeyi yanlış yanıtlamaları şeklinde yorumlanabilir. MTK, doğası gereği bilgisayar ortamında bireye uyarlanmış testler (BOBUT) uygulamalarında yaygın olarak kullanılan test kuramı olarak öne çıkmaktadır. BOBUT sistemlerinin başarısı büyük ölçüde MTK’nın sağladığı ölçme duyarlılığı, esneklik ve doğruluk özelliklerine dayanmaktadır. Şekil 2.5’te gösterildiği üzere 4PL modele ait madde karakteristik eğrileri incelendiğinde d parametresi

azaldıkça yüksek yetenek düzeyine sahip bireylerin maddeyi doğru yanıtlayabilme olasılığı da düşmektedir. Örneğin, $d = .80$ olan bir madde için bu bireylerin söz konusu maddeyi doğru cevaplandırma olasılığı en fazla %80 olarak değerlendirilmektedir.



Şekil 2.5. 4PLM İçin MKE'ler

2.3. BOBUT'un Bileşenleri

Eğitim ve psikoloji alanlarında bireylerin bilgi, beceri ve yeteneklerini değerlendirmek amacıyla uzun yıllar boyunca KTK kullanılmıştır. Ancak sabit formdaki testler, bireyler arasındaki yetenek düzeyi farklarını yeterince dikkate alamamaktadır (Weiss & Şahin, 2024). Örneğin düşük yetenek düzeyine sahip bireylere çok zor, yüksek yetenek düzeyine sahip bireylere ise çok kolay maddeler sunulabilmektedir. Bu durum, testin ölçme doğruluğunu azaltmakta ve bireyler açısından adil olmayan değerlendirmelere yol açabilmektedir (van der Linden & Glas, 2010).

BOBUT uygulamaları, bu sınırlılıkları aşmak amacıyla geliştirilmiştir. BOBUT, her bireyin yetenek düzeyine uygun maddelerin seçilmesini sağlayarak ölçme sürecini bireyselleştirmekte ve daha hassas kestirimlere olanak tanımaktadır. Bu yöntem MTK'ya dayanmakta ve sunduğu parametrik yapı sayesinde bireylerin yetenek düzeylerinin daha doğru biçimde kestirilmesini mümkün kılmaktadır. Günümüzde bilgisayar teknolojilerindeki

gelişmelerle birlikte BOBUT uygulamaları özellikle geniş ölçekli testlerde yaygın olarak kullanılmaktadır (Hambleton vd., 1991; van der Linden & Glas, 2010).

BOBUT algoritmaları, bireylerin yetenek düzeyine uygun madde seçimi yaparak test sürecini bireyselleştirmektedir. Bu yaklaşım, test süresini kısaltmakta ve bireylerin test deneyimini olumlu yönde etkilemektedir. Ayrıca BOBUT'un dinamik ve esnek yapısı sayesinde bireyler üzerindeki bilişsel yük hafifletilmekte ölçme doğruluğu ve güvenilirliği artırılmaktadır. Bu bölümde, BOBUT uygulamalarının temel adımları ayrıntılı biçimde ele alınarak testin bireyselleştirilmesini sağlayan süreçler açıklanacaktır.

2.3.1. İlk Madde Seçimi

BOBUT uygulamalarında MTK modelinin belirlenmesinden sonraki aşama bireylerin alacağı ilk maddenin seçilmesidir. İlk madde seçimi, ölçme sürecinde yüksek etkiye sahip olmasa da testi alan bireyler açısından psikolojik etkiler doğurabileceği unutulmamalıdır. Embretson ve Reise (2000) ilk maddenin çok kolay olması durumunda bireyin rehavete kapılabileceğini ve bunun dikkatsizliğe yol açabileceğini çok zor bir ilk madde durumunda ise stres kaynaklı motivasyon kaybı ve başarısızlık hissini gelişebileceğini belirtmiştir. Bu tür durumların önüne geçilmesi için ilk maddenin seçimi testin güvenilirliği açısından kritik öneme sahiptir.

Testin ilk maddesinin belirlenmesinde uzmanlar çeşitli yaklaşımlar benimsemektedir. Eğer test grubu hakkında önsel bilgi mevcutsa ilk madde grubun yetenek düzeyine göre seçilebilir (Thissen & Wainer, 2001). Bir diğer yöntem ise asıl BOBUT uygulaması öncesinde bireylere pilot bir madde seti uygulanarak bireylerin yetenek düzeylerine göre ilk maddenin seçilmesidir (van der Linden & Glas, 2010). Ancak kimi durumlarda test geliştirici, grup hakkında önsel bilgiye sahip olmayabilir ya da pilot uygulama yapma olanağı bulunmayabilir. Bu gibi durumlarda orta güçlük düzeyine sahip tek bir madde ya da bu düzeydeki maddeler arasından rastgele seçilen bir madde ile sürece başlanabilmektedir (Eggen, 1999).

2.3.2. Yetenek Kestirimi

İlk maddeye bireyin yanıt vermesinin ardından doğru ya da yanlış yanıtına bağlı olarak yetenek kestirim süreci başlatılır. Bu süreçte test geliştirme uzmanları, yaygın olarak MLE (Maximum Likelihood Estimation) ve Bayesci yöntemleri kullanmaktadır. Ancak testin

niteliği ve uygulama bağlamına göre uzmanlar farklı yetenek kestirim yaklaşımlarından da yararlanabilmektedir.

2.3.2.1. En Çok Olabilirlik Yöntemi(Maximum Likelihood Estimation- MLE)

MLE, testi alan bireylerin yetenek düzeylerini tahmin etmek amacıyla olabilirlik fonksiyonunu maksimum yapan θ değerini arar. Bu yöntem bireyin verdiği cevaplara göre en olası yetenek düzeyini belirlemeyi amaçlamaktadır. MLE eşitliği Newton-Raphson, Fisher Skorlama ve Beklenti Maksimizasyonu gibi iteratif yöntemlerle çözümlenmektedir. MLE için olabilirlik fonksiyonu aşağıdaki gibidir:

$$L(\theta) = \prod_{i=1}^n P_i(k_i|\theta)$$

Bu eşitlikte θ yetenek düzeyini; $P_i(k_i|\theta)$; Testi alan bireyin θ yetenek düzeyinde i . maddeye doğru cevap verme olasılığını; n ise madde sayısını temsil etmektedir. Bu eşitlik, optimizasyon algoritmaları kullanılarak çözüme ulaştırılmaktadır (Lord, 2012). MLE yöntemi tarafsız (unbiased) bir kestirim sağlaması nedeniyle yaygın olarak tercih edilmektedir. Ancak, bireyin tüm sorulara doğru veya tümüne yanlış yanıt verdiği mükemmel cevap örüntüsü durumlarında işlem yapılamaz ve kestirim yapılamaz. Ayrıca madde havuzunun dar olması durumunda kestirim doğruluğu düşebilmektedir (Baker, 1998).

2.3.2.2. Bayesci Yöntemler

Bayesci yöntemler, MLE'den farklı olarak önsel (prior) ve sonsal (posterior) dağılımları dikkate alarak yetenek düzeyini tahmin etmektedir. Bu yöntemlerde bireyin yetenek düzeyine ilişkin önsel bir dağılım varsayılır ve bireyin verdiği yanıtlar doğrultusunda bu dağılım sürekli olarak güncellenir. Böylece yetenek düzeyi tahmini bireyin test sürecindeki performansına bağlı olarak dinamik biçimde şekillenir.

Bayesci yöntemlerin önemli bir avantajı bireyin tüm sorulara doğru ya da tümüne yanlış yanıt verdiği durumlarda da yetenek kestirimi yapılabilmesidir. Bu özellik mükemmel cevap örüntülerinde işlem yapamayan MLE'ye göre Bayesci yöntemleri daha esnek ve uygulanabilir hale getirmektedir.

2.3.2.2.1. Maximum a Priori (MAP)

MAP kestirim, testi alan bireyin testi almadan önceki önsel bilgisine dayalı olarak maksimum tahmin değerine ulaşmaya çalışmaktadır. Başlangıçta aritmetik ortalaması 0 ve standart sapması 1 olan ve normal dağılımdan gelen önsel bilgi dağılımına göre en olası tahmin değeri bulunur. Daha sonra bireyin maddelere verdiği doğru veya yanlış yanıtlara göre oluşan olabilirlik fonksiyonu kullanarak tahmin süreci devam eder (Bock & Aitkin, 1981).

2.3.2.2.2. Expected a Posteriori (EAP)

EAP kestirimi MAP kestiriminden farklı olarak sonsal dağılıma göre yetenek kestirimi yapmaktadır. MAP kestirimi tek bir maksimum değeri hedeflerken EAP kestirimi sonsal dağılımın aritmetik ortalamasına göre kestirim yapmaktadır. EAP kestirimi bu süreçte;

$$\theta_{EAP} = \int_{-\infty}^{\infty} \theta \cdot P(\theta|k) d\theta$$

Eşitliği ile çözülmektedir. Eşitlikte θ_{EAP} sonsal dağılımın ortalamasını, $P(\theta|k)$ sonsal dağılımı, $d\theta$ entegrasyon işlemini ifade etmektedir (Ho, 2010). EAP test sürecinde yanıt dengesizliği olan durumlarda daha dayanıklıdır ve az sayıda madde yanıtlandığı durumlarda dahi kestirim yapabilmesinden dolayı test geliştirme uzmanları tarafından yaygın bir şekilde kullanılmaktadır (Hambleton & Swaminathan, 2013).

2.3.3. Madde Seçme Yöntemleri

Test geliştirme uzmanının, araştırmanın amacına uygun olarak belirlediği yetenek kestirim yönteminden sonra testi alan bireyin doğru ya da yanlış yanıt vermesi durumunda bir sonraki sorunun seçimi BOBUT uygulamalarının en kritik aşamalarından biridir. Bu süreç testin bireyselleştirilmesini sağlayan temel unsurlardan biri olarak değerlendirilir.

Test geliştirme uzmanları uygulamanın amacına, hedeflenen ölçme hassasiyetine ve test süresine göre farklı madde seçme algoritmalarını tercih edebilmektedir. BOBUT uygulamalarında en sık kullanılan madde seçme yöntemleri arasında Maksimum Fisher Bilgi (MFB), Kullback-Leibler (KL) Bilgisi, Aralık Bilgisi Ölçütü (ABÖ), En Uygun b Değeri Seçimi ve Bayesci Yöntemler yer almaktadır. Bu yöntemler bireyin yetenek düzeyine en uygun maddeyi seçerek testin ölçme duyarlılığını ve verimliliğini artırmayı amaçlar.

2.3.3.1. Maksimum Fisher Bilgi(MFB)

Maksimum Fisher Bilgi ölçütü testi alan bireyin çözdüğü her bir sorudan sonra bireyin yetenek düzeyi ile ilgili en yüksek bilgiyi verecek maddenin seçilmesi temeline dayanmaktadır. Ancak uygulamanın başında birey ile ilgili önsel bilgi mevcut değilse seçilecek sorunun yanlışlık oluşturma ihtimali olabilir, madde seçim algoritması yüksek ayırt ediciliğe sahip maddeleri seçmeye meyillidir (Han, 2009). MFB eşitliği;

$$I_i(\theta) = \frac{a_i^2 \cdot e^{-Da_i(\theta-b_i)}}{(1 + e^{-Da_i(\theta-b_i)})^2}$$

Şeklinde dir. Formülde $I_i(\theta)$ yetenek düzeyine göre fisher bilgi fonksiyonunu, a_i madde ayırt edicilik parametresini, b_i madde güçlük parametresini, D ölçekleme sabitini, $e^{-Da_i(\theta-b_i)}$ ekspansiyel terimi, $(1 + e^{-Da_i(\theta-b_i)})^2$ normalizasyon faktörünü ifade etmektedir (Birnbbaum, 1968; Veerkamp ve Berger 1997).

2.3.3.2. Bayesci Yöntemler

Bayesci yöntemler, bireyin her yanıtladığı madde sonrasında yetenek kestiriminin yeniden hesaplanmasını ve buna bağlı olarak en doğru kestirimi sağlayacak maddenin seçilmesini amaçlamaktadır. Bu süreçte bireyin önceki sorulara verdiği yanıtlarla önsel dağılımın olabilirlik fonksiyonu çarpılarak güncellenir ve bu sayede her yeni madde için daha güvenilir bir tahmin elde edilir. Güncellenen sonsal dağılım temel alınarak birey hakkında en fazla bilgi sağlayacak madde seçilerek test süreci devam eder.

Bu özelliği sayesinde Bayesci yöntemler, test boyunca doğru ve sürekli güncellenen kestirimler yapabilme olanağı sunar. Ayrıca testin az sayıda maddeden oluştuğu durumlarda dahi güvenilir sonuçlar elde edilebilmektedir. Ancak her maddenin ardından sonsal dağılımın yeniden güncellenmesi özellikle geniş ölçekli test uygulamalarında işlem yükünü artırmakta ve Bayesci yöntemleri daha az uygulanabilir hale getirebilmektedir (Van Der Linden & Pashley, 2009).

2.3.3.3. Aralık Bilgi Ölçütü(ABÖ)

Veerkamp ve Berger (1997) tarafından ortaya atılmış olan ABÖ, MFB'ye alternatif olarak geliştirilmiştir. Her bir madde için bilgi fonksiyonunu belirli bir güven aralığında ortalayarak daha dengeli bir madde seçimi yapılmasına imkân sağlamaktadır. ABÖ için

bireyin yetenek düzeyi etrafında bir güven aralığı belirlenir ve bu aralık boyunca sağlanan toplam bilgiye göre madde seçimi yapılır. ABÖ eşitliği

$$I_i = \int_{\hat{\theta}_L}^{\hat{\theta}_R} I(\theta) d\theta$$

Şeklinde ve $\hat{\theta}_R$ ve $\hat{\theta}_L$ tahmin edilen yetenek seviyesinin alt ve üst sınırını gösterir. $I(\theta)$ θ seviyesindeki bilgi fonksiyonudur. Kestirilen yetenek seviyesinin güven aralığında ele alınması standart hatayı azaltır ve test sürecinin başında daha dengeli madde seçimi yapılmasına imkan tanır. Ancak güven aralığı geniş olduğu için bilgi miktarı düşük olabilmektedir ve bu durumda seçilen maddenin bireyin yetenek kestirimi için sağladığı bilgiyi düşürebilmektedir (Chang & Ying, 1999).

2.3.4. Test Sonlandırma Kuralı

BOBUT uygulamalarında test sürecinin kontrolünü sağlamak ve süreyi yönetmek amacıyla kullanılan test sonlandırma kuralları iki temel gruba ayrılmaktadır: sabit test uzunluğu kuralı ve değişken test uzunluğu yöntemleri (Embretson & Reise, 2000).

2.3.4.1. Sabit Test Uzunluğu Kuralı

BOBUT kapsamında en temel test sonlandırma stratejilerinden biri sabit test uzunluğu kuralıdır. Bu yöntemde teste başlamadan önce her bireyin yanıtlayacağı madde sayısı önceden belirlenmiştir. Örneğin 20 sorudan oluşan bir BOBUT uygulamasında tüm bireyler 20 soruyu yanıtlar ve test bu noktada sona erer (van der Linden & Pashley, 2010). Bu yaklaşım özellikle klasik testlerin bilgisayar ortamına aktarılması sürecinde sıkça kullanılmaktadır (Lilley, 2007; Lord, 2012).

Sabit uzunlukta uygulanan BOBUT sistemleri zaman yönetimi, içerik kontrolü ve madde havuzu yönetimi gibi pratik açılardan çeşitli avantajlar sunmaktadır (Vale & Weiss, 1975). Ayrıca test için sabit süreli oturumlar planlanabilmekte ve böylece testin geçerliliği ile güvenirliliği üzerine öngörüler yapılabilmektedir (Parshall vd., 2002). Bu yöntemin uygulandığı durumlarda test yalnızca madde seçimi açısından bireyselleştirilmektedir (Kingsbury & Houser, 1993).

Arařtırmalar, sabit uzunlukta testlerin bireyler arasında farklı test bilgi fonksiyonlarına neden olabileceğini ve bazı bireylerde yetenek kestirimlerinin daha düşük güvenilirlikle sonuçlanabileceğini göstermektedir (Wainer vd., 2000).

2.3.4.2. Deęişken Test Uzunluęu Yöntemleri

BOBUT uygulamalarında test uzunluęunun bireyin performansına göre belirlendięi yaklaşımlar “deęişken test uzunluęu yöntemleri” olarak adlandırılmaktadır. Bu yöntemler, testin bireyin yetenek düzeyine göre erken veya geç sonlandırılmasına olanak tanıyarak ölçme doęruluęunu artırmakta aynı zamanda test süresi ile madde sayısını bireyin performansına uyarlamaktadır (Shapiro & Gebhardt, 2012). Sabit test uzunluęu yöntemi, farklı yetenek düzeylerindeki bireyleri yeterince ayırt edemezken deęişken test uzunluęu bu sınırlılıęı büyük ölçüde ortadan kaldırmaktadır (Vale & Weiss, 1975).

Deęişken test uzunluęu, ölçme hassasiyeti ile minimum madde kullanımı arasında denge kurmak açısından etkili stratejiler sunar. Bireyin yetenek düzeyi netleřtięinde testin sonlandırılması hem sürenin kısaltılmasını hem de bireyin daha az yorulmasını saęlar (Parshall vd., 2002). Aynı zamanda gereksiz madde sunumunun önlenmesi, madde havuzunun güvenlięini de desteklemektedir (Wainer vd., 2000). BOBUT uygulamalarında en sık kullanılan deęişken test uzunluęu stratejileri standart hata temelli ve kesme puanı temelli sonlandırma yöntemleridir (Reckase & Reckase, 2009).

2.3.4.2.1. Standart Hata Temelli Test Sonlandırma

Standart hata temelli test sonlandırma, bireyin yetenek düzeyine iliřkin kestirimin doęruluęunun esas alındıęı bir yaklaşımdır. Bu yöntemde test, bireyin yetenek tahminine ait standart hata deęeri önceden belirlenen bir eřik düzeyinin (örneğin .30 ya da .40) altına düřtüęünde sonlandırılır (Weiss, 1982; Weiss & řahin, 2024).

Yetenek kestirimi ve buna iliřkin standart hata her madde sonrasında güncellenmektedir. Belirlenen eřik deęerin altına inildięinde test sonlandırılır ve bu da ölçme sürecinin belirli bir güvenilirlięe ulařtıęının göstergesidir (van der Linden & Pashley, 2010). Bu yöntem sayesinde testin süresi ve madde sayısı bireyin performansına uyarlanır (Sands vd., 1997). Ayrıca düşük ya da yüksek yetenekli bireylerde testin kısa sürede tamamlanabilmesi ölçme verimlilięini artırırken madde havuzunun korunmasına da katkı

sağlar (Parshall vd., 2002). Özetle bu yöntem bireysel farklılıklara duyarlı, ekonomik ve psikometrik açıdan güçlü bir test sonlandırma stratejisidir.

2.3.4.2.2. Kesme Puanı Temelli Test Sonlandırma

Kesme puanı temelli test sonlandırma, bireyin yetenek kestiriminin önceden belirlenmiş bir yeterlilik sınırına ulaşp ulaşmadığına göre testin sonlandırılmasını sağlayan bir yöntemdir. Birey belirli bir başarı eşiğini geçtiğinde veya altında kaldığında test sonlandırılır (Weiss & Şahin, 2024). Bu yöntem sayesinde bireylerin yeterli ya da yetersiz olarak daha hızlı ve etkin biçimde sınıflandırılması mümkündür (van der Linden & Glas, 2000). Aynı zamanda gereksiz madde sunumu engellenerek madde havuzu korunur ve bireyin bilişsel yükü azaltılır.

Yöntemin en önemli avantajlarından biri karar odaklı değerlendirmelerde test süresini ve madde sayısını azaltma olanağı sunmasıdır (Sands vd., 1997). Ayrıca bu yöntem uyarlanabilir karar algoritmalarıyla entegre edilerek daha karmaşık sınıflandırmalara da olanak tanımaktadır. Örneğin bireyin performansı “yetersiz”, “temel” veya “yeterli” gibi birden fazla düzeyde kategorize edilebilir (Aybek, 2016). Bu tür uygulamalarda karar belirsizliğini azaltmak için karar güven aralıkları da kullanılabilir (Stocking & Lewis, 1995). Sonuç olarak kesme puanı temelli sonlandırma yalnızca test süresini kısaltmaz aynı zamanda eğitimsel ve mesleki yeterliliklerin sınıflandırılmasında etkili, esnek ve ekonomik bir çözüm sunar.

2.4. Kullanıcı Deneyimi ve Eğitim Teknolojisi

2.4.1. Kullanıcı Deneyimi Kavramı

Kullanıcı deneyimi bireylerin bir ürün, sistem veya hizmetle etkileşimlerinden doğan tüm algılarını, duygularını ve tepkilerini içeren çok boyutlu bir kavramdır. Bu deneyim verimlilik, memnuniyet, öğrenme kolaylığı gibi olumlu hislerin yanı sıra stres, hayal kırıklığı ve kaygı gibi olumsuz duyguları da kapsamaktadır (Ngadiman vd., 2021). Eğitim teknolojileri bağlamında kullanıcı deneyimi öğrencilerin bir öğrenme platformu ya da dijital uygulama ile etkileşimleri sırasında yaşadıkları duygusal, bilişsel ve işlevsel tepkileri ifade eder. Bu bağlamda literatürde öğrenme etkililiği, memnuniyet düzeyi ve hedefe ulaşma kolaylığı gibi konular ele alınmaktadır (Tullis & Albert, 2013; Weichbroth, 2020). Kullanıcı deneyimi

kullanılabilirlik (usability), etkililik (effectiveness), memnuniyet (satisfaction), bağlılık (engagement) ve öğrenme kolaylığı (learnability) gibi temel bileşenler çerçevesinde incelenmektedir (Stewart, 2009).

Kullanılabilirlik (Usability): Bir sistemin kullanıcılar tarafından ne derece kolay öğrenilip kullanılabildiğini ve hatırlanabildiğini ifade eder. Eğitim teknolojilerinde kullanılabilirlik bireylerin öğrenme süreçlerinin olumlu ilerleyebilmesi için materyallerin anlaşılır ve işlevsel olmasıyla ilişkilidir. Bu nedenle kullanılabilirlik testleri sistemin hedef kitleye uygunluğunu belirlemede kritik öneme sahiptir (Benyon, 2014). Örneğin BOBUT sistemlerinde kullanıcıların soru ekranında rahatça gezinmesi, talimatların yalın bir dille anlaşılır şekilde verilmesi ve metinlerin okunabilirliği oldukça önemlidir. Ayrıca tasarlanan BOBUT uygulamalarının dikkat dağıtmayacak şekilde tasarlanması gerekmektedir (Economides, 2020).

Etkililik (Effectiveness): Kullanıcıların bir sistemi kullanarak belirlenen hedeflere ulaşım ulaşılmadıklarını ifade eder. Eğitim teknolojilerinde bu, öğrenme materyallerinin anlaşılabilirliği ya da testlerin güvenilirliği ile doğrudan ilişkilidir. Örneğin BOBUT uygulamalarında doğru ölçme teknikleri ve bireye uyarlanmış algoritmalar, öğrencilerin gerçek öğrenme düzeylerini doğru biçimde yansıtmak açısından önem taşır (van der Linden & Glas, 2000). Etkililik aynı zamanda teknolojinin amacına ulaşım ulaşmadığını test eden temel bir ölçüt olarak kabul edilmektedir (Benyon, 2014).

Memnuniyet (Satisfaction): Kullanıcıların bir sistemle etkileşimlerinden duydukları tatmini ifade eder. Eğitim teknolojilerinde memnuniyet yalnızca teknik altyapının kalitesiyle değil aynı zamanda görsel tasarım ve kullanıcıya sunulan destekle de ilgilidir (Tullis & Albert, 2013). Örneğin bir BOBUT uygulamasında öğrencilere sağlanan motive edici ve anlamlı geribildirimler, memnuniyet düzeyini artırabilir (Economides, 2020; Vollmeyer & Rheinberg, 2005). Araştırmalar memnuniyetin yüksek olduğu sistemlerin kullanıcılar tarafından daha sık tercih edildiğini göstermektedir (Akhtar & Kovacs, 2024).

Bağlılık (Engagement): Kullanıcıların bir sistemle sürekli etkileşimde bulunma isteğini ifade eder. Eğitim teknolojilerinde bağlılık sayesinde öğrencilerin öğrenme sürecine aktif katılımı artmaktadır. Bağlılığı artırmak amacıyla oyunlaştırma, rozetler, başarı göstergeleri ve kişiselleştirilmiş hedefler gibi öğeler yaygın olarak kullanılmaktadır (Deterding vd., 2023; Hamari vd., 2014). BOBUT uygulamalarında bağlılığı artırmak için etkileşimli ve eğlenceli öğelerin yanı sıra gerçek zamanlı geribildirim ve ödül sistemleri de tercih edilmektedir.

Örneğin başarı düzeyine göre rozet kazanımı ya da performansa dayalı puan tablolarının sunulması, öğrencilerde rekabet duygusu oluşturarak teste katılımı teşvik edebilir (Sun & Hsu, 2013).

Öğrenme Kolaylığı (Learnability): Bir sistemin ilk kullanımda ne ölçüde anlaşılır ve erişilebilir olduğunu ifade eder. Kullanıcı dostu bir arayüz ve sade tasarım, kullanıcıların sistemi kolay ve hızlı biçimde öğrenmelerini sağlar (Nielsen, 1993). Eğitim teknolojilerinde farklı dijital yeterlilik düzeylerine sahip bireylerin kısa bir kılavuzla sistemi hızla öğrenebilmesi hedeflenmelidir (Norman, 2013). Araştırmalar öğrenme kolaylığı yüksek olan sistemlerin daha geniş kullanıcı kitleleri tarafından kabul gördüğünü ortaya koymaktadır. Örneğin bireyselleştirilmiş öğrenme sistemlerinde net talimatlar, sade tasarımlar ve rehber materyaller öğrenmeyi kolaylaştırabilmektedir (Clark & Mayer, 2016). BOBUT uygulamalarında öğrenme kolaylığını artırmak için bilgi yapısı, görsel hiyerarşi ve hızlı erişim araçları dikkatle tasarlanmalıdır (Shneiderman vd., 2016).

2.4.2. Eğitim Teknolojilerinde Kullanıcı Deneyiminin Önemi

Günümüzde hızla gelişen eğitim teknolojileri, bireylerin öğrenme süreçlerinin dijitalleşmesine neden olmuştur. Bu teknolojilerin etkililiğinde kullanıcı deneyimi en belirleyici faktörlerden biridir. Kullanıcı deneyimini merkeze alan tasarımlar, eğitim teknolojilerinin benimsenmesini kolaylaştırmakta ve bireylerin motivasyonunu artırmaktadır.

Kullanıcı deneyiminin iyileştirilmesi bireylerin öğrenme süreçlerine katılımını olumlu yönde etkilerken motivasyonu da desteklemektedir. Özellikle oyunlaştırma unsurlarının genç bireylerin katılım düzeyini artırdığı bilinmektedir (Deterding vd., 2011). Araştırmalar, kullanıcıların eğlenceli ve anlamlı deneyimlerle öğrenme süreçlerine daha aktif katıldığını göstermektedir. Örneğin BOBUT gibi bireye uyarlanmış test sistemlerinde öğrencilerin beceri seviyelerine uygun maddelerle karşılaşmaları, motivasyonlarını ve etkileşimlerini artırabilmektedir (Hamari vd., 2014; Sun & Hsu, 2013).

Kullanıcı dostu sistemler, yalnızca öğrenciler için değil öğretmenler için de eğitim teknolojilerinin benimsenmesini kolaylaştırır. Öğrenci başarısının kolayca izlenmesini ve raporlanmasını sağlayan sistemler, öğretim süreçlerini daha verimli hale getirmektedir (Norman, 2013). Öğrenciler açısından ise kullanıcı dostu arayüzler ve net talimatlar, öğrenme süreçlerini daha anlaşılır ve erişilebilir kılmaktadır. Örneğin yeni geliştirilen bir BOBUT uygulamasında kısa bir tanıtım videosu ile sistemin öğrenilebilmesi, kullanıcıların teknolojiye

olan kabulünü artırabilir (Nielsen, 1993). Öğretmenlerin sistemi kolaylıkla kullanabilmesi ise öğrenci farkındalığını artırarak iletişimi güçlendirmektedir (Shneiderman vd., 2016).

Eğitim teknolojileri, ölçme ve değerlendirme süreçlerini optimize edebilmek için kullanıcı deneyimini temel alan analiz olanakları sunmaktadır. BOBUT sistemlerindeki bireyselleştirilmiş algoritmalar, bireysel öğrenme ihtiyaçlarına göre özelleştirilmiş testler sunarak ölçme doğruluğunu artırmaktadır (Wainer, 2000). Kullanıcı deneyimini güçlendiren bir diğer unsur da geribildirim mekanizmalarıdır. Anlamlı ve zamanında sunulan geribildirimler, bireysel öğrenme hedeflerine ulaşmada ve eksik öğrenmeleri gidermede önemli rol oynamaktadır (Hattie & Timperley, 2007).

Kullanıcı deneyimi, eğitim teknolojilerinin benimsenmesinde en kritik unsurlardan biridir. Kullanıcı dostu olmayan sistemler, öğrenciler ve öğretmenler tarafından daha az tercih edilirken sezgisel ve erişilebilir tasarımlar daha yaygın benimsenmektedir (Rogers, 2003). Ayrıca kullanılan sistemlerin sürdürülebilirliği doğrudan kullanıcı deneyiminin kalitesiyle ilişkilidir. Karmaşık ve kullanıcıyı zorlayan tasarımlar kullanıcıları sistemden uzaklaştırırken, sade ve işlevsel tasarımlar kullanıcıların etkileşimlerini sürdürmesini sağlamaktadır (Clark & Mayer, 2016).

Etkili bir kullanıcı deneyimi, öğrenme çıktılarının kalitesini artırabilir. Öğrencilerin sürece daha aktif katılımı bilgiyi anlamalarını ve uygulamalarını kolaylaştırmaktadır (Hattie & Yates, 2014). Eğitim teknolojilerinde kullanılan etkileşimli tasarımlar ve kişiselleştirilmiş içerikler, öğrenme süreçlerini bireyselleştirerek daha iyi sonuçlar elde edilmesini mümkün kılar (Kay vd., 2019). Kullanıcı deneyimi geribildirim sistemi, test süresi-performans ilişkisi ve arayüz tasarımı gibi değişkenlerin etkili bir bütünlük içinde sunulmasıyla şekillenmektedir.

Geribildirim, öğretim süreçlerinde bireyleri yönlendirme açısından kritik öneme sahiptir ve etkili sunulduğunda kullanıcı başarısını artırır. Azevedo ve Bernard'ın (1995) yürüttüğü meta-analiz çalışmasına göre bilgisayar destekli öğrenmede anlık geribildirimlerin öğrenme çıktılarında önemli iyileşmelere yol açtığı görülmüştür. Araştırmalar, özellikle biçimlendirici ve teşvik edici olumlu geribildirimlerin öğrenci motivasyonunu artırdığını göstermektedir. BOBUT uygulamalarında kişiye özel sunulan geribildirimler, testin adil ve bireysel olduğu algısını güçlendirmektedir (Gasperson, 2014; Vispoel, 1998). Ayrıca yazılı geribildirimlerde açıklamaların detaylı ve anlaşılır olması hataların fark edilip düzeltilmesini kolaylaştırmaktadır (Paulos & Mahony, 2008).

Testi alan bireylerin dikkat süreleri ile test süresi arasındaki uyum performans üzerinde doğrudan etkilidir. BOBUT sistemlerinde test süresinin bireysel yeteneklere göre uyarlanabilmesi başarıyı ve memnuniyeti artırmaktadır (Wu vd., 2017). Örneğin testin daha kısa sürede tamamlanması, test kaygısını azaltarak daha iyi performansa imkân sağlayabilmektedir (Vispoel, 1998). Ayrıca anlık geribildirim mekanizmaları ve zaman optimizasyonu, özellikle matematik alanındaki öğrenme süreçlerinde kullanıcı motivasyonunu artırmaktadır (Wu vd., 2017).

Kullanıcı deneyimini etkileyen bir diğer bileşen ise arayüz tasarımıdır. Kullanıcı dostu bir arayüz, bireylerin test içeriğini daha iyi anlamalarına ve sürece odaklanmalarına yardımcı olur. Şenel'in (2017) görme engelli bireylerle yaptığı çalışmada erişilebilirlik esasına göre tasarlanan bir arayüzün motivasyonu artırdığı ve bağımsızlık hissini güçlendirdiği saptanmıştır. BOBUT uygulamalarının tasarımında görsel hiyerarşiye uyulması ve sezgisel navigasyon sistemlerinin oluşturulması kullanıcıların içerikle daha etkili etkileşim kurmasını sağlamaktadır. Yazılı geribildirimlerde de görsel ve metinsel öğelerin net olması kullanıcıların anlamasını kolaylaştırmakta ve geribildirim etkinliğini artırmaktadır (Lee, 2008).

2.5. Geribildirim ve Test Tasarımı

Geribildirim, bireylerin eğitim-öğretim sürecindeki performans ve öğrenme çıktılarına ilişkin davranışlarına ya da sonuçlarına yönelik sunulan bilgi ve değerlendirmelerdir. Geribildirimler öğrencilerin hatalarını düzeltmelerine, performanslarını geliştirmelerine ve karşılaşılabilecekleri problemlere karşı daha etkili stratejiler geliştirmelerine rehberlik etmektedir. Etkili geribildirim yalnızca doğru veya yanlışın belirlenmesiyle sınırlı kalmayıp bu doğrular ve yanlışlar üzerinden yeni bir öğrenme sürecinin başlatılmasını da kapsamaktadır (Hattie & Timperley, 2007; Jansen vd., 2025).

Özellikle bireyselleştirilmiş ve öğrenci tarafından anlamlı bulunan geribildirimlerin görev bilincini artırdığı ve öğrenme motivasyonunu desteklediği bilinmektedir (Burgers vd., 2015; Gan & Wang, 2025). Ayrıca olumlu duygusal tepkilere neden olan geribildirimlerin öğrencilerin öğrenmeye yönelik ilgilerini sürdürmelerine katkı sağladığı görülmüştür (Kuklick & Lindner, 2021). Etkili bir şekilde sunulan geribildirim yalnızca performansa dair bilgi vermekle kalmaz aynı zamanda öğrencilerin öz değerlendirme yapmalarına olanak tanıyarak öz düzenleme becerilerini geliştirir (DePasque & Tricomi, 2015; Stančić, 2024). Bu çerçevede öğrencilerin öz yeterlilik düzeyleri ve motivasyonları geribildirim etkinliğini

belirleyen kritik deęişkenler arasında yer almaktadır (Keller vd., 2024). Ayrıca geribildirim bireyin performansına ilişkin belirsizlikleri azaltarak test ve öğrenme sürecine dair kaygıyı hafifletmekte böylece test sürecinde bireylerin kendilerini daha rahat hissetmelerine olanak tanımaktadır (Turan vd., 2011).

Geribildirim öğrenme ortamlarındaki belirsizlikleri azaltarak sınav ve performans kaygısını düşürmekte böylece öğrencilerin kendilerini daha güvende hissetmelerine yardımcı olmaktadır. Öğretmenler açısından ise geribildirimler öğrencilerin hangi alanlarda desteęe ihtiyaç duyduğunu belirlemelerine katkı sağlamakta bu durum ise öğretim sürecinin bireyselleştirilmesini ve öğrencilerin öğrenme hedeflerine daha etkili bir biçimde yönlendirilmesini mümkün kılmaktadır (Ajjawi vd., 2023; Guo & Wei, 2019). Sonuç olarak güncel araştırmalar doğrultusunda geribildirim yapılandırılmış, öğrenciye özgü, duygusal olarak destekleyici ve motivasyonel açıdan besleyici nitelikte olması eğitim sürecindeki başarının temel belirleyicilerinden biri olarak öne çıkmaktadır.

2.5.1. Geribildirim Türleri

Biçimlendirici Geribildirim: Öğrenme süreci devam ederken öğrencilere gelişimlerini desteklemek amacıyla sunulan yönlendirici bilgilerdir. Bu tür geribildirim öğrencinin güçlü ve zayıf yönlerini fark etmesine, hatalarını anlamasına ve daha etkili öğrenme stratejileri geliştirmesine yardımcı olur (Mackintosh-Franklin, 2021). Araştırmalar biçimlendirici geribildirim motivasyonu artırdığını, öğrenme sürecine sürekli katkı sağladığını ve akademik başarı üzerinde olumlu etkiler yarattığını ortaya koymaktadır (Wang vd., 2019; Weidlich vd., 2024).

Sonuç Odaklı Geribildirim: Öğrenme sürecinin sonunda verilen ve öğrencinin genel başarı düzeyini yansıtan deęerlendirmelerdir. Genellikle sınavlar, proje teslimleri veya dönem sonu performans görevleriyle sunulur. Bu tür geribildirimler, öğrencilerin gelişim düzeylerini görmelerine yardımcı olur ancak öğrenme üzerindeki doğrudan etkisi sınırlı olabilir. Güncel uygulamalarda sonuç odaklı geribildirimlerin biçimlendirici unsurlarla harmanlanması yoluyla daha bütüncül ve etkili sonuçlar elde edilmektedir (Mohan, 2024; Svensäter & Rohlin, 2023).

Anlık Geribildirim: Öğrencinin gerçekleştirdiğı bir eylemin hemen ardından sunulan geribildirim türüdür. Özellikle bilgisayar destekli öğrenme ortamlarının yaygınlaşmasıyla birlikte sıklıkla kullanılmaktadır. Programlama ve matematik gibi uygulama temelli derslerde

yüksek etkililiğe sahiptir (Marwan vd., 2022). Bu geribildirimler, hataların anında fark edilmesini sağlayarak hızlı müdahaleye ve kalıcı öğrenmeye katkı sunar (DePasque & Tricomi, 2015; Sanchez-Santillan vd., 2023).

Pozitif & Negatif Geribildirim: Pozitif geribildirim, öğrencinin sergilediği doğru davranışların pekiştirilmesini negatif geribildirim ise hatalı öğrenmelerin düzeltilmesini amaçlar. Pozitif geribildirim motivasyon ve öz güveni artırırken negatif geribildirim dikkatli bir şekilde verilmelidir. Zira norm referanslı negatif geribildirimler öğrencilerde olumsuz duygular yaratabilirken kriter referanslı geribildirimler daha olumlu tepkiler oluşturabilmektedir (Baadte & Schnotz, 2014; Weidlich vd., 2024).

Özelleştirilmiş Geribildirim: Öğrencinin bireysel ihtiyaçları ve performans düzeyine göre uyarlanmış geribildirimlerdir. Araştırmalar kalabalık sınıflarda dijital araçlarla desteklenen özelleştirilmiş geribildirimlerin öğrenmeye anlamlı katkılar sunduğunu göstermektedir (Beneroso & Robinson, 2021). Bu tür geribildirimler, öğrencilerin kendi hızlarında öğrenmelerine olanak tanıyarak öz güvenlerini artırmaktadır (Guo & Wei, 2019).

Sözel & Yazılı Geribildirim: Sözel geribildirim birebir ya da grup ortamında sunularak öğrencilerle daha kişisel ilişkiler kurulmasını sağlar ve motivasyonu destekler. Yazılı geribildirim ise daha ayrıntılı açıklamalar içermekte olup öğrencinin kendi hızında değerlendirme yapmasına imkân verir (Atkinson vd., 2022; Molloy & Boud, 2014; Zhu vd., 2017).

Çeşitli geribildirim türleri bireylerin eğitim sürecinin farklı evrelerinde önemli katkılar sunmaktadır. Biçimlendirici ve anlık geribildirimler öğrenme sürecini desteklerken özelleştirilmiş ve pozitif geribildirimler motivasyonu artırıcı etkiler göstermektedir (Marwan vd., 2022; Weidlich vd., 2024). Buna karşın negatif geribildirimler dikkatle uygulanmalı aksi takdirde motivasyon üzerinde olumsuz etkiler doğurabilir. Geribildirim türlerinin etkili bir şekilde kullanılması hem bireysel hem de grup düzeyinde öğrenme başarılarını artırmak açısından kritik öneme sahiptir (Morris vd., 2021).

Geribildirim, yalnızca akademik başarıyı değil aynı zamanda bilişsel becerilerin gelişimini ve karar alma süreçlerini de destekleyen önemli bir öğrenme aracıdır. Özellikle çok aşamalı ve karmaşık görevlerde sunulan geribildirimler, öğrencilerin problem çözme becerilerini hızla geliştirmelerine ve stratejik kararlar almalarına katkı sunmaktadır (Svensäter & Rohlin, 2023; Wisniewski vd., 2020). Ayrıca geribildirimler öğrencilerin öğrenme

abalarını artırmakta, z gvenlerini ve ğrenmeye ynelik olumlu inanlarını da glendirmektedir (Atkinson vd., 2022; Turda vd., 2021).

Arařtırmalar, geribildirim ğrenme sırasında ortaya ıkan belirsizlikleri ortadan kaldırdığını ve ğrenme kaygısını azalttığını ortaya koymaktadır. Yapılandırılmış ve anlık geribildirimler bireyin rahatlamasına ve ğrenmeye odaklanmasına yardımcı olmaktadır (Jones vd., 2021; Tomasine, 2024). Yazılı geribildirimler, zellikle yabancı dil ğreniminde dilbilgisi hatalarının dzeltilmesi ve yazma becerilerinin geliřtirilmesinde etkili bulunmuřtur. Bu da yazılı performansın ve yazma motivasyonunun artmasını saėlamaktadır (Elola & Oskoz, 2022; Siekmann vd., 2023). Ayrıca ğrenme hedeflerine ynelik aık ve sistematik řekilde sunulan geribildirimler, bilginin daha derinlemesine kavranmasını ve kalıcı ğrenmeyi desteklemektedir (Morris vd., 2021; Zhuang vd., 2017).

2.6. İlgili Arařtırmalar

BOBUT kapsamında geliřtirilen uygulamalarda kullanıcı deneyimi, lme gvenirliėi ve teknolojik kabul edilebilirlik kritik neme sahiptir. Literatrdeki alıřmalar, BOBUT uygulamalarına iliřkin kullanıcı deneyimiyle ilgili eřitli bulgular sunmaktadır. Akhtar ve Kovács (2024) BOBUT sistemlerinin lme doėruluėu aısından sabit testlere gre stn olduėunu ortaya koymuřtur. Ancak kullanıcı deneyimi aısından sabit testlerin daha olumlu algılanabildiėi, BOBUT sistemlerinde ise test zme sresi farkındalıėı ve biliřsel ykn daha yoğun hissedildiėi belirtilmiřtir. Bu sonular Babcock ve Weiss'in (2012) alıřmasıyla rtřmekte, alıřmada durdurma kurallarının yanlış yapılandırılmasının kullanıcı gvenini olumsuz etkileyebileceėi vurgulanmaktadır.

Soe ve arkadaşları (2024) BOBUT sistemlerinin kullanıcı dostu olduėunu ve bireysel ğrenme stillerini belirlemede yksek bařarı saėladığını gstermiřtir. Ara yz tasarımı ve etkileřim kalitesi, test deneyiminin belirleyici unsurları arasında yer almaktadır. Cheng (2023) Concerto yazılımının kullanıcı dostu ara yz ve gl lme altyapısıyla hem arařtırmacılar hem de uygulayıcılar tarafından tercih edildiėini belirtmektedir. Benzer řekilde Tu ve arkadaşları (2023) flexCAT platformunun esnek ve eriřilebilir yapısıyla kullanıcı deneyimini desteklediėini vurgulamıřtır.

Van Wijk ve arkadaşları (2024) tıp ėrencileriyle yrttkleri karřılařtırmalı alıřmada BOBUT'un geleneksel testlere gre daha kısa srede tamamlandıėını ve gvenilir sonular sunduėunu ancak kullanıcılar tarafından daha zorlayıcı bulunduėunu bildirmiřtir.

Beiser ve arkadaşları (2016) depresyon taramalarında kullanılan BOBUT uygulamalarının yüksek tekrar test güvenilirliğine sahip olduğunu ve olumlu kullanıcı geri bildirimi aldığını belirtmiştir.

Uyarlanabilir sistemlerin kabul düzeyine ilişkin araştırmalar, teknolojik kabule dair önemli ipuçları sunmaktadır. Rost ve arkadaşları (2017) depresyon için geliştirilen CBT sistemlerinin genel olarak olumlu karşılandığını raporlamıştır. Szajna (1996) tarafından geliştirilen teknoloji kabul modeli, kullanıcı deneyimi ile sistem kullanım niyeti arasında anlamlı bir ilişki olduğunu göstermiştir.

Geerards ve arkadaşları (2019) bilişsel yük açısından grafiksel ve metin tabanlı geribildirimlerin test deneyimini iyileştirdiğini ancak test uzunluğunun azaltılmasının deneyime doğrudan katkı sunmadığını belirtmiştir. Baysari ve arkadaşları (2018) kullanıcı deneyimlerinin sistem kullanımının ilk altı ayında değiştiğini ve bu süreçte yapılan geliştirmelerin benimsemeyi kolaylaştırdığını göstermiştir. Kasturi ve arkadaşları (2017) ise hasta bildirimli uyarlanabilir testlerin güvenilir ve geçerli sonuçlar sunduğunu kullanıcılar tarafından kolayca kabul edildiğini bulgulamıştır. Legg ve Buhr (1992) BOBUT sistemlerinin genel olarak öğrenciler tarafından sabit testlere benzer şekilde algılandığını ancak bireyselleştirme düzeyinin deneyimi farklılaştırdığını ifade etmiştir.

Ramgovind ve Pramjeeth (2023) üniversite öğrencilerinin BOBUT uygulamalarına olumlu yaklaştığını ve sistemin bireysel öğrenme ihtiyaçlarını karşılama potansiyeline sahip olduğunu belirtmiştir. Bu bulgular, Thompson ve Weiss'in (2011) çalışmasında yer alan olumlu kullanıcı tutumlarıyla da örtüşmektedir. Awad ve arkadaşları (2022) BOBUT'un bireyselleştirilmiş doğasını temel alarak geliştirilen öğrenme ortamlarının kullanıcı etkileşimini ve kullanıcı deneyimini artırdığını vurgulamaktadır.

Kisieleska ve arkadaşları (2023) Avrupa genelinde tıp öğrencileriyle yürüttükleri araştırmada uyarlanabilir algoritmaların öğrencilerin bireysel bilgi düzeyine göre ayarlanmasının motivasyonu artırdığını ve test uzunluğuna ilişkin algıyı olumlu etkilediğini belirtmiştir. Latu ve Chapman (2002) BOBUT'un bireyin yetenek düzeyine göre madde seçimi sunarak daha bireyselleştirilmiş ve kabul edilebilir bir test deneyimi sağladığını vurgulamıştır. Oppl ve arkadaşları (2017) esnek çevrim içi CAT platformları sayesinde bireysel öğrenme profillerine uygun test düzenleme olanaklarının arttığını ifade etmiştir. Markovic ve arkadaşları (2013) bireysel öğrenme stillerine göre uyarlanan test sistemlerinin öğrenme süreçlerini kolaylaştırdığını ve kullanıcı kabulünü artırdığını göstermiştir.

BOBUT sistemlerinin kullanıcı deneyimini artırmaya yönelik diğer bir odak noktası bireylere uygun madde seçimi ile sürecin bireyselleştirilmesidir. Bu yöntem, bireylerin yetenek düzeylerine daha uygun ve güvenilir kestirimler yapılmasını sağlamaktadır. Ayrıca test süresinin kısalması da kullanıcı memnuniyetine katkıda bulunmaktadır (Zhuang vd., 2022). Ancak madde seçme algoritmalarının adil ve tutarlı olabilmesi için kullanıcı verisine dayalı model iyileştirmeleri önerilmektedir (Kwon vd., 2023).

Chang (2015) uyarlanabilir testlerin bireysel öğretim tasarımlarına katkısını vurgulamakta ve yüksek ölçme doğruluğu sunan sistemlerin kullanıcı deneyimi açısından da avantaj sağladığını belirtmektedir. Kimura (2017) CAT sistemlerinde psikometrik etkililik ile kullanıcı deneyimi arasında denge kurulması gerektiğini vurgulamakta ve yüksek ölçme hassasiyetine rağmen bazı kullanıcıların motivasyon kaybı yaşayabildiğini ifade etmektedir. Bu bulgular test tasarımında kullanıcı odaklılık ile ölçme doğruluğu arasında hassas bir denge kurulması gerektiğine işaret etmektedir.

BÖLÜM III

YÖNTEM

Bu bölümde araştırma sorularını yanıtlamak için kullanılan araştırma modeli ve varsayımları, çalışma grubu, veri toplama araçları, Concerto ve BOBUT formlarının özellikleri ve post-hoc simülasyon koşulları açıklanmış ve elde edilen verilere dayalı olarak kullanılan istatistiksel yöntemler hakkında bilgi verilmiştir.

3.1. Araştırmanın Modeli

Bu araştırma öğretmen adaylarının kullanıcı deneyimini etkileyen değişkenlerin belirlenmesi amacıyla hem nitel hem de nicel verilerin eş zamanlı olarak toplandığı ve analiz edildiği yakınsayan paralel karma yöntem deseni (convergent parallel mixed methods design) çerçevesinde gerçekleştirilmiştir. Bu model, farklı türdeki verilerden elde edilen sonuçların birleştirilmesiyle çok boyutlu olgulara ilişkin daha bütüncül bir anlayış geliştirmeyi amaçlamaktadır (Creswell & Plano Clark, 2017). Yakınsayan paralel desenin eğitim teknolojileri araştırmalarında yaygın kullanımı nitel ve nicel verilerin entegrasyonu ile kullanıcı deneyimi gibi çok boyutlu olguların derinlemesine incelenmesini kolaylaştırmaktadır (Schoonenboom & Johnson, 2017).

Bu bağlamda öğrencilerin görüşleri içerik analizi tekniğiyle nitel olarak, Likert tipi ölçekler aracılığıyla ise nicel olarak değerlendirilmiştir. İçerik analizi; verilerin kodlama yoluyla kategorilere ayrıldığı, sistematik ve yinelenebilir bir yöntemdir (Büyüköztürk vd., 2015). Nicel veriler ise tarama araştırması yöntemiyle elde edilmiş ve bu yöntemle bireylerin görüşleri belirli bir zaman dilimi içinde betimlenmiştir (Tekindal, 2021).

Araştırma süreci BOBUT kullanıcı deneyimine dair sistematik test uygulamalarının içeriğinin yapılandırılmasına ilişkin katkısıyla aynı zamanda bir AR-GE çalışması niteliği taşımaktadır. Eğitim teknolojisi alanındaki AR-GE projeleri, tasarım odaklı süreçlerle teknolojik uygulamaların geliştirilmesini, test edilmesini ve kullanıcı geri bildirimleriyle yeniden yapılandırılmasını içerir (Kozma, 2000). Bu yönüyle araştırma, yenilikçi bir eğitim teknolojisi ürünü olan BOBUT'un tasarım, simülasyon ve sahada uygulanmasına dair döngüsel geliştirme adımlarını da kapsamaktadır.

3.2. Çalışma Grubu

Araştırma iki farklı araştırma grubu ile yürütülmüştür. Madde havuzunun oluşturulması sürecinde çalışmaya katılan bireylere ilişkin demografik bilgiler Tablo 3.1’de sunulmuştur. Bu çalışmada öncelikle madde havuzunu oluşturan maddelere ait parametrelerin kestirilmesi amacıyla veri toplanmıştır. Madde parametreleri kestirildikten ve madde havuzu oluşturulduktan sonra BOBUT kullanıcı deneyimine etki eden değişkenlerin belirlenmesine yönelik testler hazırlanmış ve tekrar veri toplanmıştır.

Tablo 3.1. Madde Havuzunun Oluşturulmasındaki Çalışma Grubuna İlişkin Demografik Özellikler

Demografik Özellikler	Kategoriler	n	Yüzde (%)
Cinsiyet	Kadın	179	61.1
	Erkek	114	38.9
Üniversite	Akdeniz Üniversitesi	38	13.0
	Pamukkale Üniversitesi	255	87.0
Bölüm	Matematik	23	7.8
	Sosyal bilgiler	47	16.0
	Türkçe	37	12.6
	Sınıf	23	7.8
	PDR	45	15.4
	İngilizce	5	1.7
	Fen bilgisi	20	6.8
	Müzik	20	6.8
	Özel eğitim	26	8.9
	Okul öncesi	13	4.4
	BESYO	34	11.6

Canlı BOBUT uygulamasına katılan bireylere ilişkin demografik özelliklere Tablo 3.2’de yer verilmiştir.

Tablo 3.2. BOBUT Uygulamasında Kullanıcı Deneyiminin Araştırılmasına Yönelik Çalışma Grubuna İlişkin Demografik Özellikler

Demografik Özellikler	Kategoriler	n	Yüzde (%)
Cinsiyet	Kadın	76	58.0
	Erkek	55	42.0
Üniversite	Akdeniz Üniversitesi	105	80.2
	Pamukkale Üniversitesi	26	19.8
Bölüm	Matematik	27	20.6
	Sosyal bilgiler	27	20.6
	Türkçe	21	16.0
	Sınıf	12	9.2
	PDR	11	8.4
	İngilizce	11	8.4
	Fen bilgisi	22	16.8

3.3. Veri Toplama Araçları

Çalışmada Eğitimde Ölçme ve Değerlendirme dersine yönelik çoktan seçmeli test maddeleri hazırlanmış ve bu maddelerin kalibrasyonu gerçekleştirilmiştir. Ayrıca BOBUT uygulamaları sonrasında kullanıcı deneyiminin incelenmesine yönelik olarak katılımcı anketleri uygulanmıştır. Katılımcılardan her iki uygulamada da bölüm ve cinsiyet değişkenlerine ilişkin veriler toplanmıştır.

3.3.1. Madde Havuzuna İlişkin Uzman Görüşü

Bu çalışmada Eğitimde Ölçme ve Değerlendirme dersine ilişkin konulardan 130 çoktan seçmeli test maddesi yazılmıştır. Maddeler öncelikle iki dil uzmanına gönderilmiş ve gerekli düzeltmeler yapılmıştır. Daha sonra maddeler, dört ölçme ve değerlendirme uzmanına gönderilmiştir. Uzman görüşleri doğrultusunda gerekli düzenlemeler gerçekleştirilmiş ve 40 madde testten çıkartılarak 90 maddelik nihai madde havuzu oluşturulmuştur.

Oluşturulan madde havuzunda 21 madde temel kavramlara, 28 madde ölçme aracının niteliklerine, 21 madde ölçme araçlarına ve 20 madde test ve madde istatistiklerine ilişkin konuları kapsamaktadır.

3.3.2. Eğitimde Ölçme ve Değerlendirme Madde Havuzunun Oluşturulması ve Veri Toplama Aracına İlişkin Eksik Test Deseni

BOBUT uygulamasında kullanıcı deneyimine etki eden değişkenlerin ve olası etkilerinin belirlenmesi amacıyla Eğitimde Ölçme ve Değerlendirme alanına ilişkin çoktan seçmeli test maddeleri yazılmıştır. Yazılmış olan test maddeleri katılımcılara iki farklı yöntemde uygulanmıştır. Hazırlanan madde havuzu içerisinde test formlarının karşılaştırılabilirliğini sağlamak amacıyla 10 adet ortak madde (anchor item) belirlenmiştir. Bu maddeler, Eğitimde Ölçme ve Değerlendirme dersi kapsamında yer alan dört konunun tamamını kapsayacak şekilde seçilmiştir. Maddeler tüm test formlarında yer alarak madde parametrelerinin ortak bir test üzerinde kalibre edilmesini sağlamak amacıyla hazırlanmıştır.

Toplam 30'ar maddeden oluşan test formları her birinde 10 ortak madde ve 20 farklı test maddesi olacak şekilde yapılandırılmıştır. Bu yapı, çoklu matris örnekleme (MMS) yaklaşımına uygun olarak tasarlanmıştır. MMS tasarımı, farklı bireylerin farklı test formlarına yanıt vermesini sağlarken içerik kapsamını korur ve ölçme yükünü azaltır (Aybek vd., 2024). Bu çalışmada kullanılan test formu yapısı literatürde “unbalanced incomplete block design with a common block” olarak tanımlanan bir MMS türüne karşılık gelmektedir (Gonzalez & Rutkowski, 2010).

Bu bağlamda test uygulamasında her birey yalnızca 30 maddeye yanıt vermektedir; böylece hem uygulama süresi kısaltılmakta hem de veri toplama süreci daha verimli hâle getirilmektedir. Oluşturulan testlere ilişkin kapsam Tablo 3.3'te sunulmuştur.

Tablo 3.3. Test Kitapçıklarına İlişkin Kapsam Geçerliliği

Testler		Test Yapısı	
Test I	Ortak Maddeler(10)	Temel Kavramlar I(10)	Ölçme Aracının Nitelikleri I(10)
Test II	Ortak Maddeler(10)	Ölçme Araçları I(10)	Madde ve Test İstatistiği I(10)
Test III	Ortak Maddeler(10)	Ölçme Aracının Nitelikleri II(10)	Ölçme Araçları II(10)
Test IV	Ortak Maddeler(10)	Madde ve Test İstatistiği II(10)	Temel Kavramlar II(10)

Araştırma kapsamında oluşturulan diğer test deseni 90 maddelik havuzun iki kitapçık hâlinde uygulanmasıdır. Testin uzunluğu göz önünde bulundurularak öğrencilerin dikkat düzeylerini korumak ve bilişsel yükü azaltmak amacıyla uygulama iki oturuma ayrılmıştır. Bu

doğrultuda katılımcılar testin ilk 45 maddesini birinci oturumda, kalan 45 maddesini ise ikinci oturumda yanıtlamıştır.

Oturumlara bölünmüş bu tür test uygulamaları özellikle düşük motivasyon, dikkat dağınıklığı ve test yorgunluğu gibi olumsuzluklara karşı etkili bir önlem olarak değerlendirilmektedir. Literatürde uzun testlerde bilişsel tükenme ve performans düşüşünün özellikle ikinci yarıda madde yanıt doğruluğunu olumsuz etkileyebileceği belirtilmektedir (Paek & Young, 2005).

Her iki uygulamada da ortak maddelerin yer alması madde parametrelerinin ortak bir test temelinde kalibre edilmesine olanak tanımıştır.

3.3.3. Concerto ve BOBUT Uygulama Formlarının Özellikleri

Araştırmanın amacına uygun olarak belirlenen geribildirim ve süre değişkenleri dikkate alınarak dört farklı BOBUT tasarımı Concerto platformu kullanılarak geliştirilmiştir. Uygulama süreci öncesinde yapılan post-hoc simülasyon çalışmaları sonucunda madde seçimi için MFI (Maximum Fisher Information), yetenek kestirimi için Bayesci MAP (Maximum A Posteriori) yöntemi ve sabit 10 madde ile sınırlandırılmış test uzunluğu kullanılarak durdurma kuralı uygulanmasının geçerli ve güvenilir sonuçlar üreteceği belirlenmiştir.

Concerto uygulamaları, R yazılımını temel alarak çalışmaktadır. BOBUT sisteminin kullanılabilmesi için Apache, MySQL ve R yazılımlarının birlikte çalıştırılması gerekmektedir. Bu çalışma kapsamında uygulamaların yürütülebilmesi amacıyla Amazon'dan bir sunucu kiralanmıştır. Amazon T2-Micro tipi sunucuda Apache ve MySQL yazılımları önceden yüklü olarak gelmekte olup yalnızca R paketleri ve Concerto kurulumu ayrıca yapılmıştır.

Testlerin R kodlarıyla çalıştırılması nedeniyle sunucunun en az 1 GB RAM'e sahip olması gerekmektedir. Ancak uygulamaya aynı anda çok sayıda öğrencinin katılımı planlanıyorsa RAM kapasitesinin artırılması önerilmektedir. Ayrıca canlı BOBUT uygulamasında bireylerin teste rastgele atanmasını sağlamak amacıyla yönlendirme sayfası kullanılmış ve bu sayede gruplar arası yanlılık olasılığı minimize edilmiştir. BOBUT uygulamalarına ilişkin ekran görüntüleri Şekil 3.1–3.5 arasında sunulmuştur.

HOŞ GELDİNİZ!!!

Bu test, "BOBUT Uygulamasında Kullanıcı Deneyimine Etki Eden Değişkenlerin ve Olası Etkilerinin Belirlenmesi" isimli doktora tezi için veri toplanması amacıyla uygulanmaktadır.

Alper TOSUN - Akdeniz Üniversitesi

Doç. Dr. Alper SİNAN - Akdeniz Üniversitesi

Uygulamaya telefondan katılım gösteriyorsanız lütfen telefonunuzu yan çevirip öyle devam ediniz.

Teste başla

Şekil 2.1. BOBUT Karşılama Ekranı

Ölçme ve Değerlendirme Bilgisayar Ortamında Bireye Uyarlanmış Test

00:01:55

Soru	Güçlük(p)	Ayırt Edicilik(r)
I	0,10	0,05
II	0,20	0,45
III	0,30	0,40
IV	0,20	-0,10
V	0,50	0,50

Bir testteki maddelere ilişkin madde güçlükleri ve ayırt edicilikleri yukarıdaki tabloda verilmiştir. Tabloda verilen indekslere göre hangi maddenin güvenilirliğe katkısı en yüksektir?

I

II

III

IV

V

Page 4 / 10

İleri

Şekil 3.2. Sürenin Gösterildiği Yanıt Verilmiş Madde

Bir matematik öğretmeni 500 kişilik 7.sınıf öğrencileri arasından matematik bilgi yarışması için 5 kişi seçecektir. Öğrencilere yapılan sınav sonucunda 95 puan üzeri alan öğrenciler takıma seçilecektir. Ancak 95 puan üzerinde puan alan yalnızca 3 öğrencidir. Buna göre öğretmenin yapmış olduğu hatanın sebebi nedir?

Testin objektifliğinin düşük olması
Öğrenci başarı düzeyinin düşük olması
Standart hatanın çok yüksek çıkması
Testin kullanılabilirliğinin düşük olması
Bağıl ölçüt yerine mutlak ölçüt kullanması

Page 1 / 10

İleri

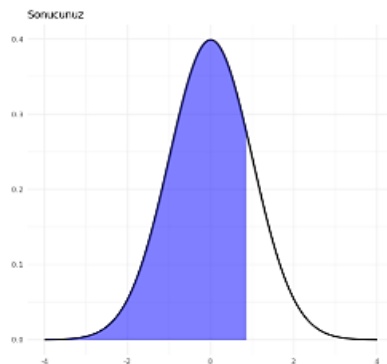
Şekil 4.3. Sürenin Gösterilmediği Yanıt Verilmiş Madde

Uygulamaya katıldığınız için teşekkürler. Almış olduğunuz testin orijinal hali 75 maddeden oluşmaktadır. Ancak bilgisayar ortamında bireyselleştirilmiş test uygulamaları sayesinde bu 75 maddelik test içerisinde sizin başarı düzeyinize en uygun olan 10 madde uygulandı. Yani bu yöntemle **%86,67** daha az madde cevaplandırmış oldunuz. Performansınıza ilişkin sonuçlar aşağıda verilmiştir.

Teste ait güvenirlik düzeyi: **0.83** olarak hesaplanmıştır. Bu değer, testin tutarlı ve güvenilir ölçümler sunduğunu göstermektedir. Ancak, test sonuçlarının en doğru şekilde değerlendirilmesi için bireysel farklılıklar, sınav anındaki koşullar ve dış faktörlerinde göz önünde bulundurulması gerektiğini unutmayın.

Başarı düzeyiniz grubun **%80.57** 'den daha yüksektir. Grafikteki mavi taralı alan daha başarılı olduğunuz kitleyi temsil etmektedir.

Yorum: Tebrikler! Testteki performansınız, konuya büyük ölçüde hakim olduğunuzu ve başarılı bir seviyede olduğunuzu gösteriyor. Ancak, daha da iyi olmak için çalışmaya devam etmelisiniz. Yeni soru tiplerini denemek, derinlemesine analiz yapmak ve bilgiyi farklı senaryolarda uygulamak öğrenmenizi daha da güçlendirecektir. Öğrenmeye açık olmaya devam ederek eksiklerinizi kapatabilir ve kendinizi daha üst seviyeye taşıyabilirsiniz!



Testiniz sonlanmıştır. Aşağıdaki bağlantıdan katılmış olduğunuz uygulamaya yönelik olarak hazırlanmış soruları cevaplandırmanız gerekmektedir. Katılımınız için teşekkürler.

Görüş Formu

Şekil 5.4. Geniş Geribildirim Sayfası

Başarı düzeyiniz size benzer bir grubun %48.93'den daha yüksektir.

Testiniz sonlamıştır. Aşağıdaki bağlantıdan katılmış olduğunuz uygulamaya yönelik olarak hazırlanmış soruları cevaplandırmanız gerekmektedir. Katılımınız için teşekkürler.

Görüş Formu

Şekil 6.5. Dar Geribildirim Sayfası

Canlı BOBUT uygulamasına katılan bireylerin karşılandığı giriş ekranına Şekil 3.1’de yer verilmiştir. Uygulama tasarımlarında kullanılan süre değişkeni, bireylerin test sırasında süreyi görüp görmemeleri durumunu ifade etmektedir (Şekil 3.2 ve Şekil 3.3).

Geribildirim değişkeni için belirlenen geniş kapsamlı geribildirimde madde havuzunun özellikleri, BOBUT sisteminin bu alandaki avantajları, testin güvenilirlik düzeyi, bireyin benzer gruplara kıyasla başarı durumu, başarıya yönelik açıklayıcı yorum ve başarı grafiği sunulmuştur (Şekil 3.4). Ayrıca testin tamamlanmasının ardından bireylerin kestirilen yetenek düzeylerine göre motive edici ve bireyselleştirilmiş geribildirimler de verilmiştir. Bu geribildirimler şunlardır:

- $\theta \leq -1.0$: “Bu testte elde ettiğiniz sonuç, değerlendirdiğimiz konu alanında şu an için gelişime açık bir noktada olduğunuzu göstermektedir. Ancak bu, ilerleyemeyeceğiniz anlamına gelmez! Öğrenme süreci zaman ve çaba gerektirir. Eksik olduğunuz konuları belirleyerek düzenli tekrar yapmanız ve ek kaynaklardan destek almanız önerilir. Öğrenme stratejileri geliştirmek, küçük adımlarla ilerlemek ve düzenli pratik yapmak başarıyı artıracaktır. Unutmayın, her bireyin öğrenme süreci farklıdır ve sürekli gelişime açıktır!”
- $-1.0 < \theta < 0.0$: “Sonuçlarınız, konuyla ilgili temel kavramlara hakim olduğunuzu ancak bazı alanlarda daha fazla çalışmanız gerektiğini göstermektedir. Eksik olduğunuz noktaları belirleyerek konu tekrarları yapmanız, yeni soru türleriyle pratik yapmanız faydalı olacaktır. Özellikle zorlandığınız konulara odaklanarak eksikliklerinizi giderdiğinizde önemli bir gelişim gösterebilirsiniz. Motivasyonunuzu kaybetmeden, planlı ve düzenli çalışmaya devam edin!”
- $0.0 < \theta < 1.0$: “Tebrikler! Testteki performansınız, konuya büyük ölçüde hakim olduğunuzu ve başarılı bir seviyede olduğunuzu gösteriyor. Ancak, daha da iyi olmak için çalışmaya devam etmelisiniz. Yeni soru tiplerini denemek, derinlemesine analiz

yapmak ve bilgiyi farklı senaryolarda uygulamak öğrenmenizi daha da güçlendirecektir. Öğrenmeye açık olmaya devam ederek eksiklerinizi kapatabilir ve kendinizi daha üst seviyeye taşıyabilirsiniz!”

- $1.0 \leq \theta$: “Harika bir iş çıkardınız! Sonuçlarınız, bu konudaki yetkinliğinizin oldukça yüksek olduğunu göstermektedir. Güçlü yönlerinizi koruyarak bilgilerinizi daha derinlemesine pekiştirebilir ve ileri seviyedeki konulara yönelebilirsiniz. Bilginizi daha etkili kullanmak için karmaşık problem çözümleriyle kendinizi zorlamaya devam edebilirsiniz. Aynı zamanda başkalarına rehberlik ederek konuları daha iyi pekiştirebilirsiniz. Başarılarınızın devamını dileriz!”

Şeklinde. Bu bağlamda geniş geribildirim sonuç odaklı, pozitif ve özelleştirilmiş bir yapıda oluşturulmuştur. Dar geribildirim için ise bireylere yalnızca benzer bir gruba göre başarı düzeyleri sunulmuştur (Şekil 3.5). Diğer bir deyişle dar geribildirim sadece sonuç odaklı bir yapıda yapılandırılmıştır. BOBUT tasarımlarına ilişkin bilgilere Tablo 3.4’te yer verilmiştir.

Tablo 3.4. BOBUT Uygulamalarına İlişkin Değişkenler

Test	Süre	Geribildirim
Test-1	Gösteriliyor	Detaylı Geribildirim
Test-2	Gösteriliyor	Dar Geribildirim
Test-3	Gösterilmiyor	Detaylı Geribildirim
Test-4	Gösterilmiyor	Dar Geribildirim

Katılımcılar BOBUT uygulaması sonucunda geribildirim sayfasında görüş formuna yönlendirilmiştir.

3.3.4. Kullanıcı Deneyimine İlişkin Görüş Formu

Katılımcıların BOBUT uygulamasındaki deneyimlerine ilişkin görüşlerin toplanması adına her BOBUT uygulaması için Google Form üzerinden kullanıcı deneyimi görüş formu oluşturulmuştur. Oluşturulan görüş formları kapsamında katılımcılara uygulama deneyimlerine yönelik altı açık uçlu soru ve beş Likert tipi soru yöneltilmiştir. Bu bağlamda kullanıcılara yöneltilen açık uçlu sorular şunlardır:

1. Test sırasında aldığınız geribildirim uzunluğu hakkında düşüncelerinizi detaylandırabilir misiniz? Bu geribildirim motivasyonunuz ve genel deneyiminiz üzerindeki etkilerini örneklerle nasıl açıklarsınız?
2. Test sırasında sürenin gösterilmesi/gösterilmemesi, zaman yönetiminizi ve stres düzeyinizi nasıl Şekillendirdi?
3. Soruları çözdüğünüz ekranda sana kalan süren gösterildi/gösterilmedi. Soruları çözerken bunun dışında başka bir bilgi görmek ister miydin?
4. Test aldığınız sistemle ilgili önerileriniz var mı? Varsa nedir?
5. Aldığınız geribildirim anlamlandırmak için ders hocasına soru sormaya gerek duyduunuz mu? Duyduysanız bu soru neydi?
6. Geribildirim sayfasından başka neler görmek isterdiniz?

Şeklindedir. Katılımcılara yöneltilmiş olan Likert yapıdaki sorular:

1. Aldığınız teste yönelik genel memnuniyet düzeyinizi puanlayınız.
 2. Test sonucunda aldığım geribildirim performansımın doğru bir değerlendirilmesi olduğunu düşünüyorum.
 3. Bu geribildirimle pek uyuşmadığımı düşünüyorum ve içeriğini yeterince objektif ve güvenilir bulmadığım için ciddiye almakta zorlanıyorum.
 4. Aldığım geribildirim, gelecekteki denemelerimdeki motivasyonumu ve gösterdiğim çabayı doğrudan etkileyecek bir faktör olacak.
 5. Gelecekte sınavlarımı bireye uyarlanmış olarak almak isterim.
- Şeklindedir.

3.3.5. Verilerin Toplanması Sürecinde Rastgelelik

Katılımcıların BOBUT uygulaması için homojen katılım gösterebilmesi adına bir web sitesi kullanılmıştır. Bu web sitesi aracılığıyla katılımcılar rastgele testlere yönlendirilmiştir. Bu sayfada katılımcılara uygulamanın amacı, BOBUT'un işleyişi ve testin konusuna ilişkin bilgiler sunulmuştur. Web sitesine ilişkin görsel Şekil 3.6'da verilmiştir.



HOŞ GELDİNİZ

- Bu test, Bilgisayar Ortamında Bireye Uyarlanmış Testlerde Kullanıcı Deneyimine Etki Eden Değişkenlerin belirlenmesi adına uygulanmaktadır.
- Test sürecinde karşılaştığınız her yeni soru, bir önceki soruya verdiğiniz yanıtın doğru veya yanlış olmasına göre otomatik olarak belirlenecektir.
- Test, Eğitimde Ölçme ve Değerlendirme dersi konuları temel alınarak hazırlanmıştır.

Alper TOSUN
Doç. Dr. Alper SİNAN

Şekil 3.6. Rastgeleliğin Sağlanması İçin Oluşturulan Web Sitesi Karşılama Ekranı

Canlı BOBUT uygulamasına toplam 131 kişi katılmıştır. Katılımcıların %20.6'sı Test-1'i (n = 27), %32.0'ı Test-2'yi (n = 42), %25.1'i Test-3'ü (n = 33) ve %22.1'i Test-4'ü (n = 29) almıştır. Testi alan katılımcıların yüzdelerik dağılımı incelendiğinde, grupların homojen dağıldığı gözlemlenmiştir.

3.4. Verilerin Toplanması

Araştırma kapsamında verilerin toplanması süreci iki şekilde yürütülmüştür. İlk olarak BOBUT uygulamasının madde havuzunun oluşturulması amacıyla kâğıt-kalem test uygulaması gerçekleştirilmiştir. Ardından canlı BOBUT uygulaması ve kullanıcı deneyimi anketi uygulanmıştır.

3.4.1. Kâğıt Kalem Uygulaması

Verilerin toplanma süreci BOBUT uygulamaları için madde havuzunun oluşturulmasına yönelik olarak kâğıt-kalem testi şeklinde yürütülmüştür. Uygulamalar 15 Aralık 2024 – 03 Ocak 2025 tarihleri arasında Pamukkale Üniversitesi ve Akdeniz Üniversitesi’nde Eğitimde Ölçme ve Değerlendirme dersini almış öğretmen adaylarına uygulanmıştır. Akdeniz Üniversitesi’ndeki 38 öğrenciye test iki oturumda tüm maddelerin uygulanmasıyla gerçekleştirilmiştir. Spor Bilimleri Fakültesi BESYO bölümünde öğrenim gören 36 öğretmen adayına yapılan uygulama da benzer şekilde iki oturumda yürütülmüştür. Pamukkale Üniversitesi’ndeki kalan 221 öğrenciye ise ortak maddeleri bulunan dört farklı kitapçıktan biri uygulanmıştır. Test kitapçıkları rastgele dağıtılmış ve böylece kopya riskinin önlenmesi amaçlanmıştır.

Çalışmaya katılan bireylere verilerin hangi amaçla toplandığı, uygulamanın ne kadar sürede tamamlanacağı ve sınavla ilgili uyulması gereken kurallar sınav yönergesi aracılığıyla açıklanmıştır. Araştırmacı uygulama süresince yedek test kitapçıkları ve gerekli kırtasiye malzemelerini hazır bulundurmuştur. Katılımcılar cevaplarını test sonunda yer alan cevaplar kısmına işaretleme yaparak vermiştir.

3.4.2. Canlı BOBUT Uygulaması

Canlı BOBUT uygulaması kâğıt-kalem testinden elde edilen sonuçlara dayanarak oluşturulan madde havuzunun farklı test formları aracılığıyla bireylere uygulanması yoluyla gerçekleştirilmiştir. Uygulamaya <https://eptlab.com/olcme> adresinden erişilmiş; test yönergesinin okunmasının ardından bireyler, BOBUT karşılama sayfasına yönlendirilerek sürece başlamışlardır. Katılımcıların maddelere yanıt vermeden geçmelerini önlemek amacıyla her maddenin ardından yalnızca yanıt verildikten sonra aktif hale gelen bir “İleri” butonu eklenmiştir. Katılımcılar uygulamaya kişisel cep telefonları, tabletleri ya da bilgisayarları aracılığıyla katılmışlardır.

Katılımcıların madde yanıtları, hangi konulardan hangi maddelerin uygulandığı, teste başlama ve bitiş süreleri gibi bilgiler MySQL veri tabanına kaydedilmiştir. Bu veriler ilgili veri tabanından araştırma süreci için alınmıştır.

3.5. Verilerin Analizi

3.5.1. Model Varsayımlarının Sınanması

Madde Tepki Kuramı (MTK) analizlerine başlamadan önce modelin temel varsayımları olan yerel bağımsızlık ve madde uyum indeksleri incelenmiştir. Yerel bağımsızlık, aynı yetenek düzeyine sahip bireylerin madde yanıtlarının birbirinden bağımsız olması gerektiği varsayımdır (Hambleton vd., 1991). Bu varsayımın değerlendirilmesinde Yen'in Q3 istatistiği kullanılmıştır. Bu istatistik, iki madde arasındaki artık (residual) korelasyonlara dayanır. Yerel bağımsızlık sağlandığında maddeler arası artık korelasyonlar sıfıra yakın olmalıdır. 0.36 ile 0.40 arasındaki Q3 değerleri, “zayıf yerel bağımlılık” göstergesi olarak kabul edilmektedir (Christensen vd., 2013; Peak & Cole, 2020).

Bu doğrultuda, çalışmada Yen'in Q3 değeri 0.37'nin üzerinde olan maddeler testten çıkartılmıştır. Yerel bağımsızlığa ilişkin testten çıkarılan maddelere ait Q3 istatistiklerine ilişkin bulgular Tablo 3.5'te sunulmuştur.

Tablo 3.5. Yerel Bağımsızlık Varsayımını Karşılamayan Maddelere İlişkin Yen'in Q3'ü

Madde-1	Madde-2	Artık korelasyonu
38.madde	16.madde	0.383
87.madde	16.madde	0.439
44.madde	33.madde	0.405
36.madde	34.madde	0.402
44.madde	36.madde	0.389
48.madde	36.madde	0.445
77.madde	43.madde	0.382
48.madde	46.madde	0.435
50.madde	48.madde	0.376

Tablo 3.4'e göre 16, 33, 34, 36, 38, 43, 44, 46, 48, 50, 77 ve 87 numaralı maddeler, Yen'in Q3 istatistiği için belirlenen 0.37 sınır değerinin üzerinde artık korelasyonlara sahip oldukları için testten çıkarılmıştır. Yerel bağımsızlık varsayımını karşılamayan maddelerin çıkarılmasının ardından kalan 78 madde üzerinden madde düzeyinde model uyumu değerlendirilmiştir.

Madde uyumunun değerlendirilmesinde en yaygın kullanılan istatistiklerden biri olan Ki-kare (χ^2) testi, her bir maddenin önerilen modele ne düzeyde uyum sağladığını belirlemektedir (de Ayala, 2009). Bununla birlikte RMSEA (Root Mean Square Error of Approximation) değerleri ile bu değerlere karşılık gelen anlamlılık düzeyleri (p değerleri) de model uyumunun değerlendirilmesinde dikkate alınmıştır. RMSEA değerinin 0.06'nın

üzerinde olması ve p değerinin 0.05'ten küçük olması durumunda, madde ile model arasında anlamlı bir uyumsuzluk olduğu kabul edilmektedir (Browne & Cudeck, 1993; Hambleton vd., 1991). Bu bağlamda kalan maddelere ilişkin model uyum indeksleri Tablo 3.6'da sunulmuştur.

Tablo 3.6. Maddelere İlişkin Uyum İndeksleri

Madde	χ^2	df	RMSEA	p	Madde	χ^2	df	RMSEA	p
1.madde	9.734	9	0.0190	0.372	49.madde	3.531	9	0.0001	0.940
2.madde	7.826	9	0.0001	0.552	51.madde	9.944	9	0.0220	0.355
3.madde	6.456	9	0.0001	0.694	52.madde	9.138	9	0.0080	0.425
4.madde	10.805	9	0.0300	0.289	53.madde	19.882	9	0.0740	0.019
5.madde	17.427	9	0.0650	0.042	54.madde	10.416	9	0.0270	0.318
6.madde	12.297	9	0.0410	0.197	55.madde	6.041	9	0.0001	0.736
7.madde	14.51	9	0.0530	0.105	56.madde	4.638	9	0.0001	0.865
8.madde	17.772	9	0.0670	0.038	57.madde	11.529	9	0.0360	0.241
9.madde	6.426	9	0.0001	0.697	58.madde	4.567	9	0.0001	0.870
10.madde	10.804	9	0.0300	0.289	59.madde	10.851	9	0.0310	0.286
11.madde	11.613	9	0.0360	0.236	60.madde	21.123	9	0.0780	0.012
12.madde	9.465	9	0.0150	0.395	61.madde	4.004	9	0.0001	0.911
13.madde	5.142	9	0.0001	0.822	62.madde	4.106	9	0.0001	0.904
14.madde	16.51	9	0.0620	0.057	63.madde	6.023	9	0.0001	0.738
15.madde	7.148	9	0.0001	0.622	64.madde	20.495	9	0.0760	0.015
17.madde	14.085	9	0.0510	0.119	65.madde	14.122	9	0.0510	0.118
18.madde	8.488	9	0.0001	0.486	66.madde	12.896	9	0.0440	0.167
19.madde	5.878	9	0.0001	0.752	67.madde	30.711	9	0.1050	0.0001
20.madde	12.455	9	0.0420	0.189	68.madde	12.607	9	0.0430	0.181
21.madde	15.628	9	0.0580	0.075	69.madde	10.139	9	0.0240	0.339
22.madde	9.977	9	0.0220	0.352	70.madde	6.96	9	0.0001	0.641
23.madde	13.342	9	0.0470	0.148	71.madde	9.522	9	0.0160	0.391
24.madde	6.27	9	0.0001	0.713	72.madde	13.078	9	0.0450	0.159
25.madde	4.105	9	0.0001	0.904	73.madde	6.602	9	0.0001	0.678
26.madde	8.932	9	0.0001	0.444	74.madde	5.442	9	0.0001	0.794
27.madde	15.185	9	0.0560	0.086	75.madde	8.039	9	0.0001	0.530
28.madde	10.903	9	0.0310	0.282	76.madde	18.128	9	0.0680	0.034
29.madde	6.248	9	0.0001	0.715	78.madde	12.459	9	0.0420	0.189
30.madde	10.016	9	0.0230	0.349	79.madde	19.751	9	0.0740	0.020
31.madde	6.08	9	0.0001	0.732	80.madde	7.459	9	0.0001	0.589
32.madde	12.481	9	0.0420	0.188	81.madde	7.904	9	0.0001	0.544
35.madde	7.473	9	0.0001	0.588	82.madde	13.96	9	0.0500	0.124
37.madde	13.844	9	0.0490	0.128	83.madde	13.932	9	0.0500	0.125
39.madde	11.38	9	0.0350	0.251	84.madde	5.22	9	0.0001	0.815
40.madde	19.858	9	0.0740	0.019	85.madde	12.762	9	0.0440	0.174
41.madde	18.298	9	0.0690	0.032	86.madde	11.07	9	0.0320	0.271
42.madde	9.036	9	0.0040	0.434	88.madde	4.082	9	0.0001	0.906
45.madde	12.333	9	0.0410	0.195	89.madde	9.156	9	0.0090	0.423
47.madde	16.63	9	0.0620	0.055	90.madde	16.196	9	0.0600	0.063

Maddelere ilişkin uyum indeksleri incelendiğinde 60 ($\chi^2 = 21.123$; $df = 9$; $RMSEA = 0.078$; $p = 0.012$), 67 ($\chi^2 = 30.711$; $df = 9$; $RMSEA = 0.105$; $p = 0.0001$) ve 79

($\chi^2 = 19.751$; $df = 9$; $RMSEA = 0.074$; $p = 0.020$) numaralı maddelerin hem istatistiksel olarak anlamlı bir şekilde uyumsuzluk gösterdiği hem de RMSEA içim belirlenmiş olan eşik değerinin üstünde değerler gösterdiği için testten çıkarılmıştır. Madde parametrelerinin kestirilmesi için madde havuzunun 75 maddeden oluşmasına karar verilmiştir. Kalan maddelere ilişkin kapsam şu Şekildedir;

- Temel kavramlar: 17 madde
- Ölçme aracının nitelikleri: 26 madde
- Ölçme araçları: 16 madde
- Madde ve test istatistiği: 16 madde

Madde parametrelerinin kestirilmesinde Rasch modeli tercih edilmiştir. Rasch model, özellikle tek parametrelili yapısı sayesinde küçük ve orta büyüklükteki örneklemelerde istikrarlı sonuçlar üretmesi nedeniyle tercih edilmiştir (Bond & Fox, 2015). Bu çalışmada yer alan 293 kişilik çalışma grubu literatürde Rasch modeline dayalı güvenilir madde kalibrasyonu için önerilen minimum 200 kişilik örneklem büyüklüğünün üzerindedir (Linacre, 1994).

Araştırmalarda küçük örneklemelerde yapılan kalibrasyonların büyük örneklemelere dayalı yetenek kestirimleriyle oldukça tutarlı sonuçlar verdiği gözlemlenmiştir (Boone, 2016; O'Neill, Gregg & Peabody, 2020).

Rasch modelinin daha küçük örneklem büyüklüğü gereksinimi ve maddeler arası karşılaştırmaya olanak sağlaması eğitimde ölçme ve değerlendirme çalışmalarında yaygın olarak tercih edilmesinin başlıca nedenleri arasında yer almaktadır (Embretson & Reise, 2013). Ayrıca Rasch modelinde elde edilen logit temelli madde güçlük parametreleri ölçme aracının geliştirilmesinde ve madde bankası oluşturulmasında önemli rol oynamaktadır (Wright & Masters, 1982). Model varsayımlarının sınanması için Aybek (2021) tarafından geliştirilen “catIRT” uygulamasında R Core Team (2023) tabanlı mirt v.1.44.0 paketi kullanılmıştır (Chalmers, 2012).

3.5.2. Post-hoc Simülasyon

BOBUT uygulaması için simülasyon koşullarının belirlenmesi adına Aybek (2021) tarafından geliştirilen “catIRT tools” uygulamasında post-hoc simülasyonları için cevap örüntülerinin oluşturulması için ve BOBUT simülasyonu için “catR” paketi kullanılmıştır (Magis & Barrada, 2017; Magis & Raiche, 2012). Simülasyon koşulları 1000 örneklem

büyüklüğü için uygulanmıştır. Simülasyonlar 500 tekrar ile uygulanmıştır. Simülasyon koşulları:

1. Durdurma kuralı: sabit uzunluk=10 madde ve 0.30, 0.40 ve 0.50 standart hata
2. İlk madde seçimi θ : rastgele θ ve $\theta=0.0$
3. MTK modeli: Rasch model
4. Madde Seçimi: MFI (Maximum Fisher Information)
5. Madde kullanım kontrolü(Exposure control): kullanma
6. En düşük θ : -3.0
7. En büyük θ : 3.0
8. θ kestirim modeli: MAP (Maximum a posteriori) ve ML (Maximum Likelihood)
9. Önsel dağılım (prior distribution): $\bar{X}=0.00$ ve $ss=1.00$
10. Ölçekleme (D): 1.7

Simülasyon sonuçlarına göre BOBUT uygulamasının ortalama kaç maddede sonlandığı, ortalama standart hata, RMSEA, ortalama sapma (BIAS) ve gerçek θ ile simülasyon sonucu elde edilen θ değerleri arasındaki korelasyon katsayıları elde edilmiştir. Bu bağlamda hangi durdurma kuralının, hangi ilk madde seçiminin, hangi θ kestirim modelinin kullanılacağı belirlenmeye çalışılmıştır.

3.5.3. BOBUT Kullanıcı Deneyimine İlişkin Görüşlerin İncelenmesi

BOBUT kullanıcı deneyimine ilişkin açık uçlu soruların değerlendirilmesi amacıyla her bir soruya katılımcılar tarafından verilen yanıtlar belirli temalara uygunluk açısından analiz edilmiştir. Bu bağlamda sorular ve cevap temalarına ilişkin gruplar Tablo 3.7’de sunulmuştur.

Tablo 3.7. Açık Uçlu Sorulara Yönelik Cevap Temaları

Soru	Tema
1. Test sırasında aldığınız geribildirim uzunluğu hakkında düşüncelerinizi detaylandırabilir misiniz? Bu geribildirim motivasyonunuz ve genel deneyiminiz üzerindeki etkilerini örneklerle nasıl açıklarsınız?	Olumlu motivasyon Geliştirilmeli /Olumsuz motivasyon
2. Test sırasında sürenin gösterilmesi/gösterilmemesi, zaman yönetiminizi ve stres düzeyinizi nasıl Şekillendirdi?	Pozitif etki Nötr etki Negatif etki
3. Soruları çözdüğünüz ekranda sana kalan süren gösterildi/gösterilmedi. Soruları çözerken bunun dışında başka bir bilgi görmek ister miydin?	Başka bir şey görmek istemezdim. Süre bilgisi Puan bilgisi Diğer
4. Test aldığınız sistemle ilgili önerileriniz var mı? Varsa nedir?	Önerim yok Ara yüz önerisi İçerik önerisi
6. Geribildirim sayfasından başka neler görmek isterdiniz?	Açıklama yeterli Ek açıklama(kaynak, konu vb.) Puanlama bilgisi

Araştırmada katılımcılara sorulan “Aldığınız geribildirimi anlamlandırmak için ders hocasına soru sormaya gerek duydunuz mu? Duyduysanız bu soru neydi?” sorusu için yalnızca gerçekten soru sormuş katılımcıların metin yanıtları dikkate alınmıştır. Soru sayısının az olması ($n < 20$) nedeniyle gelen yanıtlar üzerinde tema oluşturmak yerine doğrudan alıntılarla sunulmuştur. Ayrıca tüm sorular için cevaplardan alıntılara bulgular kısmında yer verilmiştir. Cevapların temalara göre ayrıştırılması sürecinde iki alan uzmanının görüşlerine başvurulmuş ve puanlayıcılar arası Fleiss Kappa güvenirlik katsayısı incelenmiştir. Puanlayıcılar arası güvenirliğe ilişkin bulgular Tablo 3.8’de sunulmuştur.

Tablo 3.8. Açık Uçlu Soruların Fleiss Kappa Katsayısına İlişkin Bulguları

	Uyuşma Yüzdesi	Şans Yüzdesi	Fleiss Kappa Katsayısı	<i>p</i>	%95 Güven Aralığı
1.Soru	0.984	0.535	0.967	0.0001	[0.921;1.000]
2.Soru	0.977	0.334	0.965	0.0001	[0.927;1.000]
3.Soru	0.964	0.415	0.949	0.0001	[0.897;0.999]
4.Soru	0.984	0.413	0.973	0.0001	[0.938;1.000]
6.Soru	0.992	0.333	0.988	0.0001	[0.966;1.000]

Tablo 3.8 incelendiğinde birinci soru için uzmanların uyuşma yüzdesinin %98.4 ve şans uyuşma yüzdesinin %53.5 olduğu görülmüştür. Fleiss Kappa katsayısı .967 olarak bulunmuştur. İkinci soru için uzmanların uyuşma yüzdesi %97.7; şans uyuşma yüzdesi %33.4; Fleiss Kappa katsayısı ise .965'tir. Üçüncü soru için bu değerler sırasıyla %96.4; %41.5 ve .949'dur. Dördüncü soru için uyuşma oranı %98.4; şans oranı %41.3; Fleiss Kappa katsayısı 0.973 olarak hesaplanmıştır. Altıncı soru için uyuşma yüzdesi %99.2; şans yüzdesi %33.3; Fleiss Kappa katsayısı ise .988'dir. Fleiss ve diğerleri (2003) elde edilen katsayıları yorumlarken 0,40'tan küçük değerleri zayıf, .40 ile .75 arası değerleri iyi ve .75'ten büyük değerleri çok iyi uyum olarak sınıflandırmaktadır. Buna göre uzmanlar arasında çok iyi düzeyde uyum olduğu gözlemlenmiştir.

Katılımcılara uygulanan dört farklı test koşulu, BOBUT kullanıcı deneyimini etkileyebilecek çeşitli değişkenleri içermektedir. Kullanıcı deneyimini etkileyen değişkenlerin uygulamalara göre farklılık gösterebildiği nitel araştırmalarda sıklıkla vurgulanmaktadır (Wislocki vd., 2024). Bu bağlamda her bir açık uçlu soruya verilen tematik yanıtların uygulanan test türüne göre anlamlı biçimde farklılaşıp farklılaşmadığını incelemek amacıyla ki-kare (χ^2) testi uygulanmıştır.

Ki-kare testi, özellikle kategorik veriler arasındaki ilişkinin bağımsız olup olmadığını test etmek için kullanılan yaygın ve güçlü bir istatistiksel yöntemdir (Hess & Hess, 2017). Bu yöntem temalar gibi sınıflandırılmış nitel cevapların farklı deneyim koşullarında benzer biçimde dağılıp dağılmadığını belirlemek açısından uygundur (Franke vd., 2012).

Katılımcıların BOBUT kullanıcı deneyiminin araştırılmasına ilişkin beş adet Likert anket maddesi uygulanmıştır. Bunlar;

1. Aldığınız teste yönelik genel memnuniyet düzeyinizi puanlayınız.

2. Test sonucunda aldığım geribildirim performansı doğru bir değerlendirmesi olduğunu düşünüyorum.
3. Bu geribildirimler pek uyuşmadığını düşünüyorum ve içeriğini yeterince objektif ve güvenilir bulmadığım için ciddiye almakta zorlanıyorum.
4. Aldığım geribildirim, gelecekteki denemelerimdeki motivasyonumu ve gösterdiğim çabayı doğrudan etkileyecek bir faktör olacak.
5. Gelecekte sınavlarımı bireye uyarlanmış olarak almak isterim.

Şeklinde sorulmuş olan anket sorularına ilişkin betimsel istatistikler Tablo 3.9’da verilmiştir.

Tablo 3.9. BOBUT Kullanıcı Deneyimi Anketine İlişkin Betimsel İstatistikler

		N	Min	Maks	$\bar{X}$	ss	Çarpıklık	Basıklık	p
Soru-1	Test 1	27	3	5	4.22	0.80	-0.438	-1.280	0.0001
	Test 2	42	1	5	3.66	0.97	-0.575	0.115	0.0010
	Test 3	33	3	5	4.33	0.69	-0.557	-0.712	0.0001
	Test 4	29	1	5	3.48	1.18	-0.788	0.061	0.0020
Soru-2	Test 1	27	2	5	4.44	0.84	-1.452	1.379	0.0001
	Test 2	42	1	5	3.92	1.11	-0.856	-0.136	0.0001
	Test 3	33	3	5	4.21	0.81	-0.423	-1.379	0.0001
	Test 4	29	1	5	3.48	1.35	-0.705	-0.647	0.0010
Soru-3	Test 1	27	1	5	3.48	1.67	-0.578	-1.339	0.0001
	Test 2	42	1	5	3.42	1.36	-0.297	-1.170	0.0001
	Test 3	33	1	5	3.45	1.43	-0.407	-1.226	0.0010
	Test 4	29	1	5	3.51	1.50	-0.375	-1.476	0.0001
Soru-4	Test 1	27	2	5	4.14	0.94	-0.607	-0.970	0.0001
	Test 2	42	1	5	3.45	1.32	-0.449	-0.799	0.0001
	Test 3	33	1	5	3.72	1.17	-0.767	-0.081	0.0010
	Test 4	29	1	5	3.20	1.34	-0.405	-0.923	0.0050
Soru-5	Test 1	27	3	5	4.33	0.83	-0.721	-1.156	0.0001
	Test 2	42	1	5	3.88	1.21	-1.051	0.318	0.0001
	Test 3	33	1	5	3.96	1.13	-1.177	1.080	0.0001
	Test 4	29	1	5	3.86	1.21	-0.996	0.274	0.0001

Katılımcılara yöneltilen ankette yer alan Likert tipi beş maddeden elde edilen verilere ilişkin betimsel istatistikler incelendiğinde Shapiro-Wilk normallik test sonuçlarına göre tüm

maddeler tüm test koşulları için anlamlı bulunmuştur (p değeri $> .05$) Bu bulgu, verilerin normal dağılım göstermediğini ortaya koymaktadır. Analize ait etki büyüklüklerini belirlemek amacıyla Epsilon kare değeri hesaplanmıştır. Epsilon kare için .10 küçük düzeyde, .25 orta düzeyde ve .37 büyük düzeyde etki büyüklüğünü ifade etmektedir (Büyüköztürk vd., 2018).

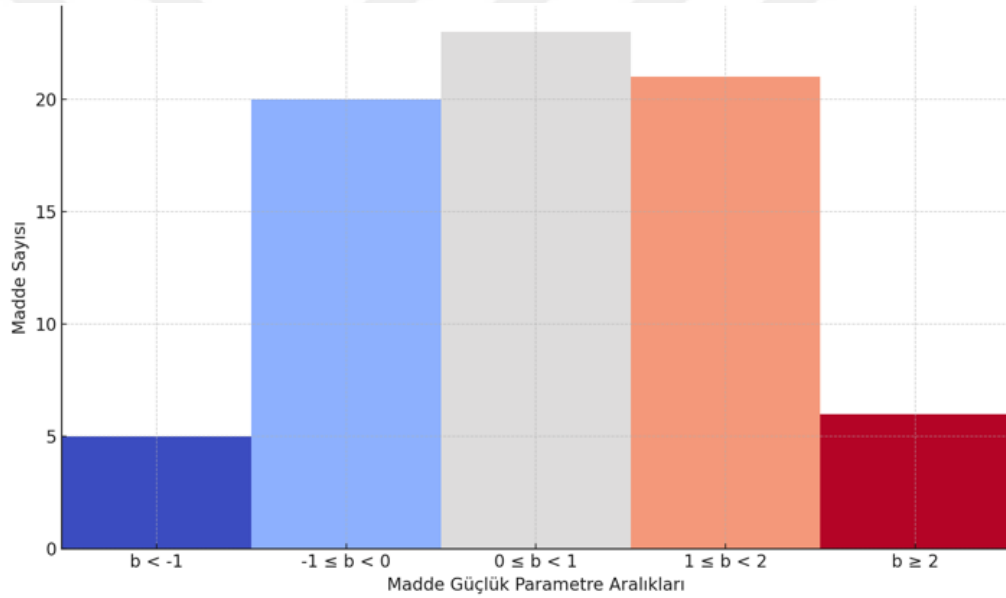


BÖLÜM IV

BULGULAR

4.1. Madde Parametreleri

BOBUT kullanıcı deneyiminin araştırılması amacıyla uygulanan çoktan seçmeli maddelere ilişkin madde parametreleri bu bölümde sunulmuştur. 1PL modelinde ayırt edicilik parametrelerinin 1'e sabitlenmesiyle elde edilen model "Rasch modeli" olarak adlandırılmaktadır. Bu modelde yalnızca madde güçlük parametreleri kestirilmektedir. Madde havuzuna ilişkin madde güçlük parametreleri incelendiğinde, değerlerin -2.07 ile +2.47 aralığında değiştiği gözlemlenmiştir. Madde güçlük parametrelerinin normal bir dağılım gösterdiği anlaşılmış olup bu dağılıma ilişkin histogram Şekil 4.1 'de sunulmuştur.



Şekil 4.1. Maddelere Ait Güçlük Parametrelerinin Dağılımı

Madde havuzuna ilişkin madde güçlük ortalaması .49, medyanı ise .70 olarak bulunmuştur. Bu durum madde havuzunun orta-üst yetenek düzeyine sahip bireylerin ölçülmesinde daha güvenilir sonuçlar verdiğini göstermektedir. Konulara ilişkin madde güçlük parametreleri Tablo 4.1–4.4'te sunulmuştur.

Tablo 4.1. Temel Kavramlar Konusuna İlişkin Maddelere Ait Madde Parametreleri

Madde	a	b	c	d
1.madde	1.0	-0.252	0.0	1.0
2.madde	1.0	-0.373	0.0	1.0
11.madde	1.0	-0.478	0.0	1.0
12.madde	1.0	1.097	0.0	1.0
13.madde	1.0	0.916	0.0	1.0
14.madde	1.0	-0.099	0.0	1.0
15.madde	1.0	1.621	0.0	1.0
17.madde	1.0	0.050	0.0	1.0
81.madde	1.0	2.205	0.0	1.0
82.madde	1.0	-0.079	0.0	1.0
83.madde	1.0	-1.609	0.0	1.0
84.madde	1.0	1.121	0.0	1.0
85.madde	1.0	-0.398	0.0	1.0
86.madde	1.0	-0.538	0.0	1.0
88.madde	1.0	0.803	0.0	1.0
89.madde	1.0	-0.834	0.0	1.0
90.madde	1.0	0.655	0.0	1.0

Temel kavramlar konusuna ait maddelere ilişkin madde güçlük parametreleri incelendiğinde, maddelerin -1.60 ile +2.20 aralığında değiştiği gözlenmiştir. Maddelerin yaklaşık %52.5'inin (n=9) negatif madde güçlük parametresine ve yaklaşık %47.5'inin (n=8) pozitif madde güçlük parametresine sahip olduğu görülmüştür. Temel kavramlar konusundan madde havuzuna alınan maddelerin ortalama madde güçlüğü .22, medyanı ise -.08 olarak bulunmuştur. Bu bulgular, konuya ait maddelerin orta düzeyde yeteneğe sahip bireylerin ölçülmesinde daha uygun sonuçlar verdiğini göstermektedir. Ayrıca bu konu ortalama madde güçlüğü açısından en kolay maddelerin bulunduğu konu olarak öne çıkmaktadır.

Tablo 4.2. Ölçme Araçlarının Nitelikleri Konusuna İlişkin Maddelere Ait Madde Parametreleri

Madde	a	b	c	d
3.madde	1.0	-0.272	0.0	1.0
4.madde	1.0	-0.393	0.0	1.0
5.madde	1.0	0.971	0.0	1.0
18.madde	1.0	2.472	0.0	1.0
19.madde	1.0	0.124	0.0	1.0
20.madde	1.0	0.788	0.0	1.0
21.madde	1.0	0.916	0.0	1.0
22.madde	1.0	1.005	0.0	1.0
23.madde	1.0	-0.211	0.0	1.0
24.madde	1.0	1.192	0.0	1.0
25.madde	1.0	1.050	0.0	1.0
26.madde	1.0	0.389	0.0	1.0
27.madde	1.0	0.960	0.0	1.0
28.madde	1.0	-0.886	0.0	1.0
29.madde	1.0	-0.929	0.0	1.0
30.madde	1.0	1.562	0.0	1.0
51.madde	1.0	-0.617	0.0	1.0
52.madde	1.0	-1.173	0.0	1.0
53.madde	1.0	0.743	0.0	1.0
54.madde	1.0	0.830	0.0	1.0
55.madde	1.0	1.157	0.0	1.0
56.madde	1.0	0.416	0.0	1.0
57.madde	1.0	1.108	0.0	1.0
58.madde	1.0	0.260	0.0	1.0
59.madde	1.0	2.199	0.0	1.0
61.madde	1.0	0.032	0.0	1.0

Ölçme aracının nitelikleri konusuna ait maddelere ilişkin madde güçlük parametreleri incelendiğinde, maddelerin -1.17 ile +2.47 aralığında değiştiği gözlenmiştir. Maddelerin yaklaşık %26.9'unun (n=7) negatif madde güçlük parametresine ve yaklaşık %73.1'inin (n=19) pozitif madde güçlük parametresine sahip olduğu görülmüştür. Ölçme aracının nitelikleri konusundan madde havuzuna alınan maddelerin ortalama madde güçlüğü .53, medyanı ise .77 olarak bulunmuştur. Bu bulgular konu maddelerinin orta-üst düzeyde yeteneğe sahip olan bireylerin ölçülmesinde daha uygun sonuçlar verdiğini göstermektedir.

Tablo 4.3. Ölçme Araçları Konusuna İlişkin Maddelere Ait Madde Parametreleri

Madde	a	b	c	d
6.madde	1.0	1.357	0.0	1.0
7.madde	1.0	0.855	0.0	1.0
31.madde	1.0	-1.721	0.0	1.0
32.madde	1.0	1.620	0.0	1.0
35.madde	1.0	-2.074	0.0	1.0
37.madde	1.0	1.426	0.0	1.0
39.madde	1.0	1.137	0.0	1.0
40.madde	1.0	1.306	0.0	1.0
62.madde	1.0	-0.737	0.0	1.0
63.madde	1.0	1.208	0.0	1.0
64.madde	1.0	0.495	0.0	1.0
65.madde	1.0	0.495	0.0	1.0
66.madde	1.0	-1.530	0.0	1.0
68.madde	1.0	-0.231	0.0	1.0
69.madde	1.0	0.701	0.0	1.0
70.madde	1.0	1.423	0.0	1.0

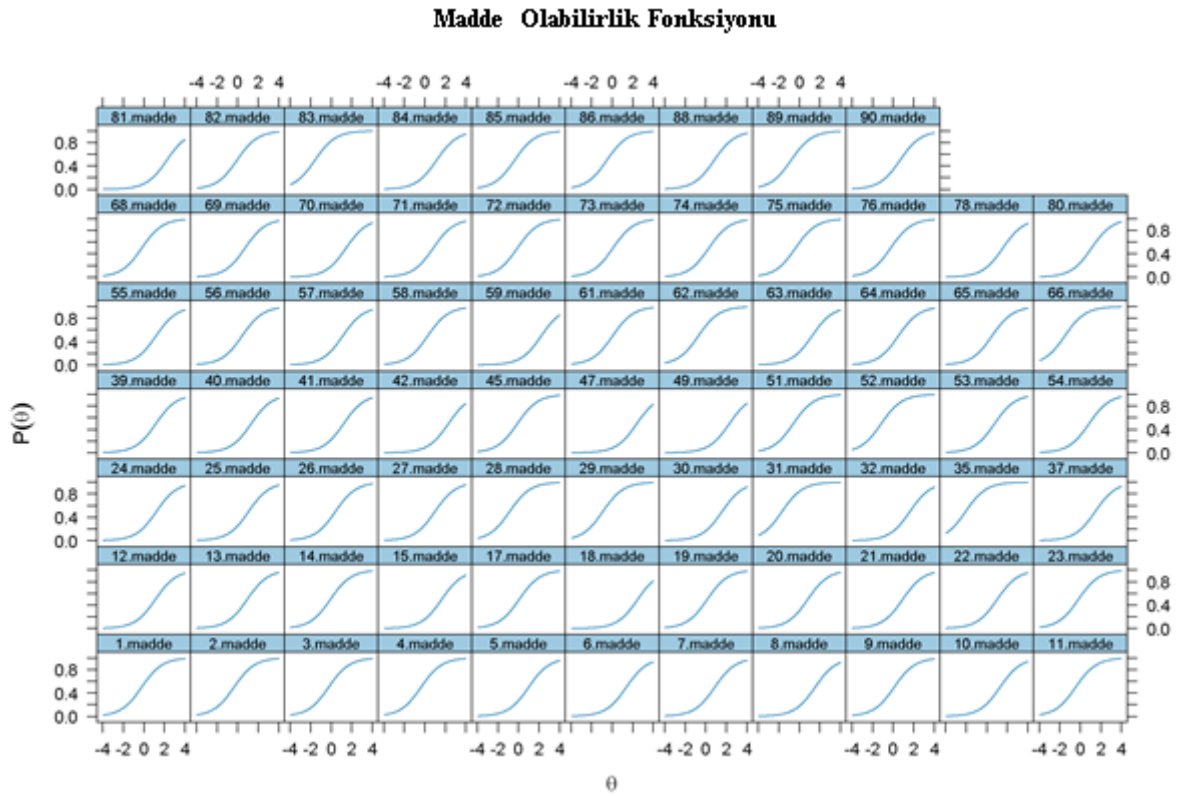
Ölçme araçları konusuna ait maddelere ilişkin madde güçlük parametreleri incelendiğinde, maddelerin -2.07 ile +1.42 aralığında değiştiği gözlenmiştir. Maddelerin yaklaşık %31.5'inin (n=5) negatif madde güçlük parametresine ve yaklaşık %68.5'inin (n=11) pozitif madde güçlük parametresine sahip olduğu görülmüştür. Ölçme araçları konusundan madde havuzuna alınan maddelerin ortalama madde güçlüğü .36, medyanı ise .78 olarak bulunmuştur. Bu bulgular, konu maddelerinin orta-üst düzeyde yeteneğe sahip olan bireylerin ölçülmesinde daha uygun sonuçlar verdiğini göstermektedir.

Tablo 4.4. Madde ve Test İstatistiği Konusuna İlişkin Maddelere Ait Madde Parametreleri

Madde	a	b	c	d
8.madde	1.0	1.503	0.0	1.0
9.madde	1.0	0.084	0.0	1.0
10.madde	1.0	1.275	0.0	1.0
41.madde	1.0	1.192	0.0	1.0
42.madde	1.0	2.270	0.0	1.0
45.madde	1.0	-0.134	0.0	1.0
47.madde	1.0	2.375	0.0	1.0
49.madde	1.0	2.270	0.0	1.0
71.madde	1.0	0.803	0.0	1.0
72.madde	1.0	-0.124	0.0	1.0
73.madde	1.0	0.056	0.0	1.0
74.madde	1.0	0.281	0.0	1.0
75.madde	1.0	-0.398	0.0	1.0
76.madde	1.0	-0.260	0.0	1.0
78.madde	1.0	1.551	0.0	1.0
80.madde	1.0	1.066	0.0	1.0

Madde ve test istatistiği konusuna ait maddelere ilişkin madde güçlük parametreleri incelendiğinde, maddelerin -0.39 ile $+2.37$ aralığında değiştiği gözlenmiştir. Maddelerin %25.0'nin ($n=4$) negatif madde güçlük parametresine ve %75.0'nin ($n=12$) pozitif madde güçlük parametresine sahip olduğu görülmüştür. Madde ve test istatistiği konusundan madde havuzuna alınan maddelerin ortalama madde güçlüğü $.86$, medyanı ise $.93$ olarak bulunmuştur. Bu bulgular, konu maddelerinin üst düzeyde yeteneğe sahip olan bireylerin ölçülmesinde daha uygun sonuçlar verdiğini göstermektedir. Ayrıca bu konu, ortalama madde güçlüğü açısından en zor maddelerin bulunduğu konudur. Çıkarılan maddelerden sonra maddeler tekrar sırayla numaralandırılmış ve madde havuzuna ilişkin grafikler bu şekilde yazdırılmıştır.

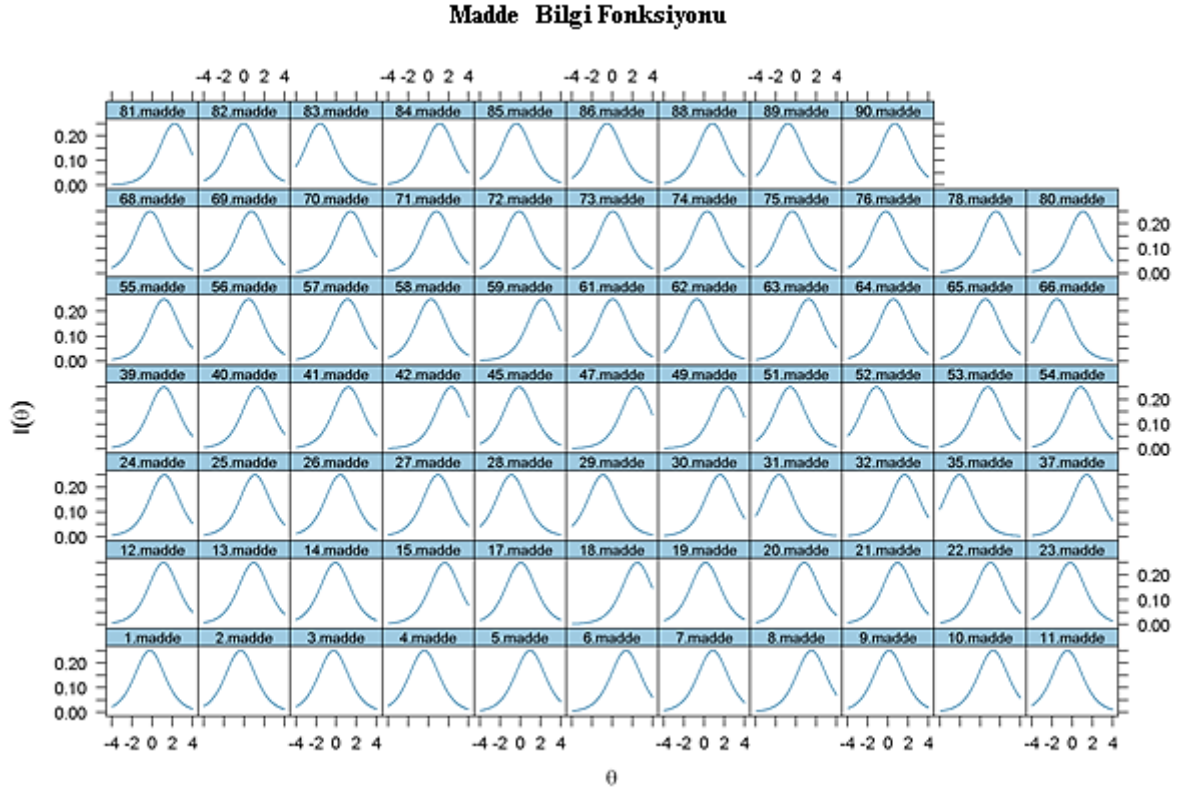
4.2. MKE ve Madde Bilgi Fonksiyonu



Şekil 4.2. Madde Havuzundaki Maddelere İlişkin Madde Karakteristik Eğrileri

Şekil 4.2’de yer alan madde karakteristik eğrileri incelendiğinde Rasch modelin kuramsal yapısına uygun olarak tüm eğrilerin eğimlerinin aynı olduğu görülmektedir. Tablo 4.1–4.4 ile uyumlu biçimde maddelere ait Madde Karakteristik Eğrileri’nin (MKE) orta-üst

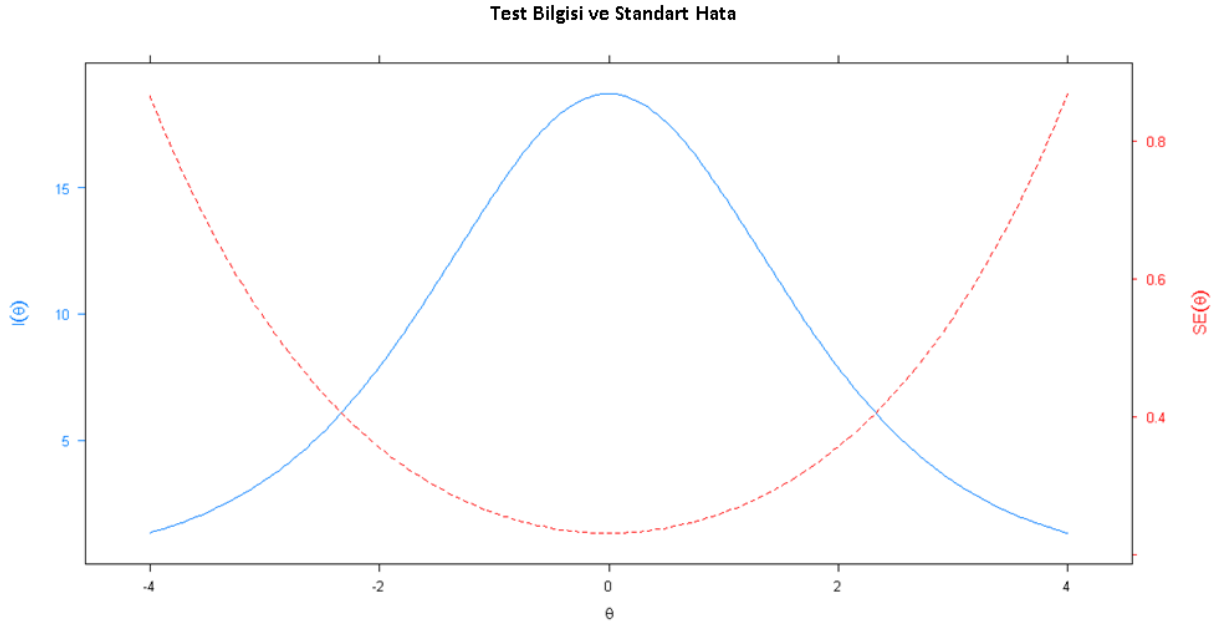
düzey θ 'ya sahip bireylerin ölçülmesi için daha uygun olduğu gözlenmektedir. Diğer bir deyişle maddelerin büyük çoğunluğunda düşük θ düzeyine sahip bireylerin maddeyi doğru yanıtlama olasılığı düşüktür. Maddelere ilişkin bilgi fonksiyonları Şekil 4.3'te sunulmuştur.



Şekil 4.3. Madde Havuzundaki Maddelere İlişkin Madde Bilgi Fonksiyonları

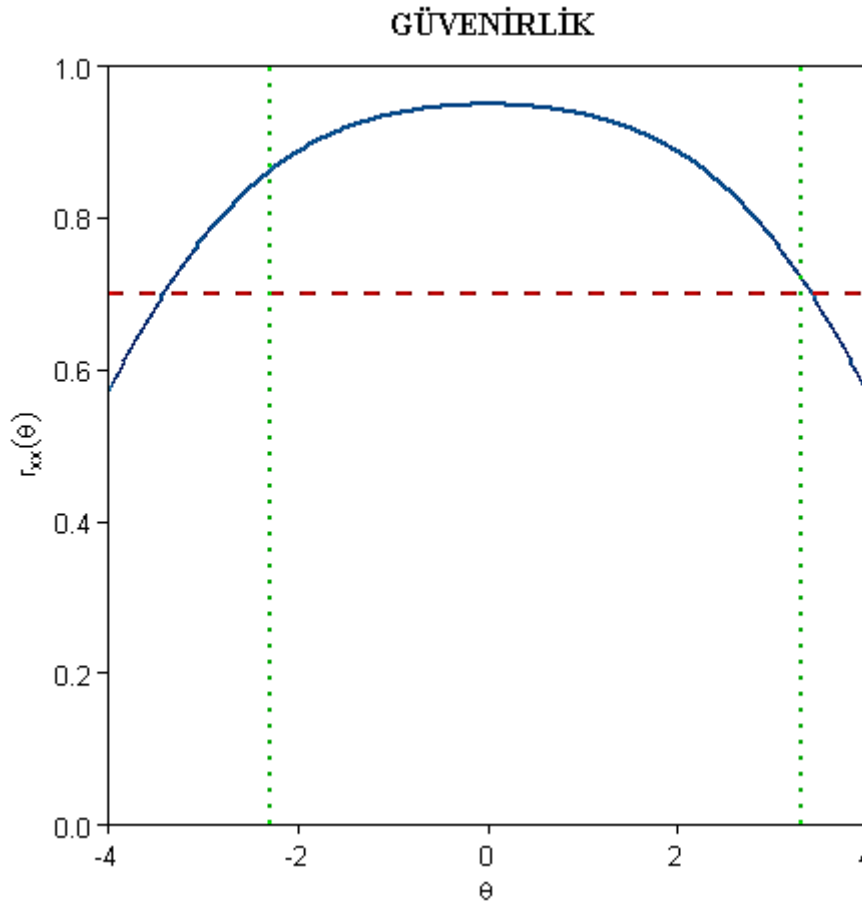
Madde bilgi fonksiyonlarına ilişkin Şekil 4.3 incelendiğinde her bir maddenin bilgi fonksiyonunun ilgili madde güçlük parametresi civarında tepe noktasına ulaştığı gözlemlenmektedir. Bu tepe noktası, maddenin hangi yetenek düzeyinde en fazla bilgi sağladığını göstermektedir. Grafikler incelendiğinde maddelerin orta-üst düzey θ 'ya sahip bireyler için daha fazla bilgi verdiği anlaşılmaktadır.

4.3. Madde Havuzu Güvenirliđi ve Hatası



Şekil 4.4. Madde Havuzuna İlişkin Test Bilgi Fonksiyonu ve Standart Hata

BOBUT madde havuzuna ilişkin test bilgi fonksiyonu ve standart hatası Şekil 4.4 üzerinden incelendiğinde mavi eğri test bilgi fonksiyonunu, kırmızı kesikli eğri ise standart hata (SE) eğrisini göstermektedir. Test bilgi fonksiyonu eğrisi madde havuzunun $-1 \leq I(\theta) < 2$ aralığında en yüksek bilgiyi sağladığını ve bu aralıkta daha güvenilir olduğunu göstermektedir. Ayrıca test bilgi fonksiyonu eğrisinin tepe noktası $\theta \approx 1$ civarındadır; bu nokta aynı zamanda en düşük standart hata düzeyine ($SE \approx .25$) sahiptir. Diğer bir deyişle, testin orta-yüksek düzeyde θ 'ya sahip bireyleri en güvenilir biçimde ölçtüğü söylenebilir. Bu bulgu Tablo 4.1–4.4 arasında yer alan madde güçlük parametreleri ile tutarlıdır.



Şekil 4.5. Madde Havuzuna İlişkin Güvenirlik

BOBUT uygulaması için kullanılan madde havuzuna ilişkin güvenilirlik eğrisinin yer aldığı Şekil 4.5 incelendiğinde testin güvenilirliğinin yetenek düzeylerine göre değişiklik gösterdiği görülmektedir. Güvenirliğin 0–2 θ düzeyinde en yüksek seviyelere ulaşması, bu aralıkta testin daha düşük standart hatayla daha fazla bilgi sağladığını bir kez daha ortaya koymaktadır. Eğri, $\theta \approx 1$ civarında tepe noktasına ulaşarak $r_{xx}(\theta) \approx .90$ düzeyine erişmiştir. Bu bulgu, orta-üst düzey θ değerlerinde testin yüksek güvenirlğe sahip olduğunu göstermektedir. Eğrinin genel ortalaması alındığında elde edilen güvenilirlik düzeyi $r_{xx}(\theta) = .71$ olup, bu değer kabul edilebilir bir güvenilirlik düzeyine işaret etmektedir. Madde güçlük parametrelerinin yaklaşık %66'sının sıfırdan büyük olması, madde havuzunun orta-üst düzey θ 'ya sahip bireylerin yetenek düzeyini belirlemede, düşük-orta düzey θ 'ya sahip bireylere kıyasla daha güvenilir olduğunu göstermektedir. Nihai madde havuzunu oluşturan konulara göre yapılan değerlendirmede; temel kavramlar konusuna ait maddelerin orta düzey θ 'ya sahip bireyleri ölçmede daha güvenilir olduğu, ölçme aracının nitelikleri ve ölçme araçları

konularındaki maddelerin orta-üst düzey θ 'ya sahip bireyleri ölçmede daha güvenilir olduğu ve madde ile test istatistiği konusuna ait maddelerin de yine orta-üst düzey θ 'ya sahip bireyleri ölçmede yüksek güvenilirlik sunduğu görülmektedir.

4.4. Post-hoc Sonuçları

Tablo 4.5. Post-Hoc Simülasyon Sonuçları

Madde Seçimi					MFI			
θ Kestirimi					MAP			
Test Başlangıç θ					$\theta = 0.0$			
					$\theta = random$			
Sonlandırma	SE = 0.30	SE = 0.40	SE = 0.50	10 Madde	SE = 0.30	SE = 0.40	SE = 0.50	10 Madde
Kuralı								
Korelasyon	0.950	0.908	0.858	0.918	0.945	0.913	0.865	0.912
BİAS	0.005	0.0003	0.016	-0.012	0.0009	-0.002	-0.014	0.026
RMSEA	0.304	0.399	0.501	0.391	0.310	0.394	0.502	0.391
Test Uzunluğu	18.77	9.163	5.423	10,00	19.49	9.513	6.148	10,00
MeanSE	0.299	0.390	0.477	0.368	0.296	0.389	0.476	0.379
θ Kestirimi					MLE			
Test Başlangıç θ					$\theta = 0.0$			
					$\theta = random$			
Sonlandırma	SE = 0.30	SE = 0.40	SE = 0.50	10 Madde	SE = 0.30	SE = 0.40	SE = 0.50	10 Madde
Kuralı								
Korelasyon	0.962	0.914	0.848	0.930	0.939	0.915	0.884	0.910
BİAS	0.004	0.005	0.009	0.023	-0.009	0.0003	-0.066	-0.004
RMSEA	0.309	0.407	0.495	0.382	0.292	0.404	0.502	0.395
Test Uzunluğu	19.38	9.015	5.383	10,00	18.251	9.751	6.179	10,00
MeanSE	0.297	0.390	0.476	0.370	0.296	0.389	0.476	0.379

BOBUT kullanıcı deneyimine etki eden değişkenlerin belirlenmesi adına uygulanacak BOBUT'un koşullarının belirlenmesi adına yapılan post-hoc simülasyon sonuçlarına Tablo 4.5'te yer verilmiştir. Simülasyonlar 500 tekrar ile yapılmıştır. Simülasyon sonucunda değişkenliği incelenen BOBUT koşulları:

- θ kestirim modeli: MAP (Maximum a posteriori) ve ML (Maximum Likelihood)
- İlk madde seçimi: θ ; rastgele θ ve $\theta=0.0$
- Durdurma kuralı: sabit uzunluk=10 madde ve 0.30, 0.40 ve 0.50 standart hata

Şeklinde. Post-hoc simülasyonunda korelasyon katsayısı, BIAS, RMSEA, kullanılan ortalama madde sayısı ve ortalama standart hata değerleri incelenmiştir. Diğer değişkenler

sabit tutulduğunda MAP ve MLE kestirimlerinin doğruluk göstergelerinin oldukça yakın olduğu görülmektedir. Bu durum kestirim yönteminin ölçme kalitesinde yarattığı farkların ihmal edilebilir düzeyde olduğunu göstermektedir. Testin başlangıcında seçilecek ilk maddenin rastgele θ veya $\theta=0.0$ şeklinde değişkenlik gösterdiği koşullar için simülasyonlar arasında göz ardı edilebilir farklar olduğu gözlemlenmiştir. Ancak $\theta=0.0$ koşulu madde kullanım sıklığının ihmal edildiği gözlemlenmiştir. Bundan dolayı rastgele θ koşulu seçilmiştir. Simülasyon sonuçları incelendiğinde $SE = 0.30$ koşulunda test uzunluğunun 18 madde altına inmediği görülmüştür. Durdurma kuralının $SE = 0.40$ olarak belirlendiği durumlarda testin dokuz maddede sonlandığı görülmüştür. Durdurma kuralının $SE = 0.50$ olarak belirlendiği durumlarda testin beş ya da altı maddede sonlandığı görülmüştür. Durdurma kuralı olarak sabit 10 madde ile sonlandırma sonuçlarının optimum koşulları sağladığını göstermektedir. Özetle 10 maddelik test durdurma kuralı, madde seçiminde MFI, yetenek kestiriminde MAP ve başlangıç θ 'sı rastgele olacak şekilde belirlenmiştir. Bu koşullara göre $r = 0.912$; $BIAS = 0.026$; $RMSEA = 0.391$; $Mean_{SE} = 0.379$ olarak belirlenmiştir.

4.5. BOBUT Uygulamasına İlişkin Betimsel İstatistikler

Tablo 4.6. Canlı BOBUT Sonuçlarına İlişkin Betimsel İstatistikler

	$\bar{\theta}$	$\bar{\theta}(\%95CI)$	$\overline{SEM}$
Test-1*	-0.22	-0.27 / -0.17	0.51
Test-2**	-0.12	-0.17 / -0.07	0.51
Test-3***	-0.02	-0.07 / 0.02	0.51
Test-4****	-0.09	-0.14 / -0.04	0.51

*;Geniş geribildirim/Süre gösteriliyor,**Dar geribildirim/Süre gösteriliyor,***Geniş geribildirim /Süre gösterilmiyor,****Dar geribildirim /Süre gösterilmiyor

Canlı BOBUT uygulaması sonucunda Tablo 4.6 incelendiğinde geniş geribildirim ve sürenin gösterildiği koşulları içeren test-1 için $\bar{\theta} = -0.22$ ve $\overline{SEM}=0.51$ olarak bulunmuştur. Test-2 için $\bar{\theta} = -0.12$ ve $\overline{SEM}=0.51$ olarak bulunmuştur. Test-3 için $\bar{\theta} = -0.02$ ve $\overline{SEM}=0.51$ olarak bulunmuştur. Test-4 için $\bar{\theta} = -0.09$ ve $\overline{SEM}=0.51$ olarak bulunmuştur. Testlere ilişkin $\bar{\theta}(\%95CI)$ değerleri incelendiğinde grupların benzer özellik gösterdikleri gözlenmiştir. Bu durum grupların yansız olarak seçildiğini göstermektedir.

4.6. BOBUT Kullanıcı Deneyimi Görüşme Formuna İlişkin Bulgular

BOBUT kullanıcı deneyiminin araştırılması adına bireylere sorulmuş olan açık uçlu sorular için cevaplar temalara ayrılmıştır. Temalara ilişkin puanlayıcı güvenilirliği hesaplanmış ve sonuçları Tablo 3.8’de verilmiştir. Buna göre bireylere sorulmuş olan “Test sırasında aldığınız geribildirim uzunluğu hakkında düşüncelerinizi detaylandırabilir misiniz? Bu geribildirim motivasyonunuz ve genel deneyiminiz üzerindeki etkilerini örneklerle nasıl açıklarsınız?” sorusuna ilişkin ki kare analizi sonuçları Tablo 4.7’de verilmiştir.

Tablo 4.7. BOBUT Kullanıcı Deneyimi Motivasyonu Düzeyine İlişkin Ki Kare

		Olumlu	Geliştirilmeli /olumsuz	Toplam
1.test	n	22	5	27
	%	81.5	18.5	100.0
2.test	n	20	22	42
	%	47.6	52.4	100.0
3.test	n	27	6	33
	%	81.8	18.2	100.0
4.test	n	15	14	29
	%	51.7	48.3	100.0
Toplam	n	84	47	131
	%	64.1	35.9	100.0

$$\chi^2 = 14.93; sd = 3; p(0.001) < 0.05$$

Tablo incelendiğinde farklı koşullara sahip olan testler ile bireylerin aldıkları geribildirime ilişkin görüşleri incelendiğinde istatistiksel olarak anlamlı bir fark vardır ($\chi^2(3) = 14.93 ; p < 0.05$). Tablodaki yüzdeliklere göre geniş geribildirime sahip ve sürenin gösterildiği test olan test-1’i alan bireylerin %81.5’inin aldıkları geribildirim motivasyonlarını olumlu etkilediği görülmektedir. Ayrıca yine geniş geribildirime sahip olan ancak sürenin gösterilmediği test olan test-2’yi alan bireylerin %81.8’inin aldıkları geribildirim motivasyonlarının olumlu etkilendiği görülmektedir. Dar geribildirime sahip ve sürenin gösterildiği test olan test-2 için bireylerin %47.6’sının aldıkları geribildirim motivasyonlarını arttırdığını, %52.4’ü ise geribildirim daha iyi geliştirilmesi gerektiğini ya da olumsuz etkilediğini söylemiştir. Dar geribildirime sahip ve sürenin gösterilmediği test olan test-4’i alan bireylerin %61.1’i aldıkları geribildirim motivasyonunu olumlu etkilediği

ve %35.9'u aldıkları geribildirim daha iyi geliştirilmesi veya motivasyonlarını olumsuz etkilediği gözlemlenmiştir. Ayrıca, “Test sırasında aldığınız geribildirim uzunluğu hakkında düşüncelerinizi detaylandırabilir misiniz? Bu geribildirim motivasyonunuz ve genel deneyiminiz üzerindeki etkilerini örneklerle nasıl açıklarsınız?” sorusuna ilişkin bireylerin görüşleri doğrudan alıntılarla desteklenmiştir.

Test-1

- “Verilen dönütün yeterli olduğunu düşünüyorum. Kendimi geliştirmem ve eksik yönlerim hakkında doğru yönlendirme yapılması açısından uygun olduğunu düşünüyorum.”(Olumlu) (T1-KD02)
- “Her şey detaylarıyla açıklanmış. Hataların ders çıkarılabilir düzeyde olduğu belirtilmiş. Gayet faydalıydı.”(Olumlu) (T1-KD14)
- “Grafik görmek ilgi çekiciydi. Uzun paragrafı okumadım ama koyu renkle yazılmış kısmı okudum.”(Olumsuz/Geliştirilmeli) (T1-KD20)
- “Çok ayrıntılı güzeldi. Ancak çok fazla teknik ve terimsel değerlendirme olduğu için anlaması zordu.”(Olumsuz/Geliştirilmeli) (T1-KD16)
- “Hangi soruları yanlış yaptığımı görmek isterdim. Daha detaylı dönütler verebilirsiniz.”(Olumsuz/Geliştirilmeli) (T1-KD12)

Test-2

- “Bir cümle şeklinde bir geribildirim aldım. Gruba göre başarıyı görmek motivasyonumun artmasını sağladı.” (Olumlu) (T2-KD01)
- “Geri bildiriminiz açık ve anlaşılırdı. Benim için olumlu bir deneyimdi.” (Olumlu) (T2-KD40)
- “Net sayımın yazılmasını yanlış yaptığım soruların doğrusunu görmek isterdim. Yüzdelik olarak iyi olmam mutlu etti.” (Olumsuz/Geliştirilmeli) (T2-KD31)
- “Motivasyonum üzerinde iyi etkisi olmadı kendimi diğerleri ile kıyaslandım” (Olumsuz/Geliştirilmeli) (T2-KD17)

Test-3

- “Test sırasında aldığım geribildirim uzunluğu gayet yeterli bir düzeydedir. Geri bildirim almak, özellikle de başarılı olduğumu ifade eden bir geri bildirim almak mutluluk verici oluyor.”(Olumlu) (T3-KD24)

- “Yeterli düzeyde bilgilendirici ve aydınlatıcı bir geri bildirim aldım. Ne çok uzun ne çok kısa. Geri bildirim kendi seviyemle ilgili tam olarak hissettiğim ve düşündüğüm şeyleri ifade ediyor.” (Olumlu) (T3-KD14)
- “Geribildirim biraz uzundu belki kısa ve çarpıcı tek bir cümle olsa daha etkili olabilirdi. Kendimi yeterli hissetmeme sebep oldu genel olarak korkulan bir ders olan ölçmeden ben de korkarım ama geribildirim demek ki yapabiliyormuşum daha iyisi de olabilir diye kaçmam yerine üzerine gitmem noktasında beni motive etti.” (Olumsuz/Geliştirilmeli) (T3-KD29)
- “Aldığım geri bildirim Tam olarak anlayamadım çünkü grafiksel bir şekilde şekillenmişti onun dışında geribildirim tatmin ediciydi. Özellikle soru sayısı beni memnun etti” (Olumsuz/Geliştirilmeli) (T3-KD12)

Test-4

- “Geribildirim uzunluğu açıklayıcı olması açısından gayet iyiydi. Geribildirim motivasyonumu arttırdı ve bende olumlu bir etki bıraktı.”(Olumlu) (T4-KD23)
- “Gayet yeterli bir açıklama idi. Mevcut bilgilerimle benzer grup içindeki yerimi görme şansı oldukça motive edici” (Olumlu) (T4-KD19)
- “Çok kısa olduğunu düşünüyorum. Soruyu ilk okuduğumda geribildirim neredeydi ki diye düşündüm akılda kalıcı ya da motive edici olduğunu düşünmüyorum” (Olumsuz/Geliştirilmeli) (T4-KD29)
- “Kısaydı ve neye göre bana benzer grup daha açıklayıcı olabilirdi” (Olumsuz/Geliştirilmeli) (T4-KD18)

Şeklindedir. Bireylere sorulan “Test sırasında sürenin gösterilmesi/gösterilmemesi, zaman yönetiminizi ve stres düzeyinizi nasıl şekillendirdi?” sorusuna ilişkin bulgular Tablo 4.8’de verilmiştir.

Tablo 4.8. BOBUT Kullanıcı Deneyimi Stres Düzeyine İlişkin Ki Kare

		Pozitif	Nötr	Negatif	Toplam
1.test	n	11	6	10	27
	%	40.7	22.2	37.0	100.0
2.test	n	6	17	19	42
	%	14.3	40.5	45.2	100.0
3.test	n	19	4	10	33
	%	57.6	12.1	30.3	100.0
4.test	n	11	13	5	29
	%	37.9	44.8	17.2	100.0
Toplam	n	47	40	44	131
	%	35.9	30.5	33.6	100.0

$$\chi^2 = 21.78; sd = 6; p(0.001) < 0.05$$

Tablo incelendiğinde bireylerin farklı koşullara sahip olan testler ile bireylerin aldıkları testte sürenin gösterilmesi/gösterilmemesi durumuna ilişkin görüşleri incelendiğinde istatistiksel olarak anlamlı bir fark vardır ($\chi^2(6) = 21.78 ; p < 0.05$). Tablodaki yüzdeliklere göre sürenin gösterildiği ve geniş geribildirim verildiği test-1 için bireylerin %40.7’si zaman yönetiminin pozitif ve stres düzeyinin negatif etkilendiği, %22.2’si herhangi bir etkinin olmadığı ve %37.0’sinin negatif etkilendiği gözlemlenmiştir. Yine sürenin gösterildiği ancak dar geribildirim verilen test-2 için bireylerin %14.3’ü zaman yönetiminin pozitif ve stres düzeyinin negatif etkilendiği, %40.5’i herhangi bir etkinin olmadığı ve %45.2’si negatif etkilendiği gözlemlenmiştir. Sürenin gösterilmediği ve geniş geribildirim verilen test3 için bireylerin %57.6’sı zaman yönetiminin pozitif ve stres düzeyinin negatif etkilendiği, %12.1’i herhangi bir etkinin olmadığı ve %30.3’ü negatif etkilendiği gözlemlenmiştir. Sürenin gösterilmediği ve dar geribildirim verilen test-4 için bireylerin %35.9’u zaman yönetiminin pozitif ve stres düzeyinin negatif etkilendiği, %30.5’i herhangi bir etkinin olmadığı ve %33.6’sı negatif etkilendiği gözlemlenmiştir. Ayrıca, “Test sırasında sürenin gösterilmesi/gösterilmemesi, zaman yönetiminizi ve stres düzeyinizi nasıl şekillendirdi?” sorusuna ilişkin bireylerin görüşleri doğrudan alıntılarla desteklenmiştir.

Test-1

- “Sürenin gösterilmesi beni olumlu etkiledi. Kendi açımdan sorulara daha rahat odaklanabildim.”(Pozitif) (T1-KD08)

- “Stresimi süre etkilemedi. ama zor sorularda süre yönetimimi yapmak adına kontrol iyi oldu”(Nötr) (T1-KD09)
- “Her soru başına verilen süre miktarı gayet idealdi. Süreyi görmek ufak bir stres oluştursa da bu şekilde olmasını oldukça beğendim.”(Negatif) (T1-KD22)
- “Stres altında hissettim. Sanki gerçek bir sınavda sınavı yetiştirememek üzerine bir panik yaşadım.”(Negatif) (T1-KD18)

Test-2

- “Süre, daha öncesinde girmiş olduğum testlere göre daha fazlaydı bundan dolayı çok stres olmadan rahat rahat soruları çözebildim. Süreyi de görüyor olmak beni daha çok rahatlattı.”(pozitif) (T2-KD42)
- “Zamanımın azaldığını görmek hızlı çözmeme gerektiğini hissettirdiği için biraz stres olmama sebep oldu. Test ilerledikçe stresim azaldı.”(nötr) (T2-KD01)
- “Süre yönetimimi olumsuz etkiledi dikkatim sürede kaldı okuduğum cümleyi anlayamadım. Orta düzey bir stres yarattı.”(negatif) (T2-KD32)
- “Süreye daha çok odaklanmama ve her süreye baktığımda en az 3 saniye sorudan kopmama neden oldu. Eğer ki süre tüm sınav için verilmiş olsaydı daha stres yaratıcı olurdu. Soru bazlı süre daha az strese sebep oldu.”(negatif) (T2-KD35)

Test-3

- “Zamanın olmaması tabi ki stres düzeyini en alt seviyelerde tuttu. Kendi var olan bilgilerini kontrol ederek, kendini analiz ederek çözme şansı sundu.”(Pozitif) (T3-KD09)
- “Test sırasında sürenin gösterilmemesi beni strese sokmadı. Strese girmedığım için soruları cevaplarken zamanı yönetmek konusunda da zorlanmadım.” (Pozitif) (T3-KD24)
- “Ne kadar zamanımın olduğunu ve kaldığını bilmemekten biraz kaygı duydum. süreyi bilmek de bir miktar kaygılandırırsa da belki kontrolün bende olduğunu bilmek iyi olabilirdi. Yani iki durumun da avantajı dezavantajı var benim için.”(Negatif) (T3-KD29)
- “Testi çözerken sürenin gösterilmesi zaman planlaması açısından faydalı olacaktır. Şahsen sınav sırasında ne kadar süre kaldığımı her istediğimde bakıp öğrenmek isterim. Bu sınavda bütün soruları hızla yanıtlayabildim. Sınav

esnasında bir soruyu yanıtlamadan sürenin bitmesi durumunda, diğer soruyu çözerken stres düzeyim yükselebilirdi.”(Nötr) (T3-KD33)

Test-4

- “Zaman gösterilmediği için zaman yönetimine dair herhangi bir kaygım olmadı. O yüzden stres yapmadım.”(Pozitif) (T4-KD03)
- “Süre strese kaynak veren bir etmen. Süre olmaması daha dikkatli çözmemi sağladı”(Pozitif) (T4-KD29)
- “Sıfıra yakın bir stresim vardı. Bu sürenin farkında olmamak ya da sürenin olmamasından kaynaklı oldu tabi ki. Zaman yönetimi ile ilgili bir planım olmadı çünkü herhangi bir sürenin olduğunun farkında değildim.”(Nötr) (T4-KD07)
- “Test sırasında sürenin gösterilmemesi zaman yönetiminde aksaklık yarattı ve stres düzeyimi olabildiğince arttırdı.”(Negatif) (T4-KD24)

Şeklinde dir. Bireylere sorulan “Soruları çözdüğünüz ekranda sana kalan süre gösterildi/gösterilmedi. Soruları çözerken bunun dışında başka bir bilgi görmek ister miydin?” sorusuna ilişkin bulgulara Tablo 4.9’da yer verilmiştir.

Tablo 4.9. BOBUT Kullanıcı Deneyimi Soru Ekranı Öğelerine İlişkin Ki Kare

		Hayır	Süre	Puan	Toplam
1.test	n	16	4	7	27
	%	59.3	14.8	25.9	100.0
2.test	n	24	6	12	42
	%	57.1	14.3	28.6	100.0
3.test	n	22	7	4	33
	%	66.7	21.2	12.1	100.0
4.test	n	13	13	3	29
	%	44.8	44.8	10.3	100.0
Toplam	n	75	30	26	131
	%	57.3	22.9	19.8	100.0

$$\chi^2 = 13.99; sd = 6; p(0.03) < 0.05$$

Tablo incelendiğinde bireylerin farklı koşullara sahip olan testler ile bireylerin aldıkları testte süre dışında görmek istediği değişkenlere ilişkin görüşleri incelendiğinde istatistiksel olarak anlamlı bir fark vardır ($\chi^2(6) = 13.99 ; p < 0.05$). Tablodaki yüzdeliklere

göre geniş geribildirime sahip olan ve sürenin gösterildiği test olan test-1'i alan bireylerin %59.3'ü başka bir şey görmek istemediklerini söylerken, %14,8 sürenin gösterilmesi gerektiğini söylemiş ve %25.9'u puanlarını görmek istemiştir. Dar geribildirime sahip olan ve sürenin gösterildiği test olan test-2'yi alan bireylerin %57.1'i başka bir şey görmek istemediklerini söylerken, %14.3'ü sürenin gösterilmesini ve %28.6'sı puanlarının gösterilmesi istemiştir. Geniş geribildirime sahip olan ve sürenin gösterilmediği test-3'ü alan bireylerin %66.7'si başka bir şey görmek istemediklerini söylerken, %21.2'si süreyi görmek istediklerini ve %12.1'si puanlarını görmek istediklerini söylemiştir. Dar geribildirime sahip olan ve sürenin gösterilmediği test-4'ü alan bireylerin %44.8'i başka bir şey görmek istemediklerini söylemiştir. %44.8'i süreyi görmek istediklerini ve %10.3'ü puanlarını görmek istediklerini söylemiştir. Ayrıca, "Soruları çözdüğünüz ekranda sana kalan süre gösterildi/gösterilmedi. Soruları çözerken bunun dışında başka bir bilgi görmek ister miydin?" sorusuna ilişkin bireylerin görüşleri doğrudan alıntılarla desteklenmiştir.

Test-1

- "Hayır gerek olmadığını düşünüyorum. Çünkü ekranda soru dışı fazla bilgi dikkat ve odaklanma sorunlarına yol açabilir."(Hayır) (T1-KD08)
- "Hangi soruya ne kadar süre harcadığımı görmek isterdim"(Süre) (T1-KD16)
- "Cevapların doğru ve yanlışlığını da görmek isterdim. Doğru yaptığımı gördüğümde motivasyonumun artacağını düşünmekteyim."(Puan) (T1-KD18)

Test-2

- "Soru sayısı görünüyordu, süre görünüyordu başka ek bir bilgiye ihtiyacım olmadı benim."(Hayır) (T2-KD42)
- "Sorulara geri dönüp dönememe hakkımın olup olmadığı görmek isterdim. Ya da güzel gidiyorsun gibi motivasyon sözleri olsun isterdim"(Temalara dahil edilmeyen cevap.) (T2-KD18)
- "Testin sonunda veya soru sonrası cevapları görmek isterim. Toplam harcanılan süre de gösterebilir."(Süre) (T2-KD12)
- "Geri dönüş veya kontrol butonu görmek isterdim. Sorulara tekrar bakamamak ileriki sorularda geçmiş soruları düşünmeme ve dikkatimin dağılmasına sebep oldu." (Temalara dahil edilmeyen cevap.) (T2-KD35)

Test-3

- “Sadece soruların gözükməsi dikkat çekici ya da dikkat dağıtıcı herhangi bir unsurun bulunmuyor oluşu beni tatmin etti”(Hayır). (T3-KD12)
- “Soru için belirlenen bir süre varsa, onu görmeyi isterdim. Ayrıca önceki sorulara verdiğim yanıtları da görmek isterdim. Genellikle bütün testin seçenek dağılımında dengeyi gözetirim.”(Süre) (T3-KD33)
- “Not vermek kaygısıyla yapılan bir test ise hayır. Ancak bilgileri hatırlamak için yapılan bir test ise verilen yanlış cevap sonrası doğrularını görmek isterdim”(Puan) (T3-KD28)
- “İpucu veyahut konuyla alakalı bir iki tüyo güzel olabilirdi.” (Temalara dahil edilmeyen cevap.) (T3-KD10)

Test-4

- “Hayır istemezdim dikkat dağıtıcı bir şey olsaydı bu yapılan testin geçerliliğini düşürebilirdi.”(Hayır) (T4-KD13)
- “Kaçınıcı soruda olduğum, çözmem gereken daha kaç sorunun olduğu ve sınavın bitimine ne kadar sürenin kaldığı bilgilerini görmek isterdim.”(Süre) (T4-KD24)
- “Doğru veya yanlış yanıtlamadığımı bilmek isterdim. Süreye dair herhangi bir ibare görmedim.”(Puan) (T4-KD04)
- “Soruların görsel olarak Daha iyi olması gerektiğini düşünüyorum. Ve bir ön bilgiden sonra sorulabilirdi.” (Temalara dahil edilmeyen cevap.) (T4-KD19)

Şeklindedir. Bireylere sorulan “Test aldığınız sistemle ilgili önerileriniz var mı? Varsa nedir?” sorusuna ilişkin bulgulara Tablo 4.10’da yer verilmiştir.

Tablo 4.10. BOBUT Kullanıcı Deneyimi Sistem Önerilerine İlişkin Ki Kare

		Yok	Ara yüz	İçerik	Toplam
1.test	n	16	7	4	27
	%	59.3	25.9	14.8	100.0
2.test	n	23	6	13	42
	%	54.8	14.3	31.0	100.0
3.test	n	18	8	7	33
	%	54.5	24.2	21.2	100.0
4.test	n	17	6	6	29
	%	58.6	20.7	20.7	100.0
Toplam	n	74	27	30	131
	%	56.5	20.6	22.9	100.0

$$\chi^2 = 3.56; sd = 6; p(0.73) > 0.05$$

Tablo incelendiğinde bireylerin farklı koşullara sahip olan testler ile bireylerin testi aldıkları sisteme ilişkin görüşleri incelendiğinde istatistiksel olarak anlamlı bir fark yoktur ($\chi^2(6) = 3.56 ; p < 0.05$). Grubun geneli incelendiğinde katılımcıların %56.5'inin bir önerisi olmadığını, %20.6'sının test ara yüzüne ilişkin önerileri olduğu ve %22.9'unun içeriğe yönelik önerileri olduğu gözlemlenmiştir. Ayrıca, "Test aldığınız sistemle ilgili önerileriniz var mı? Varsa nedir?" sorusuna ilişkin bireylerin görüşleri doğrudan alıntılarla desteklenmiştir.

Test-1

- "Herhangi bir önerim yok. Özellikle çevrimiçi bir test olması çözümü kolaylaştırdı."(Yok) (T1-KD26)
- "Aslında ideal bir sistem var. Ekstra fikir belirtmek gerekirse şu öneri olabilir şık işaretlendikten sonra sistem direk diğer soruya atabilir."(Ara yüz) (T1-KD08)
- "Ek bilgiler olabilirdi. Hatırlatıcı cümleler kullanılabilir"(İçerik) (T1-KD05)

Test-2

- "Sistem rahat ve kullanışlıydı. Herhangi bir sorun yaşamadım."(Yok) (T2-KD20)
- "Sistem cep telefonlarıyla entegre çok rahat çözüm yapılabilir. Sorunsuz ilerlendi. Tablo soruları biraz daha resim boyutunda değil standart Tablo ile iletirse görsel anlamda kolaylık sağlayacaktır."(Ara yüz) (T2-KD12)
- "Bazı sorularda ipucu gerekiyordu çünkü bazı bilgileri unutmuşuz"(İçerik) (T2-KD33)

Test-3

- “Daha çok çalışmam gerekiyor diye düşünüyorum. Genel olarak açık ama içinde dallar oluşturan bir alan.”(Yok) (T3-KD30)
- “Sorular çok sıkışık ve soluk görünüyordu. Okuması zordu. Telefonda katılmamın etkisi olup olmayacağını düşündürdü. Bu formdaki sorular gibi net ve okunur olursa daha kolay olur.”(Ara yüz) (T3-KD14)
- “Ekranın bir köşesinde zaman, diğer bir kısmında ise bütün soruları içeren cevap kağıdı görünümü eklenmesini öneririm.”(İçerik) (T3-KD33)
- “Cümle yapısı ve açıklığı arttırabilir ve soruların için zorluğuyla eşdeğer olarak ipuçları yerleştirilebilir.”(İçerik) (T3-KD18)

Test-4

- “Sistem gayet akıcıydı, herhangi bir önerim bulunmamakta.”(Yok) (T4-KD10)
- “Sistemsal bir sorun yoktu fakat soruların ve cevapların puntoları ve yazı tipleri sistemsal bir sorun yüzünden ise düzeltilmeli. Çok rahatsız ediciydi”(Ara yüz) (T4-KD27)
- “Süre arka planda işliyorsa bunu biz de görmeliyiz. Online olması daha pratik yapmamızı sağladı”(İçerik) (T4-KD29)

Şeklinde. Bireylere sorulan “Geribildirim sayfasından başka neler görmek isterdiniz?” sorusuna ilişkin görüşlere Tablo 4.11’de yer verilmiştir.

Tablo 4.11. BOBUT Kullanıcı Deneyimi Geribildirim Ekranına İlişkin Ki Kare

		Yeterli	Ek Açıklama	Puanlama	Toplam
1.test	n	17	4	6	27
	%	63.0	14.8	22.2	100.0
2.test	n	11	15	16	42
	%	26.2	35.7	38.1	100.0
3.test	n	11	11	11	33
	%	33.3	33.3	33.3	100.0
4.test	n	7	13	9	29
	%	24.1	44.8	31.0	100.0
Toplam	n	46	43	42	131
	%	35.1	32.8	32.1	100.0

$$\chi^2 = 13.30; sd = 6; p(0.03) < 0.05$$

Tablo incelendiğinde bireylerin farklı koşullara sahip olan testler ile geribildirim sayfasında başka görmek istediği bilgilere ilişkin görüşleri incelendiğinde istatistiksel olarak anlamlı bir fark vardır ($\chi^2(6) = 13.30 ; p < 0.05$). Tablodaki yüzdeliklere göre geniş geribildirime sahip olan ve sürenin gösterildiği test olan test-1'i alan bireylerin %63.0'ı geribildirim yeterli olduğunu, %14.8'i ek açıklamalar beklediğini ve %22.2'si puanlamaya ilişkin bilgiler görmek istediğini söylemiştir. Dar geribildirime sahip olan ve sürenin gösterildiği test olan test-2'yi alan bireylerin %26.2'si geribildirim sayfasının yeterli olduğunu, %35.7'si ek açıklamalar beklediğini ve %38.1'i puanlamaya ilişkin bilgiler görmek istediğini söylemiştir. Geniş geribildirime sahip olan ve sürenin gösterilmediği test-3'ü alan bireylerin %33.3'ü geribildirim yeterli olduğunu, %33.3'ü ek açıklamalar beklediğini ve %33.3'ü puanlama ile ilgili bilgiler görmek istediklerini söylemiştir. Dar geribildirime sahip olan ve sürenin gösterilmediği test-4'ü alan bireylerin %35.1'i geribildirim yeterli olduğunu, %32.8'i ek açıklamalar görmek istediklerini ve %32.1'i puanlama ile ilgili daha çok bilgi görmek istediklerini söylemiştir. Ayrıca, "Geribildirim sayfasından başka neler görmek isterdiniz?" sorusuna ilişkin bireylerin görüşleri doğrudan alıntılarla desteklenmiştir.

Test-1

- "Geri bildirim sayfası oldukça zengin bir içeriğe sahipti. Özellikle sadece yazı değil grafikle bireyin başarı durumunun gösterilmesi hem birey açısından dikkat çekici hem de test sonu değerlendirme yazısını daha ilgili bir şekilde okuma imkanı yarattı."(Yeterli) (T1-KD08)
- "Alan ile ilgili detay verilebilirdi. Alt boyut gibi düşünelim. Ağırlıklı olarak hangi konuda iyiyim hangi konuda değilim?"(Ek Açıklama) (T1-KD24)
- "Doğruları ve yanlışların hepsini görmek isterdim. Detaylı olarak hangileri doğru hangileri yanlış şeklinde."(Puanlama) (T1-KD02)

Test-2

- "Geribildirim sayfasında bulunanlar yeterli."(Yeterli) (T2-KD41)
- "Doğru/Yanlış cevapları görmek isterdim yanlış cevaplarımın doğru cevap açıklaması ilgili konuyu anlamamı sağlayacaktır. Karşılaştırıldığım grubun özelliklerini de görürsem başarı karşılaştırmasını daha net anlayabilirim."(Puanlama) (T2-KD12)
- "Geri bildirim sayfasında soruların içeriğe ders kapsamına uygunluğu hakkında görüş beklerdim"(Ek Açıklama) (T2-KD23)

Test-3

- “Geri bildirim sayfasından başka bir şey görmenin yararlı olacağını düşünmüyorum”(Yeterli) (T3-KD12)
- “Geribildirime ek olarak kaç doğru kaç yanlış yaptığımı sayı olarak görmek isterdim. Yanlış yaptığım sorulara dönüp o soruların hangileri olduğunu öğrenmek isterdim. Yanlışlarımı merak ettim açıkçası.”(Puanlama) (T3-KD24)
- “Hangi kaynaklardan çalışabileceğim ve hangi alanlarda destek almam gerektiği de yazılsa güzel olurdu”(Ek Açıklama) (T3-KD30)

Test-4

- “Eksik bir şey görmedim, yani sorunlara odaklandığım için eksiklikler görmedim”(Yeterli) (T4-KD23)
- “Ölçme değerlendirme ile ilgili sorular vardır belki yanlış olan soruların kazanımlarına göre analiz yapılabilir. Sonucunda da yönlendirme ile eksik kısımlar/konular çalışılabilir”(Ek Açıklama) (T4-KD25)
- “Daha eğlenceli yapılabilirdi bir emoji animasyon vb eklenebilirdi” (Ek Açıklama) (T4-KD29)
- “Geribildirim de yer alan "Size benzer bir grup" ifadesinin daha fazla nesnel bir açıklamasını görmek isterdim.” (Ek Açıklama) (T4-KD24)
- “Kaç soruyu doğru yaptığımız gösterilebilir. Kaç kişinin katıldığı ve ortalamanın ne olduğu gösterilebilir.”(Puanlama) (T4-KD03)

Şeklindedir. Bireylere sorulan “Aldığımız geribildirim anılandırmak için ders hocasına soru sormaya gerek duydunuz mu? Duyduysanız bu soru neydi?” sorusu için alan uzmanlarınca tema oluşturulamamış. Soru soran bireylerin görüşleri alıntılarla desteklenmiştir.

Test-1

- “Hayır sormadım ancak daha fazla detayı bilmek istedim. Grafikte gösterilen mavi alan hangi gruba tekabül ediyor? Kimden daha iyiyim?” (T1-KD24)
- “Güçlük düzeyi hakkında sormuştum. Test esnasında takıldığım için.” (T1-KD02)
- “Hayır gerek duymadım. Çünkü geri bildirim oldukça detaylı açıklama şeklinde bireye sunulmuştur.” (T1-KD08)

Test-2

- “Soru sorma gereği duymadım. Eğer sorsaydım neden sorulara tekrar bakamadığımızı sorardım.” (T2-KD35)
- “Evet. Geri bildirim hakkında görüş almak istedim.” (T2-KD39)

Test-3

- “Yani Evet aslında geri bildirimde hangi soruyu yanlış hangi soruyu doğru yaptığımı bilemediğim için bunu sorgulamak İstediğim zaman ders hocasına soru sormak isteyebilirim” (T3-KD12)
- “Hocama notumun düşüklüğünü sordum ve ne yapmam gerektiğini. Genel olarak olumlu karşıladı.” (T3-KD30)
- “Hayır soru sormaya gerek duymadım. Gayet açık ve anlaşılırdı. Benim için doğru bir geribildirim olduğunu düşündüm.” (T3-KD14)

Test-4

- “Yanlış yaptığım soruları ve çözümlerini sormak isterdim” (T4-KD12)
- “Yüzdelik olarak verildiği için başarı oranı net değildi” (T4-KD26)
- “Hayır dersin hocasına aldığım geribildirimimi anlamlandırmak için hiçbir soru sorma ihtiyacı hissetmedim çünkü aldığım geribildirim gayet açık ve net bir geribildirimdi.” (T4-KD24)

Öğrencilerin geribildirimini anlamlandırmak için soru sorma isteklerine yönelik olarak sorulmuş olan kullanıcı deneyimi görüş sorusuna bireylerin yaklaşık %7,6’sı (n=10) soru sorduklarını belirtirken yaklaşık %92,3’ü (n=121) soru sormaya gerek duymadığını belirtmiştir. Bu durumun geribildirim veya BOBUT deneyiminin eksikliğinden kaynaklandığı düşünülebilir.

4.7. BOBUT Kullanıcı Deneyimi Anketine İlişkin Bulgular

BOBUT kullanıcı deneyiminin araştırılmasına ilişkin bireylere sorulan anket formundaki sorulara ilişkin bulgular bu bölümde sunulmuştur. Bireylere sorulan “1.Aldığınız teste yönelik genel memnuniyet düzeyinizi puanlayınız.” Sorusuna ilişkin bulgulara Tablo 4.12’de yer verilmiştir.

Tablo 4.12. BOBUT Kullanıcı Deneyimi Memnuniyet Düzeyine İlişkin Kruskal Wallis-H Bulguları

Soru	Test	n	Ortalama Sıra	Kruskall Wallis-H	df	p	Fark	ϵ^2
1.Aldığınız teste yönelik genel memnuniyet düzeyinizi puanlayınız.	Test-1	27	76.67	15.292	3	0.002	3 > 2, 3 > 4	0.097
	Test-2	42	56.29					
	Test-3	33	81.23					
	Test-4	29	52.81					

Tablo incelendiğinde canlı BOBUT uygulamasına yönelik genel memnuniyet düzeyi puanları test gruplarına göre istatistiksel olarak anlamlı bir fark saptanmıştır ($H(3) = 15.292; p \text{ value} = 0.002 < \alpha = 0,05$). Memnuniyet düzeyi puanının test grupları arasındaki farkların incelenmesi adına ortalama sıralar incelendiğinde test-3'ün (81.23) ortalama sırasının test-2'nin (56.29) ortalama sırasından ve test-3'ün (81.23) ortalama sırasının test-4'ün (52.81) ortalama sırasından istatistiksel olarak anlamlı bir şekilde yüksek olduğu saptanmıştır. Etki küçük düzeydir ($\epsilon^2 = 0.097$). Bireylere sorulan “2.Test sonucunda aldığım geribildirim performansı doğru bir değerlendirilmesi olduğunu düşünüyorum.” Sorusuna ilişkin bulgulara Tablo 4.13’de yer verilmiştir.

Tablo 4.13. BOBUT Kullanıcı Deneyimi Geribildirim Doğruluğuna İlişkin Kruskal Wallis-H Bulguları

Soru	Test	n	Ortalama Sıra	Kruskall Wallis-H	df	p	Fark	ϵ^2
2.Test sonucunda aldığım geribildirim performansı doğru bir değerlendirilmesi olduğunu düşünüyorum.	Test-1	27	81.07	10.447	3	0.015	1 > 4	0.058
	Test-2	42	63.12					
	Test-3	33	70.30					
	Test-4	29	51.24					

Tablo incelendiğinde canlı BOBUT uygulamasında alınan geribildirim performansı doğru bir değerlendirmesi olduğuna yönelik görüş puanları test gruplarına göre istatistiksel olarak anlamlı bir fark saptanmıştır ($H(3) = 10.447; p \text{ value} = 0.015 < \alpha = 0.05$). Geribildirim performansına ilişkin görüş puanlarının test grupları arasındaki farkların incelenmesi adına ortalama sıralar incelendiğinde test1'in (81.07) ortalama sırasının test-4'ün (51.24) ortalama sırasından istatistiksel olarak anlamlı bir şekilde yüksek olduğu

saptanmıştır. Etki küçük düzeydir ($\epsilon^2 = 0.058$). Bireylere sorulan “3. Bu geribildirimle pek uyuşmadığımı düşünüyorum ve içeriğini yeterince objektif ve güvenilir bulmadığım için ciddiye almakta zorlanıyorum.” Sorusuna ilişkin bulgulara Tablo 4.14’de yer verilmiştir.

Tablo 4.14. BOBUT Kullanıcı Deneyimi Geribildirim Güvenine İlişkin Kruskal Wallis-H Bulguları

Soru	Test	n	Ortalama Sıra	Kruskall Wallis-H	df	p
3. Bu geribildirimle pek uyuşmadığımı düşünüyorum ve içeriğini yeterince objektif ve güvenilir bulmadığım için ciddiye almakta zorlanıyorum.	Test-1	27	67.65	0.215	3	0.975
	Test-2	42	64.33			
	Test-3	33	65.29			
	Test-4	29	67.69			

Tablo incelendiğinde canlı BOBUT uygulamasında alınan geribildirim katılmama puanlarının test gruplarına göre istatistiksel olarak anlamlı bir fark saptanamamıştır ($H(3) = 0.215; p \text{ value} = 0.975 > \alpha = 0.05$). Bireylere sorulan “4. Aldığım geribildirim, gelecekteki denemelerimdeki motivasyonumu ve gösterdiğim çabayı doğrudan etkileyecek bir faktör olacak.” Sorusuna ilişkin bulgulara Tablo 4.15’te yer verilmiştir.

Tablo 4.15. BOBUT Kullanıcı Deneyimi Geribildirim Çaba ve Motivasyona Etkisine İlişkin Kruskal Wallis-H Bulguları

Soru	Test	n	Ortalama Sıra	Kruskall Wallis-H	df	p	Fark	ϵ^2
4. Aldığım geribildirim, gelecekteki denemelerimdeki motivasyonumu ve gösterdiğim çabayı doğrudan etkileyecek bir faktör olacak.	Test-1	27	81.17	8.106	3	0.044	1 > 4	0.040
	Test-2	42	61.63					
	Test-3	33	68.97					
	Test-4	29	54.83					

Tablo incelendiğinde canlı BOBUT uygulamasında alınan geribildirim motivasyon ve çabaya etkilemede faktör olacağına ilişkin puanlarının test gruplarına göre istatistiksel olarak anlamlı bir fark saptanmıştır ($H(3) = 8.106; p \text{ value} = 0.044 < \alpha = 0.05$). Geribildirim faktörü puanlarının gruplar arasındaki farklarının incelenmesi adına ortalama sıralar incelendiğinde, test-1 (81.17) ortalama sırasının test-4 (54.83) ortalama sırasından istatistiksel olarak anlamlı bir şekilde yüksek olduğu sonucuna ulaşılmıştır. Etki küçük

düzeıdır ($\epsilon^2 = 0.040$). Bireylere sorulan “5.Gelecekte sınavlarımı bireye uyarlanmış olarak almak isterim.” Sorusuna ilişkin bulgulara Tablo 4.16’da yer verilmiştir.

Tablo 4.16. Testleri BOBUT Olarak Alma İsteğine İlişkin Kruskal Wallis-H Bulguları

Soru	Test	n	Ortalama Sıra	Kruskall Wallis-H	df	p
5.Gelecekte sınavlarımı bireye uyarlanmış olarak almak isterim.	Test-1	27	75.89	2.677	3	0.444
	Test-2	42	63.06			
	Test-3	33	64.91			
	Test-4	29	62.29			

Tablo incelendiğinde bireylerin sonraki sınavlarını bireye uyarlanmış şekilde alma isteğine yönelik puanlarının test gruplarına göre istatistiksel olarak anlamlı bir fark saptanamamıştır ($H(3) = 2.677; p \text{ value} = 0.444 > \alpha = 0.05$).

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

5.1. Sonuç ve Tartışma

Bu bölümde BOBUT kullanıcı deneyimini inceleyen nicel ve nitel araştırma bulguları sistematik bir biçimde sunulmaktadır. Testler; kullanıcı deneyiminin memnuniyet, motivasyon ve stres düzeyi gibi boyutları açısından karşılaştırılmıştır. Nicel veriler, betimsel istatistikler ve uygun parametrik olmayan yöntemler aracılığıyla özetlenmiştir. Nitel veriler ise katılımcıların test sonrasında paylaştıkları görüşlerden elde edilen temalarla desteklenmiştir. Çalışmada BOBUT kullanıcı deneyimine etki edebileceği öngörülen değişkenleri içeren dört farklı test deseni hazırlanmıştır. Testler, Eğitimde Ölçme ve Değerlendirme dersi kapsamında oluşturulmuştur. Kullanıcı deneyimini araştırmak amacıyla belirlenen koşullar geribildirim geniş ya da dar kapsamlı olması ve sürenin gösterilip gösterilmemesi şeklinde tasarlanmıştır.

Nicel bulgular, katılımcılara geniş kapsamlı geribildirim verilen ve sürenin gösterilmediği test içeriğinin genel memnuniyet düzeyinde istatistiksel olarak anlamlı biçimde daha yüksek olduğunu ortaya koymuştur. Öte yandan geniş geribildirim verilen ve süresi gösterilen test desenine katılan bireyler, sunulan geribildirim test performanslarını doğru şekilde yansıttığını ifade etmişlerdir. Ayrıca katılımcıların görüşleri incelendiğinde birçok birey geniş geribildirim verilen test desenleri için geribildirim “yeterli düzeyde açıklayıcı” ve “son derece detaylı” olarak tanımlamıştır. Sürenin gösterilip gösterilmemesi değişkenine ilişkin görüşlerde ise katılımcıların bir kısmı sürenin gösterilmesinin stres düzeyini artırdığını diğer kısmı ise zaman yönetimi açısından faydalı olduğunu ve stres kontrolüne katkı sağladığını belirtmiştir.

Alınan geribildirim motivasyon üzerine etkisinin araştırılması amacıyla bireylere yöneltilen açık uçlu sorulara verilen cevaplar incelendiğinde geniş geribildirim alan bireylerde motivasyonun daha yüksek olduğu görülmüştür. Bu durum, literatürde yer alan bazı çalışmalarla da örtüşmektedir. Örneğin De Sixte ve arkadaşları (2020) bireylere verilen geribildirimde motivasyonu artırıcı ifadeler yer verilmesinin, yalnızca içeriğin değil geribildirimin tonu ve biçiminin de motivasyon üzerinde etkili olduğunu ortaya koymuştur. Benzer şekilde Mertens ve Lindner (2025) tarafından yapılan araştırmada geniş kapsamlı geribildirim öğrencilerin duygusal durumlarını ve motivasyonlarını olumlu yönde etkilediği

belirlenmiştir. Bu çalışma kapsamında elde edilen bulguların da literatürle paralellik gösterdiği söylenebilir.

Öte yandan geniş kapsamlı geribildirim almasına rağmen bu geribildirim motivasyon üzerinde olumsuz etki yarattığını belirten ya da geribildirim geliştirilmesi gerektiğini düşünen bazı katılımcılar geribildirim uzunluğundan dolayı tamamını okumadıklarını, teknik terimlerin anlaşılmasını zorlaştırdığını ve hangi soruları doğru ya da yanlış yanıtladıklarını görmek istediklerini belirtmişlerdir. Uygulamalarda testi alan bireylerin ortalama θ düzeylerinin düşük-orta düzeyde olduğu göz önüne alındığında bu durum anlaşılabilir görünmektedir. Nitekim Nash ve Thomas (2024) tarafından yürütülen bir çalışmada da öğrencilerin geribildirimleri düzenli olarak okumamalarının düşük akademik başarı ile ilişkili olduğu bulunmuştur. Ayrıca Handley ve arkadaşları (2011) öğrencilerin geribildirimleri zaman alıcı ya da yorucu bulmalarının bu içerikleri okumama eğilimlerini artırdığını vurgulamışlardır.

Çalışmalarda ulaşılan bir diğer önemli bulgu ise geribildirim çok karmaşık geldiği veya sadece sonuç odaklı verildiği durumlarda öğrenciler tarafından görmezden gelindiğidir. Bu durum özellikle düşük performans gösteren bireylerde daha yaygındır (Jacoby vd., 1984). Hoomanfar ve arkadaşları (2021) ile Wahyudi ve arkadaşları (2024) tarafından yapılan araştırmalar da teknik ve kavramsal düzeyi yüksek olan geribildirimlerin öğrenciler tarafından zor anlaşıldığını ve bu durumun hem bilişsel hem de duygusal açıdan olumsuz etkilere neden olabileceğini ortaya koymuştur. Bu bulgular çalışmada geniş geribildirim almasına rağmen motivasyonu azalan bireylerin görüşleriyle örtüşmektedir. Düşük-orta yetenek düzeyine sahip bireyler geribildirim uzunluğundan sıkıldıklarını, kullanılan teknik ifadelerin motivasyonlarını düşürdüğünü ifade etmişlerdir.

Bireylerin doğru veya yanlış cevapladıkları soruları görme isteği BOBUT uygulamalarında sık karşılaşılan bir beklentidir. Ancak bu talep BOBUT'un en önemli ilkelerinden biri olan madde havuzunun gizliliğini tehdit etmektedir. Araştırmalar, maddelerin kullanım sıklığı dikkatli bir şekilde ayarlansa bile madde havuzunun geçerliliği ve güvenilirliğinin bu durumdan olumsuz etkilenebileceği konusunda ortak görüş bildirmektedir (Stocking & Lewis, 1995). Bu nedenle geribildirim doğru/yanlış şeklinde değil bireyin yetenek düzeyine uyarlanmış ve genel başarı düzeyini özetleyen şekilde verilmesi önerilmektedir (Lilley, 2007; Weiss & Sahin, 2024). Bu bulgular, ayrıca BOBUT

uygulamasına katılan bireylerin sistemin yapısı hakkında önceden bilgilendirilmesi gerektiğini de göstermektedir (Kalender, 2011).

Dar kapsamlı geribildirim alan bireylerin geribildirime yönelik görüşleri incelendiğinde bazı katılımcıların kendilerine benzer grubun neye göre belirlendiğini sorguladığı ya da verilen bilginin çok yüzeysel olduğu yönünde görüşler bildirdikleri görülmektedir. Literatürde geribildirim hiç verilmediği veya yetersiz olduğu durumlarda öğrencilerin motivasyon düzeylerinin etkilenmediği ya da olumsuz yönde etkilendiği gösterilmiştir (Puig vd., 2016). Ayrıca bu tür koşullarda öğrencilerin kaçınma davranışı geliştirebildiği ve öğrenme sürecinden uzaklaşabildiği belirtilmektedir (Gao & Brown, 2023).

Geribildirim kapsamının motivasyon düzeyi üzerindeki etkisini ölçmek üzere sorulan soruya verilen yanıtlar incelendiğinde dar kapsamlı geribildirim alan bireylerin motivasyon düzeylerinin daha düşük olduğu gözlemlenmiştir. Bu bulgu norm referanslı geribildirimlerin özellikle düşük yetenek düzeyine sahip öğrencilerde motivasyonu azalttığını ve duygusal olarak olumsuz etkiler oluşturduğunu bildiren çalışmalarla tutarlıdır (McCloskey & Leary, 1985; Weidlich vd., 2024). Genel olarak değerlendirildiğinde çalışmada geribildirim detaylı ya da dar kapsamlı sunulması durumunda katılımcıların verdikleri görüşlerin literatürdeki bulgularla örtüştüğü söylenebilir.

BOBUT uygulaması sırasında bireylere sürenin gösterilmesinin veya gösterilmemesinin zaman yönetimi ve stres düzeyi üzerindeki etkilerini anlamaya yönelik sorularda, süreyi gören bireylerin yaklaşık %41'i bu durumun kendilerini olumsuz etkilediğini ifade etmiştir. Buna karşın sürenin gösterilmediği grupta bu oran yaklaşık %23'tür. Süreyi gören bireylerin %31'i stres hissetmediklerini belirtirken bu oran sürenin gösterilmediği grupta %28 olarak bulunmuştur. Ayrıca sürenin gösterildiği durumda bireylerin %27'si stres düzeylerinin azaldığını belirtmiş, süre gösterilmediğinde ise bu oran %47'ye çıkmıştır.

Bu bulgular, sınav sürecinde süre bilgisinin bazı bireylerde stresi artırabileceğini ve zaman yönetimini olumsuz etkileyebileceğini ortaya koyan araştırmalarla uyumludur (Coury & Boulette, 1992). Ayrıca zaman yönetimine aşırı odaklanmanın da stres üzerinde olumsuz bir etkisi olabileceği ifade edilmektedir (Imura vd., 2016). Bu çerçevede elde edilen verilerin literatürdeki bulgularla örtüştüğü ve süre değişkeninin bireysel farklılıklar açısından dikkatle ele alınması gerektiği söylenebilir.

Sürenin gösterildiği testlere katılan bireylerin görüşleri incelendiğinde sürenin dijital olarak ekranda yer almasının dikkatlerini dağıttığını ve sınav sürecini yönetme konusunda

strese sebep olduğunu dile getirmişlerdir. Özellikle soru başına verilen süre sayacının ilk birkaç soruda öğrenciler tarafından anlaşılmadığı ve bu durumun süreyi yetiştirememeye endişesi yarattığı belirtilmiştir. Araştırmalar, dijital sayaçların öğrencilerdeki stres düzeyini ve sınav kaygısını %67 oranında artırdığını göstermektedir (Truslow vd., 2022). Bu etki, özellikle testin başlangıcında daha belirgin olmakta ve öğrencinin dikkatinin dağılmasına yol açabilmektedir (Singh vd., 2022). Bu bulgular, sürenin gösterildiği testlerde ilk sorularda bireylerin süre sayacına yönelik aktardıkları görüşlerle tutarlıdır.

Ayrıca sayacın biçimi, rengi ve sayma şeklinin de kaygıyı artırabileceği literatürde yer almaktadır. Çalışmalar, öğrencilere süre göstergesini gizleyebilme tercihi sunulduğunda sınav ortamında kendilerini daha konforlu hissettiklerini ortaya koymuştur (Hott vd., 2023). Sürenin gösterilmediği testlere katılan bireyler, soruları çözerken daha az stres yaşadıklarını ve soruları daha rahat analiz edebildiklerini ifade etmiştir. Ancak bazı öğrenciler, süreyi göremedikleri için zaman yönetiminde zorlandıklarını ve bu durumun da stres yarattığını belirtmiştir. Araştırmalar, sürenin gösterilmesi ya da gösterilmemesi durumunun öğrenci profiline göre değişkenlik gösterebileceğini ortaya koymaktadır. Süre bilgisinin olmaması bazı bireylerde stres düzeyini artırırken, bazılarında ise kaygıyı azaltıcı bir etki yaratmaktadır (Roy vd., 2023). Özellikle zaman yönetimi konusunda eğitim almış bireylerin sürenin gösterilmediği ortamlarda daha az stres yaşadıkları, bu eğitimi almamış bireylerin ise daha fazla stres yaşadığı gözlemlenmiştir (Jenaabadi vd., 2016). Çalışmada ulaşılan bulguların sürenin BOBUT kullanıcı deneyimine etkisi konusunda literatürle paralellik gösterdiği görülmektedir.

BOBUT uygulamasında süre dışında soru ekranında farklı öğelerin gösterilmesine ilişkin öğrenci görüşleri incelendiğinde, sürenin gösterildiği testleri alan bireylerin yaklaşık %58'i sürenin yeterli olduğunu düşünürken; yaklaşık %14'ü soru başına harcanan süre veya toplam süre gibi göstergeleri görmek istediğini ifade etmiştir. Ayrıca öğrencilerin yaklaşık %26'sı verdikleri cevabın doğru ya da yanlış olup olmadığını görmek istediklerini belirtmiştir. Öte yandan sürenin gösterilmediği testlere katılan bireylerin yaklaşık %55'i ekran göstergelerinin yeterli olduğunu düşünürken, yaklaşık %32'si süre bilgisini görmek istediğini ve yaklaşık %11'i de cevaplarının doğruluğuna dair bilgilendirme talep etmiştir.

Süre dışındaki tematik sınıflamaya dahil edilmeyen görüşler ayrıca incelenmiştir. Bu kapsamda bireylerin bazıları sorular arası geçişi sağlayacak navigasyon oklarının olmasını, bazıları ise ipucu verilmesini ve görsel içeriği zengin soruların yer almasını talep etmiştir.

Ancak arařtırmalar öğrencilere cevaplarının doğru/yanlıř olduğunu anlık olarak göstermenin testin psikometrik özelliklerini olumsuz etkileyebileceğini ve bireylerin sınav stratejilerinde deęiřikliklere yol açabileceğini göstermektedir (Van der Linden & Glas, 2010). Ayrıca bu uygulama madde havuzunun deřifre olmasına neden olabilir. Bununla birlikte eđer uygulanan test tanılayıcı bir amaca hizmet ediyorsa ve öğretim süreci ile bütünleřik yürütölüyorsa, anlık geribildirimlerin faydalı olabileceęi bildirilmektedir (Lilley, 2007; Weiss & řahin, 2024).

BOBUT uygulamasının temel mantıęı soruların öğrencilerin performansına yani yetenek düzeyine göre sečilerek bireye özel olarak sunulmasıdır. Bu nedenle öğrencilere sorular arasında gečil imkânı verecek olan navigasyon okları, testin bireye uyarlanmış yapısını bozacaęı için BOBUT uygulamalarında yer almamalıdır (Kalender, 2011; van der Linden, 2000).

Bireylerin sorularda daha fazla görsel materyal görme isteęi son derece olası bir durumdur. Görsel öğelerin öğrencilerin performansında olumlu etki bıraktıęı, çeřitli çalışmalarda da kanıtlanmış bir bulgudur (Sigurdsson & Ring, 2013). Ancak BOBUT uygulamalarının temel mantıęı dikkate alındıęında bireyin karřısına çıkan sorular bireyin yetenek düzeyine göre şekillenmektedir. Bu nedenle canlı BOBUT uygulamalarında bazı bireyler daha çok görsel içerik barındıran sorularla karřılařırken bazıları daha az görsel içerięe sahip ya da hiç görsel içerięe sahip olmayan sorularla karřılařabilir (Doęan & Gündeęer, 2018; Kalender, 2011). Bu bağlamda testi alan her bireyin aynı soru biçimiyle karřılařmaması doęaldır.

Ayrıca BOBUT'un temel felsefesi her bireye aynı soruların deęil ölçme güvenirlięi açısından eř deęer maddelerin sunulmasıdır. Dolayısıyla test adaleti, bireylere aynı içerikte testlerin uygulanmasıyla deęil farklı ama aynı ölçme nitelięine sahip soruların sunulmasıyla saęlanır (van der Linden & Pashley, 2000). Bu yapı bireyselleřtirilmiş ve adil bir deęerlendirme süreci sunmayı amaçlamaktadır.

Öte yandan dijital ortamda hazırlanmış sınavlarda ipucu içeren ara yüzlerin öğrenciler tarafından olumlu karřılandıęı ve özellikle düşük yetenek düzeyindeki bireylerin soruya farklı açılardan yaklařmasına olanak saęladıęı ifade edilmektedir (Beek vd., 2019; Chen & Tseng, 2021). Ancak ipuçlarının doğru cevaba doğrudan yönlendirmesi veya fazla ipucu verilmesi, bireylerin gerçek yetenek düzeyini yansıtmayan yanıtlar vermelerine neden olabilmekte ve bu durum ölçme geçerlilięini zayıflatmaktadır (Weiss & Kingsbury, 1984). Ayrıca yüksek

yetenek düzeyine sahip bireylerin ipuçlarını çoğunlukla dikkate almayarak doğrudan çözüm üretmeye yöneldikleri görülmektedir (Aybek, 2016; Van der Linden & Glas, 2010).

Çalışmadan elde edilen bulgular doğrultusunda öğrencilerin talep ettiği navigasyon oklarının BOBUT'un bireye uyarlanmış yapısına aykırı olduğu, doğru-yanlış geribildirimlerinin madde havuzu güvenliğini zedeleyebileceği ancak tanılayıcı testlerde bu tür geribildirimlerin verilebileceği söylenebilir. Öte yandan ipucu kullanımının testin amacı ve hedef kitlesi doğrultusunda değerlendirilmesi gerektiği ancak ipuçlarının ölçme geçerliliğini zedelemeyecek biçimde tasarlanmasının son derece önemli olduğu sonucuna varılmıştır.

Araştırma kapsamında katılımcılara Concerto platformuna yönelik sistemsel öneriler de sorulmuş ve farklı test türlerini alan öğrenciler arasında belirgin bir görüş farkı saptanmamıştır. Öğrencilerin yaklaşık %56'sı herhangi bir önerisi olmadığını belirtirken, yaklaşık %20'si arayüz ile ilgili, %22'si ise içerikle ilgili önerilerde bulunmuştur. Arayüze yönelik yapılan yorumlar arasında arayüzün soluk olması, yazı puntolarındaki farklılıkların okuma güçlüğü oluşturması ve seçenek işaretlendikten sonra otomatik olarak bir sonraki soruya geçilmesinin tercih edilebileceği gibi ifadeler yer almıştır.

İçerik anlamında öneride bulunan öğrenci görüşleri incelendiğinde sürenin gösterilmediği testleri alan bireylerin süre bilgisini görmek istedikleri ve madde güçlüğüne paralel şekilde ipuçlarının sunulması gerektiği yönünde görüş bildirdikleri görülmüştür. Araştırmalar dijital ortamda geliştirilen testlerde açık ve sade tasarımların öğrenme süreci üzerinde olumlu etkiler yarattığını göstermektedir (Lilley, 2007). Ayrıca BOBUT uygulamalarında algılanan adalet ve kontrol hissinin oluşturulabilmesi için arayüz tutarlılığı, görsel netlik ve etkileşim hızı gibi faktörler son derece önemlidir (Van der Linden & Glas, 2010). Arayüz tasarımlarının uygun şekilde düzenlenmesi kullanıcı deneyimini olumlu yönde etkilemektedir (Madariaga vd., 2021). Öğrenci görüşleri incelendiğinde birkaç birey dışında arayüz fontu ve tasarımının okumada zorlayıcı olduğu ifade edilmiştir. Bu durum bireylerin uygulamaya telefon aracılığıyla katılmış olabileceği ve uygulama öncesinde ekranı yan çevirmeye yönelik yönergeleri okumamış olabileceklerini düşündürmektedir.

Araştırmalar madde güçlüğüne paralel biçimde ipuçlarının verilmesinin biçimlendirici değerlendirmelerde ölçme geçerliliğini zedelemekten motivasyonu artırabildiğini, buna karşılık değer biçme amacı taşıyan değerlendirmelerde ipuçlarının dikkatli bir şekilde kullanılmasının zorunlu olduğunu göstermektedir (Parshall vd., 2000; Wainer vd., 2007). Ayrıca cevabın seçilmesinden sonra bir sonraki soruya doğrudan geçilmesine yönelik öğrenci

görüşleri incelendiğinde bu uygulamanın özellikle basit bilgi düzeyindeki sorularda testin akıcılığını artırabileceği ve test süresinin kısaltılmasına katkı sağlayabileceği ifade edilmektedir (Lilley, 2007). Ancak bu tür otomatik geçişler öğrencilerde kontrol hissini zayıflatabilmekte ve özellikle yüksek yetenek düzeyine sahip bireylerin cevabı verdikten sonra gözden geçirme ihtiyacı duyması durumunda hatalı tıklamaların telafisinin mümkün olmaması riskini doğurmaktadır (Weiss & Şahin, 2024). Araştırmalar ayrıca bu tür katı kontrol yapılarına sahip test uygulamalarının özellikle yüksek önem düzeyine sahip sınavlarda bireylerin bilişsel yükünü ve stres düzeyini artırabileceğini ortaya koymaktadır (Stocking & Swanson, 1993).

Çalışmada testlerin tamamının Concerto platformu aracılığıyla hazırlanmış olması ve farklı testleri alan bireylerin sistemle ilgili benzer görüşler ifade etmesi testler arasında sistemsel bir yanlılık oluşmadığını göstermektedir. BOBUT kullanıcı deneyiminin araştırılmasında kullanılan platformdan kaynaklanan teknik farklılıkların olmaması testlerin eşit koşullarda sunulduğunu ve yapısal geçerliliğin sağlandığını desteklemektedir. Van der Linden & Glas (2010) BOBUT uygulamalarında platformun nötr bir etki taşıması gerektiğini ve bu sayede ölçme sonuçlarının yalnızca bireyin yetenek düzeyine bağlı olacağını ifade etmektedir. Sistemsel olarak kullanıcı deneyiminin düşük olduğu platformlarda yetenekten bağımsız faktörlerin test sürecini şekillendirdiği bulgusu da literatürde yer almaktadır (Weiss & Şahin, 2024).

BOBUT kullanıcı deneyiminin araştırılması kapsamında bireylere geribildirim sayfasında gördüklerinden farklı olarak başka neleri görmek istedikleri de sorulmuştur. Elde edilen bulgulara göre geniş kapsamlı geribildirim alan test grupları ile dar kapsamlı geribildirim alan test grupları arasında istatistiksel olarak anlamlı bir fark saptanmıştır. Geniş kapsamlı geribildirim alan bireylerin yaklaşık %48'i geribildirim sayfasını yeterli bulurken, yaklaşık %23'ü ek açıklamalar görmek istediğini, yaklaşık %27'si ise puanlamaya ilişkin bilgiye yer verilmesini talep etmiştir. Öte yandan dar kapsamlı geribildirim alan bireylerin yaklaşık %34'ü geribildirim sayfasının yeterli olduğunu belirtirken, %39'u ek açıklama, %34'ü ise puanlama hakkında bilgi görmek istediklerini ifade etmiştir.

Geniş kapsamlı geribildirim alan bireylerin ek açıklama isteklerine yönelik ifadeleri incelendiğinde bireylerin hangi konularda daha başarılı olduklarını görmek ve hangi kaynaklardan destek alabilecekleri konusunda bilgi sahibi olmak istedikleri görülmüştür.

Puanlamaya ilişkin birey görüşleri incelendiğinde katılımcıların hangi soruları doğru veya yanlış cevapladıklarını görmek istedikleri ve norma dayalı olarak verilen sonuçlarda ilgili grubun özelliklerine dair bilgi edinmek istedikleri anlaşılmaktadır. Dar kapsamlı geribildirim alan bireylerin görüşlerine bakıldığında puanlamaya ilişkin açıklamalar talep eden katılımcıların doğru veya yanlış cevapladıkları maddeleri görmek istedikleri ve bu maddelere ilişkin açıklamalar bekledikleri görülmektedir. Ayrıca uygulamaya kaç kişinin katıldığı ve grubun ortalama puanının ne olduğu gibi bilgiler de öğrenciler tarafından sıklıkla talep edilmiştir.

Ek açıklama talep eden öğrencilerin görüşleri incelendiğinde soruların kapsama ve kazanımlara uygunluğuna ilişkin bilgi verilmesini ve başarısız oldukları konu ve kavramlara yönelik yönlendirme yapılmasını bekledikleri görülmüştür. Norma dayalı değerlendirmede gruba ilişkin daha fazla açıklama talep eden bireyler aynı zamanda animasyon, emoji ve video gibi dijital geribildirim öğelerinin sunulmasını da istemiştir. Araştırmalar bireylerin dijital ortamlarda aldıkları geribildirimlerin özellikle daha fazla detay içermesini ve bireyselleştirilmiş şekilde verilmesini tercih ettiğini göstermektedir (Ryan vd., 2019). Ayrıca öğrenciler yalnızca genel sonuç bilgilerini değil hangi konularda başarılı ya da başarısız olduklarını ve bu sonuçların hangi öğrenme çıktılarıyla eşleştirilebileceğini gösteren detaylı geribildirimleri daha değerli bulmaktadır (Rubner & Saunders, 2014; Smith vd., 2019).

Lilley (2007) testlerde verilecek geribildirimlerin geniş kapsamlı ve öğrencilerin gelişimine katkı sunacak biçimde yapılandırılması gerektiğini vurgulamaktadır. Özellikle geribildirim biçimlendirici rolü göz önünde bulundurularak hazırlanması ve Bloom (1956) taksonomisine göre bilişsel düzeylere uygun şekilde verilmesi öğrencilerin akademik gelişimlerini planlamaları ve kavramsal eksikliklerini fark etmeleri açısından büyük önem taşımaktadır (Lilley & Barker, 2007).

Araştırmalar bireylerin aldıkları geribildirimlerin norma dayalı olarak verilmesinin öğrenciler tarafından daha anlamlı bulunduğunu göstermektedir (Sedrakyan vd., 2025). Aybek (2016) ve Kuzgun (2006), norma dayalı değerlendirmelerin özellikle yüzdeler puanlar üzerinden verildiğinde bireylerin kendi performanslarını grup bağlamında değerlendirmelerine olanak tanıdığını ve bu tür geribildirimlerin öz-yeterlilik ve mesleki yönelim algılarını etkileyebildiğini belirtmiştir.

Parshall ve arkadaşları (2000) BOBUT uygulamalarında madde ya da konu düzeyinde verilecek geribildirimlerin bireysel öğrenmeyi desteklediğini ancak değer biçmeye yönelik

değerlendirmelerde verilecek geribildirim dikkatli biçimde yapılandırılması gerektiğini ifade etmektedir. Bu bağlamda öğrencilerin öğrenme eksikliklerine yönelik materyallerle desteklenmesi araştırmacıların sıklıkla odaklandığı bir diğer noktadır. Araştırmalar bu tür geribildirimlerin özellikle biçimlendirici değerlendirme amacıyla uygulanan testlerde kullanılması gerektiğini ve bu sayede testin etkililiğinin artabileceğini göstermektedir (Bodea vd., 2013).

Chenyuan ve arkadaşları (2021) tarafından yürütülen çalışmada ise öğrencilerin teste ilişkin yetenek düzeyleri ve öğrenme stilleri dikkate alınarak sunulan geribildirim materyallerinin öğrenme verimliliğini ve motivasyonu artırdığı ifade edilmiştir. Öğrenciler alacakları geribildirimlerde video, animasyon ve yüz ifadesi içeren öğeleri yalnızca yazılı biçimde verilen geribildirimlere kıyasla daha ilgi çekici ve motive edici bulduklarını belirtmişlerdir (Killingback vd., 2020).

Bireylerin doğru veya yanlış cevaplarına ilişkin bilgi alma istekleri testin amacı doğrultusunda farklılık gösterebilmektedir. Ancak genel olarak bu tür bilgilerin paylaşılması madde havuzunun güvenliği açısından sakıncalı bulunmakta ve önerilmemektedir. Bu bağlamda öğrencilerin geribildirim sayfasına ilişkin beklentileri değerlendirildiğinde geniş ve açıklamalı olarak hazırlanan, aynı zamanda görsel materyallerle desteklenen geribildirimlerin öğrencinin dikkatini çekerek etkililiğini artırabileceği düşünülmektedir.

Geribildirimlerin konu bazında paylaşılması öğrencilerin eksik veya yanlış öğrenmelere sahip oldukları konuları belirlemelerine ve öğrenme süreçlerini buna göre planlamalarına katkı sağlayabilir. Ancak bu tür bir uygulama BOBUT'un doğasına uygun olarak son derece dikkatli bir şekilde yapılandırılmalıdır. Konu bazlı geribildirimler sunulabilmesi için çok boyutlu bir madde havuzuna ve her konuya ilişkin yeterli sayıda maddeye sahip olunması gereklidir (Van der Linden & Glas, 2010). BOBUT'un en önemli avantajlarından biri olan daha az sayıda maddeyle bireyin yetenek düzeyinin tahmin edilebilmesi bu noktada bir sınırlılık yaratabilir. Özellikle kısa testlerde her konuya eşit sayıda madde düşmeyeceği için öğrencilerde az sayıda soruya dayanarak yapılan yorumların güvenilirliği hakkında şüpheler oluşabilir.

Ayrıca çok boyutlu madde havuzlarının her bir boyutta ayrı ayrı güvenilirlik analizine tabi tutulması gerektiği literatürde vurgulanmaktadır (Segall, 1996). Bu sorunun çözümü için önerilen yaklaşımlardan biri her boyut için farklı standart hata eşikleri belirlenerek testin durdurulmasıdır. Bu durum kestirimlerin doğruluğunu artırabilir (Weiss & Şahin, 2024).

Veldkamp ve Van der Linden (2002) çalışmalarında her bir boyut için hata varyanslarının Fisher bilgi matrisinden türetilerek hesaplanabileceğini belirtmişlerdir. Böylece yalnızca genel yetenek değil her bir alt boyuta ilişkin güvenilirlik tahminleri de elde edilebilmektedir.

Öğrencilerin katıldıkları test ve sonrasında aldıkları geribildirimlere yönelik soru sorma istekleri incelendiğinde özellikle detaylı geribildirim içeren testlerde grafiklerin nasıl okunacağı, testin zorluk düzeyi ve hangi soruların doğru veya yanlış cevaplandığı gibi konularda sorular yönelttikleri görülmüştür. Dar kapsamlı geribildirim alan öğrenciler ise hangi maddeleri doğru veya yanlış cevapladıklarını göremediklerinden ve test sırasında önceki sorulara dönemediklerinden şikâyet etmişlerdir. Ayrıca norma dayalı sunulan geribildirimi anlamlandırmakta zorluk çektiklerini dile getirmişlerdir.

Araştırmalar öğrencilerin aldıkları geribildirimleri yorumlamada zorlandıkları durumlarda soru sorma eğiliminde olduklarını göstermektedir (Mirador, 2002). Özellikle grafiksel biçimde verilen geribildirimlerin anlaşılması bazı öğrenciler için problem oluşturabilmektedir (Corrin & De Barba, 2015). Norma dayalı değerlendirmelerde ise öğrenciler çıktılarını nasıl değerlendirecekleri konusunda yetersizlik hissedebilmekte ve bu durum da geribildirimle ilgili soru sorma ihtiyacını doğurmaktadır (Doe, 2015).

Öğrencilerin doğru ve yanlış cevaplarına ilişkin bilgi alma talepleri biçimlendirici değerlendirme amacıyla hazırlanan testlerde değerlendirilebilir ancak bu bilgilerin paylaşılması madde havuzunun güvenliğini zedeleyebileceğinden genellikle önerilmemektedir. Ayrıca test esnasında önceki maddelere geri dönüşe izin verilmemesi BOBUT'un bireye uyarlanmış yapısının bir gereğidir. Bu nedenle bu uygulamanın testin doğası gereği olduğu ve öğrencilerin bu konuda bilgilendirilmesinin son derece önemli olduğu sonucuna varılmıştır.

Bireylerin aldıkları geribildirimlerin doğruluğuna ilişkin inançları incelendiğinde geniş kapsamlı geribildirim alan bireylerin dar kapsamlı geribildirim alan bireylere kıyasla geribildirimlerin doğruluğuna daha fazla inandıkları ve bu farkın istatistiksel olarak anlamlı olduğu gözlemlenmiştir. Buna karşın geribildirim içeriğini objektif ve güvenilir bulmama ya da geribildirimi ciddiye almakta zorlanma durumları açısından testler arasında anlamlı bir fark saptanmamıştır.

Araştırmalarda yeterince açıklayıcı olmayan veya yüzeysel biçimde sunulan geribildirimlerin güvenilirliğinin öğrenciler tarafından sorgulandığı bilinmektedir (Davis & Dargusch, 2015; Lee & Cha, 2022). Ayrıca geribildirimlerin etkisi öğrencilerin bu bilgileri ne

kadar deęerli ve gvenilir bulduklarına baęlı olarak řekillenmektedir (Liu & Gumah, 2020). Bireyler geribildirimini olumlu bir řekilde algılamadıklarında ya da duygusal durumlarıyla eřiřtięinde bu bilgilerin faydasını anlamlandıramamakta veya grmezden gelmektedirler (Mayordomo vd., 2022). Bu baęlamda arařtırma bulgularının literatrle rtřtę grlmektedir. Geniř kapsamlı geribildirim alan bireylerin performanslarının doęru deęerlendirildięine inanması geribildirim yeterince detaylı ve aıklayıcı olduęunu gstermektedir. Geribildirimle uyumama durumunda ise test tr fark etmeksizin bireylerin duygusal olarak olumsuz etkilenmedikleri anlařılmaktadır.

Bireylerin gelecekteki sınavlarını BOBUT biiminde alıp almak istemediklerine dair grřleri incelendięinde test trleri arasında istatistiksel olarak anlamlı bir farklılık saptanmamıřtır. Farklı dzeylerdeki test deneyimlerine raęmen bireye uyarlanmış testler ve klasik testler arasında genel memnuniyet aısından fark olmadığı birok alıřmada belirtilmiřtir (Kisiełewska vd., 2024). Guzman ve arkadaşları (2007) bireye uyarlanmış testlerin pedagojik aıdan avantaj saęladığını vurgulasa da ęrenci gruplarının test deneyimlerine ynelik belirgin bir tercih farklılıęı ortaya koymamıřtır. Benzer řekilde bazı arařtırmalar ęrenci performansı ile test tercihi arasında anlamlı bir iliřki olmadığını gstermiřtir (Olsen vd., 1989). Testin bireye uyarlanmış ya da geleneksel biimde uygulanması performans, kaygı ya da motivasyon aısından belirgin farklılıklara yol amamaktadır (Ling vd., 2017).

Bu kapsamda test tr aısından tercihlerde byk bir farklılık gzlenmesi de BOBUT'un bireye uyarlanmış yapısı, pedagojik faydaları, psikometrik doęruluęu ve teknolojik stnlkleri nedeniyle eęitim sistemlerinde daha yaygın hale gelmesi gerektięine dair bilimsel bir ynelim sz konusudur. BOBUT uygulamaları byk veriyle entegre alıřabilme, sınav sonularının gerek zamanlı iřlenebilmesi ve bireyselleřtirilmiş ęrenmeyi desteklemesi gibi avantajlarıyla geleneksel test uygulamalarının tesinde eęitimsel faydalar sunmaktadır (Thompson, 2017). Ayrıca test sresinin kısaltılması daha isabetli yetenek kestirimi yapılabilmesi ve anlık geribildirim saęlanabilmesi gibi zellikleri sayesinde BOBUT, geleneksel testlere kıyasla daha verimli bir lme sreci sunmaktadır (Nguyet & Tu, 2023).

Arařtırmalar bireylerin BOBUT uygulamalarına karřı genel olarak olumlu tutumlara sahip olduklarını ancak deneyim eksiklięinin bu tutumların benimsenmesini sınırlandırdığını gstermektedir (Goto vd., 2023). Hicks (1989) tarafından gerekleřtirilen alıřmada TOEFL

testinin bilgisayar tabanlı bir versiyonuna katılan bireylerin %72'si bu tür bir teste ilk kez girmelerine rağmen %59'u bilgisayar tabanlı testleri tercih etmiş ya da nötr kalmıştır. Bu bulgu bilgisayar tabanlı test deneyiminin ilk uygulamalarda nasıl algılandığını ortaya koymaktadır. Diğer yandan Chew (1999) bireylerin BOBUT'a ilişkin deneyim eksikliklerinin performanslarını ve test algılarını doğrudan etkilediğini belirtmektedir. Bu açıdan bireylerin gelecek testlerinde BOBUT'u tercih edebilmeleri için yeterli deneyimi henüz edinmedikleri söylenebilir.

Son olarak bu uygulamaların bilgisayar tabanlı yapısından dolayı sürekli gelişime ve güncellemeye açık olması kâğıt kalem testlerine göre bir başka avantaj olarak öne çıkmaktadır. Gelecekte öğrencilerin BOBUT uygulamasını tam olarak anlamaları ve deneyim kazanmalarıyla birlikte test tercihlerinin BOBUT lehine değişeceği öngörülmektedir.

Araştırma kapsamında ulaşılan bulgular ve literatür ışığında değerlendirildiğinde BOBUT kullanıcı deneyimini olumlu yönde etkileyebilecek bir geribildirim sayfasının öğrencilerin motivasyonunu zedelemeyecek biçimde tasarlanması gerekmektedir. Bu sayfanın başarı durumuna ilişkin açıklayıcı ifadeler içermesi, motive edici cümlelerle desteklenmesi ve kullanılan teknik terimlerin anlaşılabilir biçimde sunulması oldukça önemlidir. Özellikle düşük yetenek düzeyine sahip bireyler için hazırlanan geribildirimlerde motivasyonel ifadelerin kişiselleştirilerek ve dikkat çekici şekilde sunulması kullanıcı deneyimini geliştirebilir. Bu tür düzenlemeler geribildirim sayfalarında daha sık görsel materyal kullanımını da beraberinde getirebilir.

Öğrencilerin BOBUT uygulamaları sırasında en sık dile getirdiği beklentilerden biri olan doğru ve yanlış cevapların test sonunda gösterilmesi konusu testin amacı doğrultusunda farklılık gösterebilir. Biçimlendirici değerlendirme amacı taşıyan testlerde doğru-yanlış bilgilerinin paylaşılması mümkünken genel anlamda bu uygulamanın madde havuzunun gizliliğini riske atacağı için tercih edilmemesi gerektiği düşünülmektedir. Bununla birlikte BOBUT uygulamalarında sade ve anlaşılır bir arayüz kullanımı ile birlikte sistemsel farklılıklardan kaynaklanabilecek yanlışlıkların en aza indirilmesi de ölçmenin geçerliliği ve güvenilirliği açısından büyük önem taşımaktadır.

Testin amacı ve soruların bilişsel düzeyine bağlı olarak seçenek işaretlemeye dayalı otomatik geçiş özelliğinin kullanılıp kullanılmaması araştırmacının tercihlerine göre değişebilir. Benzer şekilde madde güçlüğüne paralel ipucu sunulması da uygulamanın hedeflerine göre farklılık gösterebilecek bir uygulamadır. Bu bağlamda öğrencilerin BOBUT

hakkında daha fazla bilgi sahibi olmaları ve bu sistemle daha fazla deneyim kazanmaları hem test sürecinin hem de alınan geribildirimlerin daha sağlıklı anlamlandırılmasına katkı sağlayacaktır.

BOBUT kullanıcı deneyimini geliştirmeye yönelik önerilerden biri de süre göstergesinin bireysel tercihe bırakılmasıdır. Araştırmalar, sürenin gösterilmesinin veya gizlenmesinin bireyden bireye değişen stres tepkilerine yol açtığını ortaya koymaktadır. Bu nedenle sürenin gösterilip gösterilmeyeceğine öğrencinin karar verebilmesi test ortamındaki stres düzeyini dengelemede etkili olabilir. Bunun yanında bireylere test süresine ilişkin bilgi verilmesi de zaman yönetimi açısından kendilerini daha güvende hissetmelerini sağlayabilir. Zaman yönetimi eğitimlerinin yaygınlaştırılması da bu konuda öğrencilerin sınav sırasında yaşadıkları stresi kontrol edebilmeleri için önemli bir destek unsuru olacaktır.

5.2. Öneriler

5.2.1. Araştırmacılara Öneriler

- BOBUT kullanıcı deneyiminin geliştirilmesine yönelik olarak gelecekte yapılacak araştırmalarda, kullanıcı deneyimini etkileyebilecek farklı değişkenlerin sistematik biçimde incelenmesi önem taşımaktadır. Bu tür değişkenlerin belirlenmesi BOBUT uygulamalarının pedagojik ve psikometrik açıdan daha etkili hale getirilmesine katkı sunacaktır.
- Araştırmanın farklı öğrenci gruplarında ve örneklemelerde tekrarlanması, elde edilen bulguların genellenebilirliğini artıracak ve BOBUT uygulamalarının farklı popülasyonlar üzerindeki etkilerini karşılaştırmalı olarak değerlendirmeye imkân tanıyacaktır. Bu durum, sistemin kültürel ve bireysel değişkenlere göre nasıl algılandığını ortaya koymak açısından da önemli bir katkı sağlayacaktır.
- Çok boyutlu yapılara sahip madde havuzlarının kullanıldığı BOBUT uygulamalarında, öğrencilere konu bazında geribildirim verilmesi kullanıcı deneyimini daha derinlemesine anlamak açısından yararlı olacaktır. Bu tür bir yaklaşım, öğrencilerin hangi konu alanlarında eksiklik yaşadıklarını fark etmelerini kolaylaştırarak öğrenme sürecinin iyileştirilmesine katkı sağlayabilir.
- Geribildirimlerin Bloom'un taksonomisine dayalı olarak yapılandırılması öğrencilerin aldıkları geribildirimi daha rahat eğitim süreçleriyle ilişkilendirmelerine olanak

taniyacaktir. Bu sayede geribildirimler, öğrencilerin bilişsel gelişim düzeylerine uygun şekilde sunulabilir ve öğrenme çıktılarının daha etkili bir biçimde değerlendirilmesi sağlanabilir.

- BOBUT sistemlerinin arayüz tasarımı evrensel kullanıcı deneyimi ilkelerine göre yeniden şekillendirilerek kullanılmalıdır. Bu sayede kullanıcıların teknolojik kabul düzeylerinin artması sistemin işlevsel faydasının daha belirgin hale gelmesi ve öğrencilerin test ortamına yönelik olumlu tutum geliştirmesi mümkün olacaktır. Kullanıcı odaklı tasarım yaklaşımları özellikle dijital ortamlarda geliştirilen ölçme araçlarının başarısı için kritik öneme sahiptir.
- Araştırmacılar BOBUT kullanıcı deneyimini derinlemesine incelemek amacıyla bilişsel tanı modellerini entegre ederek, test sürecinde bireylerin hangi bilgi alanlarında zorlandıklarını belirleyebilir ve buna dayalı tanılayıcı geri bildirimler geliştirebilirler. Bu sayede ölçme süreci yalnızca yetenek tahminiyle sınırlı kalmaz aynı zamanda öğrenme eksikliklerinin belirlenmesine de hizmet edilebilir. Bu bağlamda BOBUT kullanıcı deneyiminin araştırılmasına fayda sağlayacaktır.

5.2.2. Uygulayıcılara Öneriler

- BOBUT uygulamaları biçimlendirici değerlendirme amacıyla gerçekleştiriliyorsa öğrencilere doğru ve yanlış yanıtladıkları sorulara ilişkin geribildirim verilmesi öğrenme süreçlerinin desteklenmesi açısından faydalı olabilir. Bu tür bir geribildirim, öğrencilerin hatalarını fark etmelerini ve eksik öğrenmelerini gidermelerini kolaylaştırabilir. Ancak bu yaklaşımın madde havuzunun güvenliğini tehlikeye atabileceği ve testin geçerliliğini zedeleyebileceği unutulmamalıdır. Bu nedenle uygulayıcıların testin amacı doğrultusunda bu tercihi dikkatle değerlendirmeleri gerekmektedir.
- BOBUT uygulamalarında sunulan geribildirimlerin açık, anlaşılır ve motive edici bir dille yapılandırılması büyük önem taşımaktadır. Özellikle düşük yetenek düzeyine sahip bireylerin geribildirimlerden maksimum fayda sağlayabilmesi için sunulan içeriklerin sade ve yönlendirici olması gerekmektedir. Ayrıca görsel materyaller, güdüleyici ifadeler, animasyonlar ve diğer çoklu ortam öğeleriyle zenginleştirilmiş geribildirimlerin öğrencilerin ilgisini çekerek motivasyonlarını artıracığı öngörülmektedir.

- Test süresinin gösterilmesi ya da gösterilmemesi durumlarının öğrenciler üzerindeki etkileri bireysel farklılıklara bağlı olarak değişiklik göstermektedir. Bazı öğrenciler süre bilgisini görmekten dolayı stres yaşarken bazıları sürenin gösterilmemesini kaygı verici bulabilmektedir. Bu bağlamda sürenin ekranda gösterilip gösterilmeyeceğine dair kontrolün öğrenciye bırakılması, bireysel tercihlere saygı gösterilmesi ve test deneyiminin daha esnek hale getirilmesi açısından yararlı olacaktır. Bu yaklaşım zaman yönetimi konusunda öğrencilerin daha rahat hissetmesini sağlayarak test sürecindeki stres düzeyini azaltabilir.



KAYNAKÇA

- Aera, A. P. A. (1999). *Standards for educational and psychological testing*. American Educational Research Association.
- Ajjawi, R., Bearman, M., Molloy, E., & Noble, C. (2023). The role of feedback in supporting trainees who underperform in clinical environments. *Frontiers in Medicine*, 10, 1121602.
- Akhtar, H., & Kovacs, K. (2024). Measurement precision and user experience with adaptive versus non-adaptive psychometric tests. *Personality and Individual Differences*, 225, 112675.
- Alipour, M., Moghaddam, M. T., Vaidhyathan, K., & Kjærgaard, M. B. (2023). Emoticontrol: emotions-based control of user-interfaces adaptations. *Proceedings of the ACM on Human-Computer Interaction*, 7(EICS), 1-29.
- Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory*. Waveland Press.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Prentice Hall/Pearson Education.
- Armstrong, R. D., Jones, D. H., & Wu, L. (1992). An automated test development of parallel tests from a seed test. *Psychometrika*, 57(2), 271-288.
- Atılğan, H., Kan, A., & Aydın, B. (2017). *Eğitimde ölçme ve değerlendirme*.
- Atkinson, A., Watling, C. J., & Brand, P. L. (2022). Feedback and coaching. *European journal of pediatrics*, 181(2), 441-446.
- Aybek, E. C. (2016). *Kendini değerlendirme envanteri'nin Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT) olarak uygulanabilirliğinin araştırılması*.
- Aybek, E. C., & Çıkrıkçı, R. N. (2018). Kendini değerlendirme envanteri'nin bilgisayar ortamında bireye uyarlanmış test olarak uygulanabilirliği. *Turkish Psychological Counseling and Guidance Journal*, 8(50), 117-141.
- Aybek, E. C., Arıkan, S., & Ertaş, G. (2024). A practical guide to item bank calibration with multiple matrix sampling. *International Journal of Assessment Tools in Education*, 11(4), 647-659.
- Aybek, E.C. (2021). catIRT tools: A "Shiny" application for item response theory calibration and computerized adaptive testing simulation. *Journal of Applied Testing Technology*, 22(1), 12-24.

- Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, 13(2), 111-127.
- Baadte, C., & Schnotz, W. (2014). Feedback effects on performance, motivation and mood: are they moderated by the learner's self-concept?. *Scandinavian Journal of Educational Research*, 58, 570 - 591. <https://doi.org/10.1080/00313831.2013.781059>.
- Babcock, B., & Weiss, D. J. (2012). *Termination criteria in computerized adaptive tests: Do variable-length CAT's provide efficient and effective measurement?* International Association for Computerized Adaptive Testing, 1, 1-18. Unpublished Doctoral Thesis, The University of Texas, Austin.
- Baker, F. B. & Kim S. H. (2004). *Item response theory: Parameter estimation techniques*.
- Baker, F. B. (1998). Item response theory: parameter estimation techniques. *Applied Psychological Measurement*, 22(1), 89-91.
- Baker, F. B. (2001). The basics of item response theory. For full text: <http://ericae.net/irt/baker>.
- Baker, F.B. & Ho-Kim, S.H. (2017). *The basics of Item response theory using R*. Springer International Publishing.
- Barrett, M. D., Jiang, B., & Feagler, B. E. (2022). A smart authoring system for designing, configuring, and deploying adaptive assessments at scale. *International Journal of Artificial Intelligence in Education*, 32(1), 28-47.
- Baykul, Y. (2015). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Pegem Akademi.
- Baysari, M. T., Hardie, R. A., Lake, R., Richardson, L., McCullagh, C., Gardo, A., & Westbrook, J. (2018). Longitudinal study of user experiences of a CPOE system in a pediatric hospital. *International journal of medical informatics*, 109, 5-14.
- Beiser, D., Vu, M., & Gibbons, R. (2016). Test-retest reliability of a computerized adaptive depression screener. *Psychiatric Services*, 67(9), 1039-1041.
- Beneroso, D., & Robinson, J. (2021). A tool for assessing and providing personalised formative feedback at scale within a second in engineering courses. *Education for Chemical Engineers*, 36, 38-45.
- Bergstrom, B. A. (1992). *Ability Measure Equivalence of Computer Adaptive and Pencil and Paper Tests: A Research Synthesis*.

- Betz, N. E., & Weiss, D. J. (1973). *An Empirical Study of Computer-Administered Two-Stage Ability Testing*.
- Birnbaum, A. (1968). *Some latent trait models*. Statistical theories of mental test scores.
- Bloom, B. (1956). *A taxonomy of cognitive objectives*. McKay.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443-459.
- Bodea, C. N., Dascalu, M. I., & Lytras, M. D. (2013). Advanced personalized feedback in e-assessment systems with recommender engine. In *Information Systems, E-learning, and Knowledge Management Research: 4th World Summit on the Knowledge Society, WSKS 2011*, Mykonos, Greece, September 21-23, 2011. Revised Selected Papers 4 (pp. 179-184). Springer Berlin Heidelberg.
- Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.
- Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how?. *CBE—Life Sciences Education*, 15(4), rm4.
- Brown, J. M., & Weiss, D. J. (1977). An adaptive testing strategy for achievement test batteries. *Research Report 77-6*.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological methods & research*, 21(2), 230-258.
- Burgers, C., Eden, A., Van Engelenburg, M., & Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. *Comput. Hum. Behav.*, 48, 94-103. <https://doi.org/10.1016/j.chb.2015.01.038>.
- Büyüköztürk Ş., Çokluk, Ö., & Köklü, N. (2018). *Sosyal Bilimler İçin İstatistik*. Pegem Akademi.
- Büyüköztürk, Ş., Akgün, Ö. E., Demirel, F., Karadeniz, Ş., & Çakmak, E. K. (2015). *Bilimsel araştırma yöntemleri*.
- Camilli, G. (1994). *Methods for identifying biased test items*. Sage.
- Chalmers, R.P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29. doi:10.18637/jss.v048.i06
- Chang, H. H. (2015). Psychometrics behind computerized adaptive testing. *Psychometrika*, 80(1), 1-20.

- Chang, H. H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20(3), 213-229.
- Chang, H. H., & Ying, Z. (1999). A-stratified multistage computerized adaptive testing. *Applied psychological measurement*, 23(3), 211-222.
- Chang, H. H., & Zhang, J. (2002). Hypergeometric family and item overlap rates in computerized adaptive testing. *Psychometrika*, 67(3), 387-398.
- Chang, W., Cheng, J., Allaire, J.J., Xie, Y., & McPherson, J. (2018). shiny: Web application framework for R. R package version 1.2.0. <https://CRAN.R-project.org/package=shiny>
- Chapman, R., Weiss, D. J., & Suen, K. Y (2024). Effect of routing errors on the psychometric properties of multistage tests. In *Research for Practical Issues and Solutions in Computerized Multistage Testing* (pp. 147-166). Routledge.
- Chen, M., Nah, Y. H., Waschl, N., Poon, K., & Chen, P. (2022). Developing and piloting a computerized adaptive test for a culturally appropriate measure of adaptive behavior. *Journal of Psychoeducational Assessment*, 40(2), 238-254.
- Chen, S. Y., & Tseng, Y. F. (2021). The impacts of scaffolding e-assessment English learning: A cognitive style perspective. *Computer Assisted Language Learning*, 34(8), 1105-1127.
- Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika*, 74(4), 619-632.
- Cheng, Y. (2023). Concerto Software for Computerized Adaptive Testing—Free Version. *Measurement: Interdisciplinary Research and Perspectives*, 21(3), 194-202.
- Chenyuan, S., Jinbao, T., Fangyao, G., & Wenwen, F. (2021, September). Learning resource recommendation method based on knowledge graph and neural network. In *Proceedings of the 2021 4th International Conference on Artificial Intelligence and Pattern Recognition* (pp. 599-606).
- Cheville, A. L., Wang, C., Yost, K. J., Teresi, J. A., Ramirez, M., Ocepek-Welikson, K., ... & Weiss, D. J. (2021). Improving the delivery of function-directed care during acute hospitalizations: methods to develop and validate the Functional Assessment in Acute Care Multidimensional Computerized Adaptive Test (FAMCAT). *Archives of Rehabilitation Research and Clinical Translation*, 3(2), 100112.
- Chew, L. C. (1999). *Computerised adaptive testing: An investigation of factors affecting students' test performance*.

- Choi, Y., & McClenen, C. (2020). Development of adaptive formative assessment system using computerized adaptive testing and dynamic bayesian networks. *Applied Sciences*, 10(22), 8196.
- Christensen, K. B., Kreiner, S., & Mesbah, M. (Eds.). (2013). *Rasch models in health*. John Wiley & Sons.
- Clark, R. C., & Mayer, R. E. (2016). *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*. Wiley.
- Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement*. Mayfield Publishing Co.
- Corrin, L., & De Barba, P. (2015, March). How do students interpret feedback delivered via dashboards?. In *Proceedings of the fifth international conference on learning analytics and knowledge* (pp. 430-431).
- Coury, B. G., & Boulette, M. D. (1992). Time stress and the processing of visual displays. *Human factors*, 34(6), 707-725.
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.
- Crocker, L., & Algina, J. (2006b). Introduction to item response theory. İçinde *Introduction to Classical and Modern Test Theory* (s. 339-375). Mason, Ohio Cengage Learning.
- Cronbach, L. J., & Gleser, G. C. (1957). *Psychological tests and personnel decisions*.
- Çelen, Ü., & Aybek, E. C. (2013). Öğrenci başarısının öğretmen yapımı bir testle klasik test kuramı ve madde tepki kuramı yöntemleriyle elde edilen puanlara göre karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*.
- Davey, T. (2011). *A guide to computer adaptive testing systems*. Council of Chief State School Officers.
- Davey, T., & Parshall, C. G. (1995). *New algorithms for item selection and exposure control with computerized adaptive testing*.
- Davis, S. E., & Dargusch, J. (2015). Feedback, iterative processing and academic trust: Teacher education students' perceptions of assessment feedback. *Australian Journal of Teacher Education*, 40(1), 10.

- De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.
- De Sixte, R., Mañá, A., Ávila, V., & Sánchez, E. (2020). Warm elaborated feedback. Exploring its benefits on post-feedback behaviour. *Educational Psychology, 40*(9), 1094-1112.
- DeMars, C. (2010). *Item response theory*. Oxford University Press.
- Demirtaşlı, N. Ç. (t.y.). Test geliştirmede yeni yaklaşımlar: örtük özellikler kuramı-temel özellikleri varsayımları, modeller ve sınırlılıkları. *Ankara University Journal of Faculty of Educational Sciences (JFES), 28*(2), 161-173.
- DePasque, S., & Tricomi, E. (2015). Effects of intrinsic motivation on feedback processing during learning. *NeuroImage, 119*, 175-186.
<https://doi.org/10.1016/j.neuroimage.2015.06.046>.
- Desjardins, C. D., & Bulut, O. (2018). *Handbook of educational measurement and psychometrics using R*. CRC Press.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining "gamification". In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9-15).
- Doe, C. (2015). Student interpretations of diagnostic feedback. *Language Assessment Quarterly, 12*(1), 110-135.
- Ebenbeck, N., Bastian, M., Mühling, A., & Gebhardt, M. (2024). Duration versus accuracy—what matters for computerised adaptive testing in schools?. *Journal of Computer Assisted Learning, 40*(6), 3443-3453.
- Economides, A. A. (2005). *Adaptive feedback evaluation*. University of Macedonia.
- Eggen, T. J. (1999). Item selection in adaptive testing with the sequential probability ratio test. *Applied Psychological Measurement, 23*(3), 249-261.
- Eggen, T. J. (2018, December). Multi-segment computerized adaptive testing for educational testing purposes. In *Frontiers in Education* (Vol. 3, p. 111). Frontiers Media SA.
- Elola, I., & Oskoz, A. (2022). Reexamining feedback on L2 digital writing. *Studies in Second Language Learning and Teaching, 12*(4), 575-595.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists multivariate*.

- Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists*. Psychology Press.
- Eroğlu, M., Eroğlu, M. G., & Kelecioğlu, H. (2015). Bireyselleştirilmiş bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının ölçme kesinliği ve test uzunluğu açısından karşılaştırılması. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, 28(1), 31-52.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions*.
- Franke, T. M., Ho, T., & Christie, C. A. (2012). The chi-square test: Often used and more often misinterpreted. *American journal of evaluation*, 33(3), 448-458.
- Gan, Z., & Wang, X. (2025). Perceived teacher feedback practices, student feedback motivation and engagement in English learning: a survey of Chinese university students. *Applied Linguistics Review*, 16(1), 455-476.
- Gao, X., & Brown, G. T. (2023). The relation of students' conceptions of feedback to motivational beliefs and achievement goals: Comparing Chinese international students to New Zealand domestic students in higher education. *Education Sciences*, 13(11), 1090.
- Gasperson, S. M. (2014). *The effects of item-by-item feedback during a computer adaptive test*. North Carolina State University.
- Geerards, D., Pusic, A., Hoogbergen, M., Van Der Hulst, R., & Sidey-Gibbons, C. (2019). Computerized quality of life assessment: a randomized experiment to determine the impact of individualized feedback on assessment experience. *Journal of medical Internet research*, 21(7), e12212.
- Germanakos, P., Dimitrova, V. G., & Kleanthous, S. (2020). Article Collection on the Human Aspects in Adaptive and Personalized Interactive Environments. *Frontiers in Artificial Intelligence*, 3, 606068.
- Glas, C. A., & van der Linden, W. J. (2001). *Modeling variability in item parameters in item response models*.
- Gonzalez, E., & Rutkowski, L. (2010). Principles of multiple matrix booklet designs and parameter recovery in large-scale assessments. *IEA-ETS Research Institute Monograph*, 3, 125-156.

- Goto, T., Kano, K., & Shiose, T. (2023, July). Students' acceptance on computer-adaptive testing for achievement assessment in Japanese elementary and secondary school. In *Frontiers in Education* (Vol. 8, p. 1107341). Frontiers Media SA.
- Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational measurement*, 21(4), 347-360.
- Guo, W., & Wei, J. (2019). Teacher feedback and students' self-regulated learning in mathematics: A study of Chinese secondary students. *The Asia-Pacific Education Researcher*, 28, 265-275.
- Guzmán, E., Conejo, R., & Pérez-de-la-Cruz, J. L. (2007). Adaptive testing for hierarchical student models. *User Modeling and User-Adapted Interaction*, 17, 119-157.
- Gündeğer, C., & Doğan, N. (2018). Bireyselleştirilmiş bilgisayarlı sınıflama testlerinde madde havuzu özelliklerinin test uzunluğu ve sınıflama doğruluğu üzerindeki etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 33(4), 888-896.
- Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work?--a literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences* (pp. 3025-3034). Ieee.
- Hambleton, R. K. (1994). Item response theory: a broad psychometric framework for measurement advances 1, 2. *Psicothema*, 6(3), 535-556.
- Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media.
- Hambleton, R. K., Swaminathan, H., & Rogers, J. (1991). *Fundamentals of Item Response Theory*. SAGE Publications.
- Han, K. T. (2009). *A gradual maximum information ratio approach to item selection in computerized adaptive testing*. Reston (VA): Graduate Management Admission Council.
- Handley, K., Price, M., & Millar, J. (2011). Beyond 'doing time': Investigating the concept of student engagement with feedback. *Oxford Review of Education*, 37(4), 543-560.
- Hart, D. L., Cook, K. F., Mioduski, J. E., Teal, C. R., & Crane, P. K. (2006). Simulated computerized adaptive test for patients with shoulder impairments was efficient and produced valid measures of function. *Journal of clinical epidemiology*, 59(3), 290-298.

- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77, 112 - 81. <https://doi.org/10.3102/003465430298487>.
- Hattie, J., & Yates, G. C. R. (2014). *Visible Learning and the Science of How We Learn*. Routledge.
- Hess, A. S., & Hess, J. R. (2017). Understanding tests of the association of categorical variables: the Pearson chi-square test and Fisher's exact test. *Transfusion*, 57(4), 877-879.
- Ho, T. (2010). *A Comparison of Item Selection Procedures Using Different Ability Estimation Methods in Computerized Adaptive Testing Based on Generalized Partial Credit Model*. (Doctoral Dissertation, Graduate School of the University of Texas at Austin).
- Hoomanfar, M. H., Jafarigohar, M., & Jalilifar, A. (2021). Hindrances to L2 graduate students' incorporation of written feedback into their academic texts. *ZABANPAZHUI (Journal of Language Research)*, 12(37), 219-246.
- Hott, J. R., Basit, N., Gao, Z., Truslow, E., & Goulmamine, N. (2023, March). Providing a Choice of Time Trackers on Online Assessments. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (pp. 715-721).
- Imura, T., Takamura, M., Okazaki, Y., & Tokunaga, S. (2016). Construction of the time management scale and examination of the influence of time management on psychological stress response. *Shinrigaku Kenkyu: The Japanese Journal of Psychology*, 87(4), 374–383.
- Jacoby, J., Mazursky, D., Troutman, T., & Kuss, A. (1984). When feedback is ignored: Disutility of outcome feedback. *Journal of Applied Psychology*, 69(3), 531.
- Jansen, T., Höft, L., Bahr, J. L., Kuklick, L., & Meyer, J. (2025). Constructive feedback can function as a reward: Students' emotional profiles in reaction to feedback perception mediate associations with task interest. *Learning and Instruction*, 95, 102030.
- Jenaabadi, H., Nastiezaie, N., & Jalalzaei, S. (2016). The effect of time management training on student's test anxiety. *Journal of nursing education*, 5(1), 12-22.
- Jones, D., Nelson, J., & Opitz, B. (2021). Increased anxiety is associated with better learning from negative feedback. *Psychology Learning & Teaching*, 20, 76 - 90. <https://doi.org/10.1177/1475725720965761>.
- Kalender, İ. (2011). *Effects of different computerized adaptive testing strategies on recovery of ability* (Doctoral dissertation, Middle East Technical University).

- Kasturi, S., Szymonifka, J., Burket, J. C., Berman, J. R., Kirou, K. A., Levine, A. B., ... & Mandl, L. A. (2017). Validity and reliability of patient reported outcomes measurement information system computerized adaptive tests in systemic lupus erythematosus. *The Journal of rheumatology*, 44(7), 1024-1031.
- Kay, R. H., LeSage, A., & Knaack, L. (2019). Assessing learning, engagement, and satisfaction in online learning environments. *Canadian Journal of Learning and Technology*.
- Keller, M. V., Dresel, M., & Daumiller, M. (2024). Do achievement goals and self-efficacy matter for feedback use?. *Learning and Instruction*, 93, 101948.
- Keskin, B., & Gunay, M. (2021, October). A survey on computerized adaptive testing. In *2021 Innovations in Intelligent Systems and Applications Conference (ASYU)* (pp. 1-6). IEEE.
- Kezer, F., & Koç, N. (2009). *Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması*.
- Killingback, C., Drury, D., Mahato, P., & Williams, J. (2020). Student feedback delivery modes: A qualitative study of student and lecturer views. *Nurse Education Today*, 84, 104237.
- Kimura, T. (2017). The impacts of computer adaptive testing from a variety of perspectives. *Journal of educational evaluation for health professions*, 14.
- Kingsbury, C. G., & Zara, A. R. (1991). A comparison of procedures for content-sensitive item selection in computerized adaptive tests. *Applied measurement in education*, 4(3), 241-261.
- Kingsbury, G. G., & Houser, R. L. (1993). Assessing the utility of item response models: Computerized adaptive testing. *Educational Measurement: Issues and Practice*, 12(1), 21-27.
- Kingsbury, G. G., & Weiss, D. J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In *New horizons in testing* (pp. 257-283). Academic Press.
- Kisiełewska, J., Millin, P., Rice, N., Pego, J. M., Burr, S., Nowakowski, M., & Gale, T. (2024). Medical students' perceptions of a novel international adaptive progress test. *Education and Information Technologies*, 29(9), 11323-11338.

- Kline, T. J. (2005). *Psychological testing: A practical approach to design and evaluation*. Sage publications.
- Koğar, H. (2021). *R ile geçerlik ve güvenirlik analizleri*. PEGEM Akademi.
- Kolen, M. J. (2004). *Test equating, scaling, and linking*.
- Kozma, R. (2000). Reflections on the state of educational technology research and development. *Educational technology research and development*, 48(1), 5-15.
- Kuklick, L., & Lindner, M. A. (2021). Computer-based knowledge of results feedback in different delivery modes: Effects on performance, motivation, and achievement emotions. *Contemporary Educational Psychology*, 67, 102001.
- Kuzgun, Y. (2006). *Meslek gelişimi ve danışmanlığı*. Nobel.
- Kwon, S., Kim, S., Lee, S., Kim, J. Y., An, S., & Kim, K. (2023, October). Addressing Selection Bias in Computerized Adaptive Testing: A User-Wise Aggregate Influence Function Approach. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (pp. 4674-4680).
- Landahl, J. (2024). 'The punched cards were sent yesterday, we hope they arrive undamaged.'Computers and international large-scale assessments during the 1960s and 1970s. *Learning, Media and Technology*, 49(1), 93-108.
- Larkin, K. C., & Weiss, D. J. (1975). *An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing*.
- Latu, E., & Chapman, E. (2002). Computerised adaptive testing. *British Journal Of Educational Technology*, 33(5).
- Lee, H. W., & Cha, Y. M. (2022). The effects of digital written feedback on paper-based tests for college students. *The Asia-Pacific Education Researcher*, 31(4), 489-497.
- Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of second language writing*, 17(2), 69-85.
- Legg, S. M., & Buhr, D. C. (1992). Computerized adaptive testing with different groups. *Educational Measurement: Issues and Practice*, 11(2), 23-27.
- Lewis, C., & Sheehan, K. (1990). Using Bayesian decision theory to design a computerized mastery test. *ETS Research Report Series*, 1990(2), i-48.
- Lilley, M. (2007). *The development and application of computer-adaptive testing in a higher education environment* (Doctoral dissertation).

- Lilley, M., & Barker, T. (2007). Students' perceived usefulness of formative feedback for a computer-adaptive test. In *Proceedings of the 5th European Conference on elearning: ECEL* (p. 208). Academic Conferences Limited.
- Lilley, M., Barker, T., & Britton, C. (2004). The development and evaluation of a software prototype for computer-adaptive testing. *Computers & Education*, 43(1-2), 109-123.
- Lilley, M., Barker, T., & Britton, C. (2004). *The generation of automated student feedback for a computer-adaptive test*.
- Lin, H. F. (2007). Measuring online learning systems success: Applying the updated DeLone and McLean model. *Cyberpsychology & behavior*, 10(6), 817-820.
- Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch measurement transactions*, 7, 328.
- Linacre, J. M. (2000). *Computer-adaptive testing: A methodology whose time has come*.
- Linden, W. J. (2005). *Linear models for optimal test design*. Springer Science+ Business Media, Incorporated.
- Lindquist, E. F. (1940). *Statistical analysis in educational research*.
- Ling, G., Attali, Y., Finn, B., & Stone, E. A. (2017). Is a computerized adaptive test more motivating than a fixed-item test?. *Applied psychological measurement*, 41(7), 495-511.
- Liu, W., & Gumah, B. (2020). How perceived value of feedback influences its impact on self-efficacy. *Social Behavior and Personality: an international journal*, 48(6), 1-9.
- Lord, F. (1952). *A theory of test scores*. Psychometric monographs.
- Lord, F. M. (1968). Some test theory for tailored testing. *ETS Research Bulletin Series*, 1968(2), i-62.
- Lord, F. M. (2012). *Applications of item response theory to practical testing problems*. Routledge.
- Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. IAP.
- Luecht, R. M., & Nungester, R. J. (1998). Some practical examples of computer-adaptive sequential testing. *Journal of Educational Measurement*, 35(3), 229-249.
- Luo, X. (2020). Automated test assembly with mixed-integer programming: The effects of modeling approaches and solvers. *Journal of Educational Measurement*, 57(4), 547-565.

- Mackintosh-Franklin, C. (2021). An evaluation of formative feedback and its impact on undergraduate student nurse academic achievement. *Nurse Education in Practice*, 50, 102930.
- Madariaga, L., Nussbaum, M., Gutiérrez, I., Barahona, C., & Meneses, A. (2021). Assessment of user experience in video-based learning environments: From design guidelines to final product. *Computers & Education*, 167, 104176.
- Magis, D. & Barrada, J.R. (2017). Computerized adaptive testing with R: Recent updates of the package catR. *Journal of Statistical Software, Code Snippets*, 76(1), 1-19. doi:10.18637/jss.v076.c01
- Magis, D. & Raiche, G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software*, 48(8), 1-31. doi:10.18637/jss.v048.i08
- Makransky, G., & Glas, C. A. (2010). An automatic online calibration design in adaptive testing. *Journal of Applied Testing Technology*, 1-20.
- Makransky, G., Mortensen, E. L., & Glas, C. A. (2013). Improving personality facet scores with multidimensional computer adaptive testing: An illustration with the NEO PI-R. *Assessment*, 20(1), 3-13.
- Mariz Awad, J. S., Hamdi, N., & Abdennadher, S. (2022). *A-Learning: A Computerized Adaptive Learning Expert System*.
- Marković, S., Jovanović, Z., Jovanović, N., Jevremović, A., & Popović, R. (2013). Adaptive distance learning and testing system. *Computer Applications in Engineering Education*, 21(S1), E2-E13.
- Martin, A. J., & Lazendic, G. (2018). Computer-adaptive testing: Implications for students' achievement, motivation, engagement, and subjective test experience. *Journal of educational psychology*, 110(1), 27.
- Marwan, S., Akram, B., Barnes, T., & Price, T. W. (2022). Adaptive immediate feedback for block-based programming: Design and evaluation. *IEEE Transactions on Learning Technologies*, 15(3), 406-420.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.

- Mayordomo, R. M., Espasa, A., Guasch, T., & Martínez-Melo, M. (2022). Perception of online feedback and its impact on cognitive and emotional engagement with feedback. *Education and Information Technologies*, 27(6), 7947-7971.
- McColskey, W., & Leary, M. R. (1985). Differential effects of norm-referenced and self-referenced feedback on performance expectancies, attributions, and motivation. *Contemporary Educational Psychology*, 10(3), 275-284.
- Mertens, U., & Lindner, M. A. (2025). Computer-Based Answer-Until-Correct and Elaborated Feedback: Effects on Affective-Motivational and Performance Outcomes. *Journal of Computer Assisted Learning*, 41(2), e13112.
- Messick, S. (1980). Test validity and the ethics of assessment. *American psychologist*, 35(11), 1012.
- Messick, S. (1989). *Validity. em r. linn (org.), educational measurement*. (13-103). American Council on Education and Macmillan Publishing Company.
- Messick, S., & Linn, R. L. (1989). Educational measurement. *Validity*, 13-103.
- Mirador, J. F. (2002). What's the Story? How Higher Education Students Interpret Assessment Feedback. *The Asia-Pacific Education Researcher*, 11(2), 1-1.
- Miraz, M. H., Ali, M., & Excell, P. S. (2021). Adaptive user interfaces and universal usability through plasticity of user interface design. *Computer Science Review*, 40, 100363.
- Mohan, M., Nunez, C. M., & Kuchenbecker, K. J. (2024). Closing the loop in minimally supervised human–robot interaction: formative and summative feedback. *Scientific Reports*, 14(1), 10564.
- Molloy, E., & Boud, D. (2014). Feedback Models for Learning, Teaching and Performance. 413-424. https://doi.org/10.1007/978-1-4614-3185-5_33.
- Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, 9(3), e3292.
- Nash, R. A., & Thomas, J. M. (2024). Searching students' reflective writing for linguistic correlates of their tendency to ignore instructors' feedback. *Assessment in Education: Principles, Policy & Practice*, 31(1), 75-90.
- Ngadiman, N., Sulaiman, S., Idris, N., Samingan, M. R., & Mohamed, H. (2021). Systematic review on software quality in educational applications. *IEEE Access*, 9, 60187-60200.

- Nguyet, P. T. T., & Tu, T. P. C. (2023, December). Overview of Computerized Adaptive Testing. In *2023 IEEE/ACIS 8th International Conference on Big Data, Cloud Computing, and Data Science (BCD)* (pp. 181-185). IEEE.
- Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.
- Norman, D. A. (2013). *The Design of Everyday Things*. Basic Books.
- O'Neill, T. R., Gregg, J. L., & Peabody, M. R. (2020). Effect of sample size on common item equating using the dichotomous Rasch model. *Applied Measurement in Education*, 33(1), 10-23.
- Oladele, J. I., & Ndlovu, M. (2021). A review of standardised assessment development procedure and algorithms for computer adaptive testing: applications and relevance for Fourth Industrial Revolution. *International Journal of Learning, Teaching and Educational Research*, 20(5), 1-17.
- Oladele, J., & Ndlovu, M. (2024). *Simulated Computer Adaptive Testing Administration with Remote Proctoring for Offsite Assessment in Mathematics*. Authorea Preprints.
- Olsen, J. B., Maynes, D. D., Slawson, D., & Ho, K. (1989). Comparisons of paper-administered, computer-administered and computerized adaptive achievement tests. *Journal of Educational Computing Research*, 5(3), 311-326.
- Oppl, S., Reisinger, F., Eckmaier, A., & Helm, C. (2017). A flexible online platform for computerized adaptive testing. *International Journal of Educational Technology in Higher Education*, 14, 1-21.
- Oswald, F. L., Shaw, A., & Farmer, W. L. (2015). Comparing simple scoring with IRT scoring of personality measures: The Navy Computer Adaptive Personality Scales. *Applied Psychological Measurement*, 39(2), 144-154.
- Paek, I., & Cole, K. (2019). *Using R for item response theory model applications*. Routledge.
- Paek, I., & Young, M. J. (2005). Investigation of student growth recovery in a fixed-item linking procedure with a fixed-person prior distribution for mixed-format test data. *Applied Measurement in Education*, 18(2), 199-215.
- Pan, Y., Livne, O., Wollack, J. A., & Sinharay, S. (2023). Item selection algorithm based on collaborative filtering for item exposure control. *Educational Measurement: Issues and Practice*, 42(4), 6-18.
- Parshall, C. G., Davey, T., & Pashley, P. J. (2000). Innovative item types for computerized testing. *Computerized adaptive testing: Theory and practice*, 129-148.

- Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. Springer.
- Parshall, C. G., Spray, J. A., Kalohn, J. C., Davey, T., Parshall, C. G., Spray, J. A., ... & Davey, T. (2002). *Considerations in Computer-Based Testing* (pp. 1-12). Springer.
- Patsula, L. N. (1999). *A comparison of computerized adaptive testing and multistage testing*. University of Massachusetts Amherst.
- Popham, W. J. (1990). *Modern educational measurement: A practitioner's perspective*. (No Title).
- Postigo, Á., Cuesta, M., Pedrosa, I., Muñiz, J., & García-Cueto, E. (2020). Development of a computerized adaptive test to assess entrepreneurial personality. *Psicologia: Reflexão e Crítica*, 33, 6.
- Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *Assessment & Evaluation in Higher Education*, 33(2), 143-154.
- Puig Terrón, M., Balagué i Canadell, J., & Solé Pla, J. (2016). The role of feedback in the bologna process, according to goal setting theory: an exploratory study of students' perceptions. *International Journal of Business, Humanities and Technology*, 6(4), 35-39.
- R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ramgovind, P., & Pramjeeth, S. (2023). Students' perceptions of Computerised Adaptive Testing in higher education. *The Independent Journal of Teaching and Learning*, 18(2), 109-126.
- Reckase, M. D., & Reckase, M. D. (2009). Historical background for multidimensional item response theory (MIRT). *Multidimensional item response theory*, 57-77.
- Reise, S. P., & Revicki, D. A. (2014). *Handbook of item response theory modeling*. Taylor & Francis.
- Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item response theory: Fundamentals, applications, and promise in psychological research. *Current directions in psychological science*, 14(2), 95-101.
- Rogers, E. M., Singhal, A., & Quinlan, M. M. (2014). Diffusion of innovations. In *An integrated approach to communication theory and research* (pp. 432-448). Routledge.
- Roid, G. H., & Haladyna, T. M. (1982). A technology for test-item writing. (No Title).

- Rost, T., Stein, J., Löbner, M., Kersting, A., Luck-Sikorski, C., & Riedel-Heller, S. G. (2017). User acceptance of computerized cognitive behavioral therapy for depression: systematic review. *Journal of medical Internet research*, 19(9), e309.
- Roy, T., Saroka, K. S., Hossack, V. L., & Dotta, B. T. (2023). The Effects of Exam-Induced Stress on EEG Profiles and Memory Scores. *Behavioral Sciences*, 13(5), 373.
- Rubner, G., & Saunders, F. (2014). *Student perspectives on formative feedback: An exploratory comparative study*. University of Manchester
- Ryan, T., Henderson, M., & Phillips, M. (2019). Feedback modes matter: Comparing student perceptions of digital and non-digital feedback modes in higher education. *British Journal of Educational Technology*, 50(3), 1507-1523.
- Sanchez-Santillan, M., Fernandez-Medina, C., Perez-Perez, J. R., & Paule-Ruiz, M. (2023). An empirical evaluation of the formative feedback supported by dashboard in the context of compilation error. *Computer Applications in Engineering Education*, 31(5), 1289-1305.
- Sands, W. A., Waters, B. K., & McBride, J. R. (1997). *Computerized adaptive testing: From inquiry to operation*. American Psychological Association.
- Schoonenboom, J., & Johnson, R. B. (2017). How to construct a mixed methods research design. *Kolner Zeitschrift für Soziologie und Sozialpsychologie*, 69(Suppl 2), 107.
- Scullard, M. G. (2007). *Application of item response theory based computerized adaptive testing to the strong interest inventory*. University of Minnesota.
- Sedrakyan, G., Borsci, S., Abdi, A., van den Berg, S. M., Veldkamp, B. P., & van Hillegersberg, J. (2025). Feedback digitalization preferences in online and hybrid classroom: Experiences from lockdown and implications for post-pandemic education. *Journal of Research in Innovative Teaching & Learning*, 18(1), 56-75.
- Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61(2), 331-354.
- Segall, D. O. (2005). Computerized adaptive testing. *Encyclopedia of social measurement*, 1(429-438), 4.
- Shapiro, E. S., & Gebhardt, S. N. (2012). Comparing computer-adaptive and curriculum-based measurement methods of assessment. *School Psychology Review*, 41(3), 295-305.

- Shneiderman, B., Plaisant, C., Cohen, M. S., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2016). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*.
- Siekman, L., Parr, J., Van Ophuysen, S., & Busse, V. (2023). Text quality and changing perceptions of teacher feedback and affective-motivational variables: a study with secondary EFL students. *Frontiers in Education*. <https://doi.org/10.3389/feduc.2023.1171914>.
- Sigurdsson, S. O., & Ring, B. M. (2013). Evaluating preference for graphic feedback on correct versus incorrect performance. *Journal of Organizational Behavior Management*, 33(2), 128-136.
- Singh, A., Garg, P., & Singh, H. (2022, November). Effect of timers on the keystroke pattern of the student in a computer based exam. In *2022 IEEE 6th Conference on Information and Communication Technology (CICT)* (pp. 1-6). IEEE.
- Smith, A. C., Ralph, B. C., MacLeod, C. M., & Smilek, D. (2019). Test feedback and learning: Student preferences and perceived influence. *Scholarship of Teaching and Learning in Psychology*, 5(4), 255.
- Soe, N., Chondamrongkul, N., & Temdee, P. (2024, June). A web-based computerized adaptive testing with Rasch analysis for identifying VAK learning styles. In *2024 5th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)* (pp. 1-5). IEEE.
- Stančić, M., Milin, V., & Gašić, A. (2024). Peer feedback in higher education: How students' roles affect providing and receiving feedback. *Studies in Educational Evaluation*, 83, 101408.
- Stewart, T. (2009). *Usability evaluation*.
- Stocking, M. L., & Lewis, C. (1995). A new method of controlling item exposure in computerized adaptive testing. *ETS Research Report Series*, 1995(2), i-29.
- Stocking, M. L., & Lewis, C. (1998). Controlling item exposure conditional on ability in computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, 23(1), 57-75.
- Stocking, M. L., & Swanson, L. (1993). A method for severely constrained item selection in adaptive testing. *Applied Psychological Measurement*, 17(3), 277-292.

- Sun, J. C.-Y., & Hsu, S.-W. (2013). The effect of gamification on students' learning and engagement: a meta-analysis. *Educational Technology Research and Development*.
- Sünbül, Ö. (2020). Madde tepki kuramı (MTK). İçinde *Psikolojide Ölçme ve Ölçek Geliştirme II: Ölçme Araçlarının Psikometrik Nitelikleri ve Ölçme Kuramları* (pp. 122-141). Ankara: Pegem Akademi.
- Svensäter, G., & Rohlin, M. (2023). Assessment model blending formative and summative assessments using the SOLO taxonomy. *European Journal of Dental Education*, 27(1), 149-157.
- Sympson, J. B., & Hetter, R. D. (1985, October). Controlling item-exposure rates in computerized adaptive testing. In *Proceedings of the 27th annual meeting of the Military Testing Association* (pp. 973-977).
- Szajna, B. (1996). Empirical evaluation of the revised technology acceptance model. *Management science*, 42(1), 85-92.
- Şenel, S. (2017). *Bilgisayar ortamında bireye uyarlanmış testlerin görme engelli öğrencilere uygunluğunun incelenmesi*.
- Tekindal, S. (2021). *Nicel, nitel, karma yöntem araştırma desenleri ve istatistik*. Nobel Yayınevi.
- ter Beek, M., Opdenakker, M. C., Spijkerboer, A. W., Brummer, L., Ozinga, H. W., & Strijbos, J. W. (2019). Scaffolding expository history text reading: Effects on adolescents' comprehension, self-regulation, and motivation. *Learning and Individual Differences*, 74, 101749.
- Theunissen, T. J. J. M. (1985). Binary programming and test design. *Psychometrika*, 50(4), 411-420.
- Thissen, D., & Mislevy, R. J. (2000). Testing algorithms. In *Computerized adaptive testing* (pp. 101-133). Routledge.
- Thissen, D., & Wainer, H. (2001). An overview of test scoring. *Test scoring*, 13-32.
- Thompson, G. (2017). Computer adaptive testing, big data and algorithmic approaches to education. *British journal of sociology of education*, 38(6), 827-840.
- Thompson, N. A., & Weiss, D. A. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research, and Evaluation*, 16(1).

- Truslow, E., Goulmamine, N., Hott, J. R., & Basit, N. (2022, March). Analyzing Student Experience of Time Trackers on Assessments. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2* (pp. 1113-1113).
- Turan, S., Üner, S., & Elcin, M. (2011). The impact of standardized patient feedback on student motivational levels. *Balkan Medical Journal*, 28(1).
- Turda, E. S., Ferent, P., Claudia, C., & Sebastian Turda, E. (2021). The impact of teacher's feedback in increasing student's self-efficacy and motivation. *European Proceedings of Social and Behavioural Sciences*, 104.
- Vale, C. D., & Weiss, D. J. (1975). A Study of Computer-Administered Stradaptive Ability Testing. *Research Report 75-4*.
- van der Linden, C., & Glas, A. W. (2010). *Computerized Adaptive Testing: Theory and Practice*.
- van der Linden, W. J. (1998). Bayesian item selection criteria for adaptive testing. *Psychometrika*, 63(2), 201-216.
- van der Linden, W. J. (2016). *Handbook of Item Response Theory*, Volume Two: Statistical Tools. Chapman & Hall/CRC.
- Van der Linden, W. J., & Chang, H. H. (2003). Implementing content constraints in alpha-stratified adaptive testing using a shadow test approach. *Applied Psychological Measurement*, 27(2), 107-120.
- Van der Linden, W. J., & Glas, C. A. (Eds.). (2000). *Computerized adaptive testing: Theory and practice* (Vol. 13). Dordrecht: Kluwer Academic.
- Van der Linden, W. J., & Glas, C. A. (Eds.). (2010). *Elements of adaptive testing* (Vol. 10, pp. 978-0). New York: Springer.
- Van Der Linden, W. J., & Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. In *Handbook of modern item response theory* (pp. 1-28). Springer New York.
- van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of Modern Item Response Theory*. Springer.
- Van der Linden, W. J., & Pashley, P. J. (2009). Item selection and ability estimation in adaptive testing. In *Elements of adaptive testing* (pp. 3-30). New York, NY: Springer New York.

- Van der Linden, W. J., & van der Linden, W. (Eds.). (2016). *Handbook of item response theory* (Vol. 1). CRC press.
- van der Linden, W. J., & Veldkamp, B. P. (2004). Constraining item exposure in computerized adaptive testing with shadow tests. *Journal of Educational and Behavioral Statistics*, 29(3), 273-291.
- van der Linden, W. J., & Boekkooi-Timminga, E. (1989). A maximin model for test design with practical constraints. *Psychometrika*, 54(2), 237-247.
- Van Wijk, E. V., Donkers, J., De Laat, P. C., Meiboom, A. A., Jacobs, B., Ravensloot, J. H., ... & Bremers, A. J. (2024). Computer adaptive vs. non-adaptive medical progress testing: feasibility, test performance, and student experiences. *Perspectives on Medical Education*, 13(1), 406.
- Vanduhe, V. Z., Nat, M., & Hasan, H. F. (2020). Continuance intentions to use gamification for training in higher education: Integrating the technology acceptance model (TAM), social motivation, and task technology fit (TTF). *Ieee Access*, 8, 21473-21484.
- Veerkamp, W. J., & Berger, M. P. (1997). Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics*, 22(2), 203-226.
- Veldkamp, B. P. (2003). Item selection in polytomous CAT. In *New Developments in Psychometrics: Proceedings of the International Meeting of the Psychometric Society IMPS2001*. Osaka, Japan, July 15–19, 2001 (pp. 207-214). Springer Japan.
- Veldkamp, B. P., & van der Linden, W. J. (2000). *Designing item pools for computerized adaptive testing* (pp. 149-162). Springer Netherlands.
- Veldkamp, B. P., & van der Linden, W. J. (2002). Multidimensional adaptive testing with constraints on test content. *Psychometrika*, 67(4), 575-588.
- Vispoel, W. P. (1998). Psychometric characteristics of computer-adaptive and self-adaptive vocabulary tests: the role of answer feedback and test anxiety. *Journal of Educational Measurement*, 35(2), 155-167.
- Wahyudi, A., Sari, F., Pamuji, A., & Dharmawan, Y. Y. (2024). From Critique to Insight: Student Voices on English Writing Feedback. *Voices of English Language Education Society*, 8(1), 282-292.
- Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. Cambridge University Press.

- Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., & Mislevy, R. J. (2000). *Computerized adaptive testing: A primer*. Routledge.
- Wang, C., Weiss, D. J., & Suen, K. Y. (2021). Hypothesis testing methods for multivariate multi-occasion intra-individual change. *Multivariate behavioral research*, 56(3), 459-475.
- Wang, J., Li, Y., Yang, S., Dong, S., & Li, J. (2023). Waiting experience: Optimization of feedback mechanism of voice user interfaces based on time perception. *IEEE Access*, 11, 21241-21251.
- Wang, T., & Vispoel, W. P. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(2), 109-135.
- Wang, Z., Gong, S., Xu, S., & Hu, X. (2019). Elaborated feedback and learning: Examining cognitive and motivational influences. *Comput. Educ.*, 136, 130-140. <https://doi.org/10.1016/J.COMPEDU.2019.04.003>.
- Weichbroth, P. (2020). Usability of mobile applications: a systematic literature study. *Ieee Access*, 8, 55563-55577.
- Weidlich, J., Fink, A., Jivet, I., Yau, J., Giorgashvili, T., Drachsler, H., & Frey, A. (2024). Emotional and motivational effects of automated and personalized formative feedback: The role of reference frames. *Journal of Computer Assisted Learning*, 40(6), 2735-2752.
- Weiss, D. J. (1974). *Strategies of Adaptive Ability Measurement*.
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied psychological measurement*, 6(4), 473-492.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of educational measurement*, 21(4), 361-375.
- Weiss, D. J., & Sahin, A. (2024). *Computerized Adaptive Testing: From Concept to Implementation*. Guilford Publications.
- Wilson, M. (2023). *Constructing measures: An item response modeling approach*. Routledge.
- Wislocki, K. E., Naderi, G., Schueller, S. M., & Zalta, A. K. (2024). Assessing similarities and differences in thematic content across online mental health communities dedicated to trauma-related mental health conditions. *Psychological Trauma: Theory, Research, Practice, and Policy*.

- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.03087>.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA press.
- Wu, H. M., Kuo, B. C., & Wang, S. C. (2017). Computerized dynamic adaptive tests with immediately individualized feedback for primary school mathematics learning. *Journal of Educational Technology & Society*, 20(1), 61-72.
- Xu, L., Jin, R., Huang, F., Zhou, Y., Li, Z., & Zhang, M. (2020). Development of computerized adaptive testing for emotion regulation. *Frontiers in Psychology*, 11, 561358.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of educational measurement*, 30(3), 187-213.
- Zeller, R. A., & Carmines, E. G. (1980). *Measurement in the social sciences: The link between theory and data*. Cambridge University Press.
- Zhu, B., Kaber, D., Zahabi, M., & , J. (2017). Effect of feedback type and modality on human motivation. *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2838-2843. <https://doi.org/10.1109/SMC.2017.8123057>.
- Zhuang, Y., Feng, W., & Liao, Y. (2017). Want More? Learn Less: Motivation Affects Adolescents Learning from Negative Feedback. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00076>.
- Zhuang, Y., Liu, Q., Huang, Z., Li, Z., Jin, B., Bi, H., ... & Wang, S. (2022, July). A robust computerized adaptive testing approach in educational question retrieval. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 416-426).

EK-1. ETİK KURUL VE MEB ARAŞTIRMA İZİN BELGELERİ

Evrak Tarih ve Sayısı: 20.02.2025-1146510



T.C.
AKDENİZ ÜNİVERSİTESİ REKTÖRLÜĞÜ
Sosyal ve Beşeri Bilimler Bilimsel Araştırma ve Yayın Etiği Kurulu
KURUL KARARI



TOPLANTI TARİHİ : 20.02.2025
TOPLANTI SAYISI : 04
KARAR SAYISI : 78

Üniversitemiz Eğitim Fakültesi Eğitim Bilimleri Bölümü öğretim üyesi **Doç. Dr. Alper SİNAN**'ın danışmanlığını, **Alper TOSUN**'un araştırmacılığını üstlendiği, *"BOBUT Uygulamasında Kullanıcı Deneyimine Etki Eden Değişkenlerin ve Olası Etkilerinin Belirlenmesi"* konulu çalışmanın, fikri hukuki ve telif hakları bakımından metot ve ölçeğine ilişkin sorumluluğun başvurucuya ait olmak üzere, proje süresince uygulanmasının etik olarak **uygun olduğuna** oy birliği ile karar verilmiştir.

Prof. Dr. Hilmi DEMİRKAYA
Kurul Başkanı

Başkan
Prof. Dr.
Hilmi DEMİRKAYA

Başkan Yrd.
Prof. Dr.
Sibel PAŞAOĞLU YÖNDEM

Üye
Prof. Dr.
Ebru İÇİGEN

Üye
Prof. Dr.
Nurşen ADAK

Üye
Prof. Dr.
Taner KORKUT

Üye
Prof. Dr.
Gökhan AKYÜZ

Üye
Prof.
Ceren HEPYÜCEL

ÖZ GEÇMİŞ

Kişisel Bilgiler

Adı Soyadı: Alper TOSUN

Eğitim Durumu

Lisans: Pamukkale Üniversitesi, Eğitim Fakültesi, Sosyal Bilgiler Öğretmenliği (2015-2019)

Yüksek Lisans: Akdeniz Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Yüksek Lisans Programı (2019-2021)

Doktora: Akdeniz Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Doktora Programı (2021-)

Bilimsel Faaliyetler

Makaleler

Tosun A.,SİNAN A., KINA, Ö. F. .(2022).Kümeleme Analizi İle Akran Zorbalığı Tutumlarına Göre Ülkelerin Sınıflandırılması. Elektronik Dil ve Eğitim Dergisi,(3)52-60.

Çatal, E. , Sinan, A. , Kına, Ö. F. & Tosun, A. (2022). Yabancı Öğrencilerin Türkiye’de Üniversiteye Giriş Sınavı Başarılarının Karşılaştırılmalı Analizi . Uluslararası Türk Kültür Coğrafyasında Sosyal Bilimler Dergisi , 7 (2) , 240-254

Bildiriler

Tosun A, Sinan A, Bıçak B.(2021). Eğitimde Fırsat Eşitliği Açısından İllerin Yaşam ve Eğitim İndekslerine Göre Kümelenmesi. IV. Uluslararası Eğitim Araştırmaları Kongresi, İzmir.

Kitap Bölümleri

Sinan, A., Tosun, A., & Bıçak, B. (2023). Eğitimdeki fiziksel ve sosyal imkânlar açısından Türkiye'de illerin kümeleme analizi ile incelenmesi. In H. Şahin & A. Bostan Sarıoğlu (Eds.), Eğitim Bilimleri Araştırmaları (1. baskı, ss. 280). Livre De Lyon. ISBN: 978-2-38236-647-9.

Tosun A., Gündoğan K., & Sinan A.(2024). Klasik Test Kuramı ve Madde Tepki Kuramı Yöntemlerinin DMF İçerisinde Karşılaştırılması(Bidge Yayınlarında yayınlanacak-Eğitim Bilimlerinde Veri Analizi Çalışmaları)

Tosun A. & Sinan A.(2024) Bilgisayar Ortamında Bireye Uyarlanan Test Uygulamalarında Kullanıcı Deneyimine Etki Eden Faktörler. Veriden Kavrama: Eğitimde Ölçme ve Değerlendirme. Bidge Yayınları(2024)

Kına Ö., Tosun A. & Sinan A.(2024) İlköğretime Hazırlık Kapsamında Kullanılan Erken Dönem Ölçme Araçlarının İncelenmesi. Erken Çocuklukta Oyun ve Öğrenme: Erken Çocukluk Eğitiminin Temelleri. Bidge Yayınları(2024)

BENZERLİK RAPORU

BOBUT UYGULAMASINDA KULLANICI DENEYİMİNE ETKİ EDEN DEĞİŞKENLERİN VE OLASI ETKİLERİNİN BELİRLENMESİ

ORJİNALLİK RAPORU

%6

BENZERLİK ENDEKSİ

%5

İNTERNET KAYNAKLARI

%4

YAYINLAR

%2

ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1

acikbilim.yok.gov.tr

İnternet Kaynağı

%1

2

dspace.ankara.edu.tr

İnternet Kaynağı

%1

3

acikerisim.akdeniz.edu.tr

İnternet Kaynağı

<%1

4

Submitted to Akdeniz University

Öğrenci Ödevi

<%1

5

acikerisim.akdeniz.edu.tr:8080

İnternet Kaynağı

<%1

6

dspace.akdeniz.edu.tr

İnternet Kaynağı

<%1

7

www.openaccess.hacettepe.edu.tr:8080

İnternet Kaynağı

<%1

8

"The Wiley Handbook of Psychometric Testing", Wiley, 2018

Yayın

<%1

BİLDİRİM

Hazırladığım tezin tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin kâğıt ve elektronik kopyalarının Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü arşivlerinde saklanmasına izin verdiğimi onaylarım.

☐ Tezimin/Raporumun tamamı her yerden erişime açılabilir.

☐ Tezim/Raporum sadece Akdeniz Üniversitesi yerleşkelerinden erişime açılabilir.

☐ Tezimin/Raporumun 6 ay süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

Haziran 2025

Alper TOSUN