



**T.C.
HARRAN ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

**TÜRKİYE'DE BİLGİSAYAR ORTAMINDA BİREYE UYARLANMIŞ TEST
(BOBUT) ÇALIŞMALARI: SİSTEMATİK BİR DERLEME**

**EĞİTİM BİLİMLERİ ANA BİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

Rumeysa EKİZ KUZUCU

Şanlıurfa

2024

T.C.
HARRAN ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

TÜRKİYE'DE BİLGİSAYAR ORTAMINDA BİREYE UYARLANMIŞ TEST
(BOBUT) ÇALIŞMALARI: SİSTEMATİK BİR DERLEME

RUMEYSA EKİZ KUZUCU

EĞİTİM BİLİMLERİ ANA BİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

Tez Danışmanı: Doç. Dr. SEDAT ŞEN

Şanlıurfa
2024

TEŞEKKÜR

Öncelikle, tez danışmanım Doç. Dr. Sedat Şen'e, çalışmama gösterdiği ilgi, değerli bilgi ve tecrübeleriyle bana rehberlik ettiği ve sağladığı destek için sonsuz teşekkürlerimi sunarım. Kendisinin yönlendirmeleri ve teşvikleri olmadan bu tezin ortaya çıkması mümkün olmazdı.

Tez savunma jürimde bulunan Doç. Dr. Derya Evran ve Doç. Dr. Murat Doğan Şahin'e, kıymetli katkıları ve değerlendirmeleri için teşekkür ederim. Ayrıca, Harran Üniversitesi Eğitimde Ölçme ve Değerlendirme Bilim Dalındaki saygıdeğer hocalarıma, bana verdikleri eğitim ve destekleri için minnettarım.

Eşim Köksal Kuzucu'ya, bu zorlu süreçte gösterdiği sabır, anlayış ve sürekli desteği için teşekkür ederim. Hem kızımızla ilgilenerek bana çalışma zamanı yarattığın hem de motivasyonumu yüksek tutmamı sağladığın için sonsuz teşekkürler. Bu çalışmayı, senin manevi desteğin sayesinde tamamlayabildim. Ayrıca, kendimi geliştirmemde bana ilham veren kızım Elife de teşekkür ederim.

Bu süreçte değerli katkıları, yardımları ve önerileriyle bana destek sağlayan sevgili arkadaşım İbrahim Taşkiran'a da teşekkür etmek isterim.

Sevgili anneme, her zaman yanımda olduğu, bana inandığı ve sevgisini esirgemediği için özel teşekkürlerimi sunarım. Annemin sağladığı motivasyon ve destek, bu sürecin üstesinden gelmemde büyük rol oynamıştır.

Son olarak, bu akademik yolculukta bana inanan, destek olan ve cesaret veren herkese en derin şükranlarımı sunarım.

İÇİNDEKİLER

ÖZET	i
ABSTRACT.....	ii
ŞEKİLLER DİZİNİ	iii
ÇİZELGELER DİZİNİ	iv
SİMGELER.....	v
KISALTMALAR	vi
1. GİRİŞ	1
1.1. Problem Durumu.....	1
1.2. Araştırmanın Amacı ve Önemi	3
1.3. Araştırma Problemi	4
1.4. Araştırmanın Kuramsal Temelleri ve İlgili Araştırmalar.....	4
1.4.1. Teknoloji ve Ölçme	4
1.4.2. Kâğıt Kalem Testlerinden Bilgisayara Geçiş	5
1.4.3. Bireye Uyarlanmış Test Fikrinin Ortaya Çıkışı ve BOBUT'un Gelişim Süreci	6
1.4.4. BOBUT'un Tanımı ve Çalışma Prensipleri.....	7
1.4.5. Madde Tepki Kuramı (MTK).....	8
1.4.5.1. Madde Tepki Kuramı (MTK) nedir?.....	9
1.4.5.2. MTK'nin Temel Kavramları ve Varsayımları	10
1.4.5.3. MTK Modelleri	13
1.4.5.4. Bireye Uyarlanmış Testler ve MTK	16
1.4.6. BOBUT Temel Adımları ve Algoritmalar	16
1.4.6.1. Madde Havuzu Oluşturma	17
1.4.6.2. Test Başlatma Kuralı ve İlk Madde Seçimi.....	18
1.4.6.3. Madde Seçim Yöntemleri.....	18
1.4.6.4. Yetenek Kestirim Yöntemleri.....	20
1.4.6.5. Test Sonlandırma Kuralı	21
1.4.7. Farklı BOBUT Türleri.....	23
1.4.7.1. Çok Aşamalı BOBUT (Multistage Testing-ca-MST).....	23
1.4.7.2. Bilişsel Tanıya Dayalı BOBUT (Cognitive Diagnostic Computerized Adaptive Testing - CD-CAT).....	24
1.4.7.3. Çok Boyutlu BOBUT (Multidimensional Computerized Adaptive Testing - MCAT).....	24
1.4.7.4. Sınıflandırmaya Dayalı BOBUT (Computerized Adaptive Classification Testing – CACT).....	24
1.4.8. BOBUT'un Avantajları ve Sınırlılıkları.....	25
1.4.9. BOBUT'un Uygulama Alanları	27
1.4.10. BOBUT Yazılımları	27
2. ÖNCEKİ ÇALIŞMALAR	31
2.1. BOBUT Konulu Yurt İçinde Yapılan Çalışmalar	31
2.2. BOBUT Konulu Yurt Dışında Yapılan Çalışmalar.....	32
2.3. BOBUT Konulu Yapılan Önceki Derlemeler, Bibliyometrik Analizler ve Meta-Analiz Çalışmaları	34
3. GEREÇ VE YÖNTEM	36
3.1. Araştırmanın Deseni	36
3.2. Veri Toplama Süreci	36
3.3. Dâhil Etme ve Hariç Tutma Kriteri.....	40
3.4. Veri Dosyasının Oluşturulması ve Kodlama Güvenirliği	41
3.5. Verilerin Analizi	44

4. BULGULAR.....	46
4.1. Çalışmaların Yıllara Göre Dağılımı.....	53
4.2. Çalışmaların Yayın Türlerine Göre Dağılımı	54
4.3. Çalışmaların Yapılma Amacına Göre Dağılımı	54
4.4. Çalışmaların Yayın Diline Göre Dağılımı	55
4.5. Çalışmaların Örneklem Gruplara Göre Çeşitliliği	56
4.6. Çalışmalarda Kullanılan Farklı BOBUT Türleri.....	56
4.7. Çalışmalarda Canlı BOBUT Uygulaması ve Simülasyon Oranları	57
4.8. Çalışmalarda Kullanılan BOBUT Uygulama, Simülasyon ve Madde Kalibrasyon Yazılımları	58
4.9. Çalışmaların Örneklem Büyüklüklerinin Dağılımı	60
4.10. Çalışmaların Madde Havuzu Büyüklüklerinin Dağılımı	61
4.11. Çalışmalarda Ölçülen Yapıların Çeşitliliği	62
4.12. Çalışmalarda Ölçülen Yapıların Boyut Durumları	63
4.13. Çalışmalarda Kullanılan En Yaygın MTK Modellerinin Dağılımı.....	64
4.14. Çalışmalarda En Yaygın Kullanılan Yöntemler	65
4.14.1. En Yaygın Kullanılan Madde Seçim Yöntemleri	65
4.14.2. En Yaygın Kullanılan Madde Kestirim Yöntemleri	66
4.14.3. En Yaygın Kullanılan Yetenek Kestirim Yöntemleri	66
4.14.4. En Yaygın Kullanılan Test Başlatma Kuralları	67
4.14.5. En Yaygın Kullanılan Test Sonlandırma Kuralları	68
4.15. Çalışmalarda Madde Kullanım Sıklığı ve İçerik Dengeleme Yapılma Oranları	69
4.16. Çalışmaların Genel Sonuçlarına Göre Durumu	69
5. TARTIŞMA	71
6. SONUÇLAR	81
7. ÖNERİLER.....	84
7.1. Gelecekte Yapılacak BOBUT Çalışmalarına Yönelik Öneriler.....	84
7.2. Benzer Konuda Çalışacak Araştırmacılara Yönelik Öneriler	85
KAYNAKLAR.....	87
ÖZGEÇMİŞ	100
EKLER.....	101

ÖZET
YÜKSEK LİSANS TEZİ

**TÜRKİYE'DE BİLGİSAYAR ORTAMINDA BİREYE UYARLANMIŞ TEST
(BOBUT) ÇALIŞMALARI: SİSTEMATİK BİR DERLEME**

RUMEYSA EKİZ KUZUCU

**HARRAN ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

Tez Danışman: Doç. Dr. SEDAT ŞEN
Yıl: 2024, Sayfa : 99

Bu araştırma, Türkiye'de Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT) konulu çalışmaları sistematik bir şekilde derleyerek, bu alandaki mevcut durumu analiz etmeyi amaçlamaktadır. Çalışma, Türkiye'de yapılan BOBUT çalışmalarının hangi yıllarda yoğunlaştığını, hangi tekniklerin kullanıldığını, çalışmaların yayın türlerini, örneklem çeşitliliğini ve diğer metodolojik özelliklerini incelemektedir. Araştırma kapsamında, literatürde yer alan BOBUT çalışmaları belirli kriterlere göre seçilerek analiz edilmiştir. Çalışmalarda en yaygın kullanılan BOBUT türleri, madde seçim ve yetenek kestirim yöntemleri, test başlatma ve sonlandırma kuralları gibi teknik detaylar ele alınmıştır. Ayrıca, BOBUT'un sağladığı avantajlar ve sınırlılıklar da kapsamlı bir şekilde değerlendirilmiştir. Elde edilen bulgular, Türkiye'de BOBUT konulu araştırmaların son yıllarda artış gösterdiğini ve yapılan çalışmaların büyük bir kısmının eğitim alanında yoğunlaştığını ortaya koymaktadır. Çalışmalarda en yaygın kullanılan madde seçim yöntemi olarak Maksimum Fisher Bilgisi (MFI), yetenek kestirim yöntemi olarak Expected A Posteriori (EAP), test başlatma kuralı olarak nötr teta başlangıcı ve test sonlandırma kuralı olarak standart hata eşik değeri yöntemleri öne çıkmaktadır. Ayrıca, çalışmaların çok büyük bir kısmında madde kestirim yöntemi hakkında bilgi verilmemiş, içerik dengeleme ve madde kullanım sıklığı kontrolü çok az çalışmada yapılmıştır. Sonuç olarak, bu derleme çalışması, Türkiye'de BOBUT konulu araştırmaların mevcut durumunu özetlemekte ve bu alandaki bilgi birikimini derinleştirmektedir. Elde edilen veriler, gelecekte yapılacak araştırmalara yön göstermesi ve BOBUT yönteminin daha etkin kullanılmasına katkı sağlaması açısından önemli bir kaynak oluşturmaktadır.

ANAHTAR KELİMELEER: Bilgisayar Ortamında Bireye Uyarlanmış Test, bireyselleştirilmiş ölçme ve değerlendirme, sistematik derleme

ABSTRACT
MASTER’S THESIS

**COMPUTERIZED ADAPTIVE TESTING (CAT) STUDIES IN TURKEY: A
SYSTEMATIC REVIEW**

RUMEYSA EKİZ KUZUCU

**HARRAN UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
MEASUREMENT AND EVALUATION IN EDUCATION PROGRAM**

Thesis Supevisor: Assoc. Prof. Dr. Sedat ŞEN
Year: 2024, Page: 99

This study aims to systematically compile and analyze the current state of research on Computerized Adaptive Testing (CAT) in Turkey. The study examines the years in which CAT research in Turkey has concentrated, the techniques used, the types of publications, sample diversity, and other methodological characteristics. Within the scope of the research, CAT studies in the literature were selected and analyzed based on specific criteria. The most commonly used types of CAT, item selection and ability estimation methods, test initiation and termination rules, and other technical details were examined. Additionally, the advantages and limitations of CAT were comprehensively evaluated. The findings reveal that there has been an increase in CAT research in Turkey in recent years, with the majority of studies concentrated in the field of education. The most commonly used item selection method is Maximum Fisher Information (MFI), the ability estimation method is Maximum Likelihood Estimation (MLE), the test initiation rule is the neutral theta starting point, and the test termination rule is the standard error threshold. Furthermore, a significant portion of the studies lacked information on item calibration methods, and content balancing and item usage frequency control were addressed in very few studies. In conclusion, this review summarizes the current state of CAT research in Turkey, deepening the knowledge base in this field. The data obtained provide valuable insights for future research and contribute to the more effective use of CAT methods.

KEYWORDS: Computerized Adaptive Testing, individualized assessment and evaluation, systematic review

ŞEKİLLER DİZİNİ

Şekil 1.1. İkili Puanlanan Maddelerde BOBUT'un İşleyişi.....	8
Şekil 1.2. Madde Karakteristik Eğrisi.....	11
Şekil 1.3. MTK Modellerinin Sınıflandırılması.....	14
Şekil 1.4. BOBUT Adımları.....	16
Şekil 1.5. Madde Seçim Yöntemleri.	19
Şekil 1.6. Yetenek Kestirim Yöntemleri.....	20
Şekil 1.7. Test Sonlandırma Kuralları.....	21
Şekil 1.8. BOBUT Uygulamalarında Kullanılan Yazılımlar.....	28
Şekil 3.1. PRISMA Akış Süreci.....	38
Şekil 3.2. Kodlama Değişkenleri.....	40
Şekil 4.1. BOBUT Çalışmalarının Yıllara Göre Değişimi	52
Şekil 4.2. BOBUT Uygulamalarında Ölçülen Yapıların Çeşitliliği.....	62
Şekil 4.3. Çalışmaların Genel Sonuçları.....	69



ÇİZELGELER DİZİNİ

Çizelge 3.1. Literatür Taraması Süreci	38
Çizelge 3.2. Dahil Edilen Çalışmaların Kaynakları	39
Çizelge 4.1. Araştırma Dahilinde Analiz Edilen Çalışmalara Ait Özellikler	46
Çizelge 4.2. İncelenen Çalışmaların Yıllara Göre Dağılımı	52
Çizelge 4.3. Çalışmaların Yayın Türlerine Göre Dağılımı	53
Çizelge 4.4. Çalışmaların Amacına Göre Dağılımı	54
Çizelge 4.5. Çalışmaların Yayın Diline Göre Dağılımı	54
Çizelge 4.6. Çalışmalarda Kullanılan Örneklem Gruplarının Dağılımı.....	55
Çizelge 4.7. Çalışmalarda Kullanılan Farklı BOBUT Türlerinin Dağılımı.....	56
Çizelge 4.8. Çalışmaların Canlı BOBUT ve Simülasyonda Yapılma Oranları	56
Çizelge 4.9. Çalışmalarda Kullanılan Yazılımlar	58
Çizelge 4.10. Canlı BOBUT Uygulamalarında Kullanılan Örneklem Büyüklüğünün Dağılımı.....	59
Çizelge 4.11. Simülasyon Uygulamalarında Kullanılan Örneklem Büyüklüğünün Dağılımı.....	60
Çizelge 4.12. Çalışmalarda Kullanılan Madde Havuzu Büyüklükleri	61
Çizelge 4.13. Çalışmalarda Ölçülen Boyut Durumları	62
Çizelge 4.14. Çalışmalarda Kullanılan MTK Modellerinin Dağılımı	63
Çizelge 4.15. Çalışmalarda En Yaygın Kullanılan Madde Seçim Yöntemleri	64
Çizelge 4.16. Çalışmalarda En Yaygın Kullanılan Madde Kestirim Yöntemleri.....	65
Çizelge 4.17. Çalışmalarda En Yaygın Kullanılan Yetenek Kestirim Yöntemleri.....	66
Çizelge 4.18. Çalışmalarda En Yaygın Kullanılan Test Başlatma Kuralları.....	66
Çizelge 4.19. Çalışmalarda En Yaygın Kullanılan Test Sonlandırma Kuralları	67
Çizelge 4.20. Çalışmalarda Yapılan Madde Kullanım Sıklığı ve İçerik Dengeleme Oranları.....	68

SİMGELER

% : Yüzde

f : Frekans

M : Medyan

X : Ortalama

θ : Yetenek



KISALTMALAR

3PL	: 3-Parameter Logistic Model
BOBUT	: Bilgisayar Ortamında Bireye Uyarlanmış Test
CACT	: Computerized Adaptive Classification Testing
CAT	: Computerized Adaptive Testing
CD-CAT	: Cognitive Diagnostic Computerized Adaptive Testing
DINA	: Deterministic Inputs, Noisy "And" gate
DINO	: Deterministic Inputs, Noisy "Or" gate
EAP	: Expected A Posteriori
GDI	: Generalized Discrimination Index
GRM	: Graded Response Model
JML	: Joint Maximum Likelihood Estimation
KLI	: Kullback-Leibler Information
KTK	: Klasik Test Kuramı
MAP	: Maximum A Posteriori
MCAT	: Multidimensional Computerized Adaptive Testing
MFI	: Maximum Fisher Information
MFI-CB	: Maximum Fisher Information cut-point based
MFI-EB	: Maximum Fisher Information estimated-ability based
ML	: Maximum Likelihood
MLE	: Maximum Likelihood Estimation
MMLE	: Marginal Maximum Likelihood Estimation
MPWI	: Maximum Posterior Weighted Information
MTK	: Madde Tepki Kuramı
PWKL	: Posterior-Weighted Kullback-Leibler
WLE	: Weighted Likelihood Estimation

1. GİRİŞ

1.1. Problem Durumu

Ölçme ve değerlendirme süreçleri, yüzyıllardır bireylerin bilgi ve beceri düzeylerini belirlemek için kullanılmaktadır. Türkiye'de hala en popüler yöntem olmasına rağmen, kâğıt kalem testleri artık çağın gerekliliklerine cevap verememektedir. Milli Eğitim Bakanlığı (MEB) tarafından açıklanan verilere göre, 2024 yılı Mart ayı itibarıyla ülkemizde ilk ve ortaöğretimde 19 milyon öğrenci bulunmaktadır (MEB, 2024). Buna paralel olarak sınıflardaki öğrenci sayıları artmakta ve öğrencilerin bireysel farklılıklarını göz önünde bulundurmak zorlaşmaktadır. Bu durum, öğrencilerde var olan potansiyellerin gözden kaçırılmasına ve güvenilir bir ölçme ve değerlendirme yapılmasının sınırlandırılmasına neden olmaktadır. Geleneksel kâğıt kalem testleri, bazı durumlarda pratik ve erişilebilir bir yöntem olsa da öğrencilerin bireysel farklılıklarını ve yeteneklerini tam olarak ortaya koymada yeterli olmayabilir. Ayrıca, bu testlerin uygulanması ve değerlendirilmesi sırasında, büyük grupların sınav sonuçlarının hızla analiz edilmesi zor olabilir. Bununla birlikte, testler günlük yaşamın birçok alanında önemli bir rol oynamaktadır; gelişim süreçlerinin izlenmesi, hastalıkların tanınması, becerilerin ölçülmesi, mesleki yeterliliklerin tespiti ve kurum ya da okul giriş sınavları gibi. Bu kadar yaygın kullanım, bireyin hayatındaki önemini de arttırmaktadır. Ancak, bu önemli uygulamaların geleneksel yöntemlerle sınanmasının bireylerdeki farklılıkları ortaya çıkarmakta yetersiz olduğu artık kabul edilmektedir.

Bireyselleşme, eğitim alanında oldukça önemli bir kavramdır; bu bağlamda, ölçme ve değerlendirme alanında da bireyselleştirilmiş ölçme kavramı öne çıkmaktadır. Dünya genelinde, madde tepki kuramını (MTK) temel alan birey merkezli teknoloji yardımıyla yapılan sınavlar her geçen gün yaygınlaşmaktadır. Amerika'da "Hiçbir Çocuk Geride Kalmasın" (NCLB) reformu, çocuklara yetenekleri düzeyinde bireyselleştirilmiş eğitim ve ölçme süreci yaşatarak eğitimde eşitliği sağlamayı hedefleyen örnek çalışmalardandır (National Center for Learning Disabilities, 2005).

Türkiye'de 2017'den itibaren ehliyet sınavları ve Yabancı Dil Sınavı (YDS) bilgisayar tabanlı testlerle yapılsa da merkezi sınavlarda hala klasik test kuramına

(KTK) dayalı kâğıt kalem testleri kullanılmaktadır (MEB, 2017). Özellikle pandemi döneminde yaşanan uzaktan eğitim süreci, bireye uyarlanmış teknoloji destekli test uygulamalarının bir zaruriyet olduğunu göstermiştir. Bu test yöntemlerinin ülkemizde yaygınlaşmasının hem testi alan hem de uygulayan açısından daha faydalı olacağı öngörülmektedir.

Bireye uyarlanmış testlerin tarihi gelişimine baktığımızda, bilgisayar teknolojisinin eğitim sistemine dahil olmasıyla bazı köklü değişiklikler yaşanmıştır. 1980’lerde psikometri alanında yaşanan ilerlemelerle ve MTK’nin birleşmesiyle, bilgisayar ortamında bireye uyarlanmış testler (BOBUT) ortaya çıkmıştır (Hambleton, 1991). BOBUT, testi alan her kişi için maddenin zorluğunu bireyin yetenek düzeyine göre uyarlayarak bireye uyarlanmış, daha hassas ve güçlü ölçümler yapabilmektedir (Kingsbury ve Hauser, 2004). Özellikle eğitim sürecinin ölçme ve değerlendirme aşamasında, öğrencilerin motivasyonunu yüksek tutması ve sınavları uygulama, puanlama ve değerlendirme aşamalarında eğitimcilere sunduğu kolaylıklar gibi avantajları bulunmaktadır (Tseng, 2016).

BOBUT, aynı beceriyi kâğıt kalem testlerine kıyasla daha kısa sürede ve daha az maddeyle ölçerek zamandan tasarruf sağlamaktadır (Rudner, 1998). Ayrıca, zaman ve mekândan bağımsız olarak esnek bir ölçme imkânı sunmaktadır (Wainer, 2000). Kâğıt kalem testi uygulanan sınavlarda, test maddeleri sınavın ardından katılımcılarla paylaşıldığından, her sınav için yeniden madde hazırlama süreci zorluk yaratabilmektedir. BOBUT uygulamasında ise, kullanılan algoritmalar sayesinde maddeler daha verimli kullanılabilir. BOBUT; GRE, GMAT, TOEFL IBT, PTE Academic gibi dünya çapında bilinen birçok büyük ölçekli testte kullanılmaktadır (Kalender, 2009). BOBUT’u oluşturan dört ana bileşen bulunmaktadır: madde havuzu, madde seçim algoritması, yetenek tahmin yöntemi ve test durdurma kuralı (Stone ve Davey, 2011). Uygulama, ortalama yetenek düzeyine uygun ilk maddenin bireye sunulmasıyla başlamakta ve her yanıt karşılığında yetenek tahmini güncellenmektedir. Bu döngü, belirlenen durdurma kuralı gerçekleşene kadar devam etmektedir (Kingsbury ve Hauser, 2004). BOBUT uygulamaları, MTK’nin varsayımları doğrultusunda işlemekte ve bireylerin puanlarını aynı ölçek üzerinden standardize etmektedir. Bu sayede, farklı maddeleri yanıtlayan bireyler arasında karşılaştırma yapma imkânı sunmaktadır (Hambleton, 1991).

Son yıllarda, dünya genelinde BOBUT konulu yapılan çalışmaların sayısında artış olduğu görülmektedir. Google Akademik üzerinden “computerized adaptive testing” anahtar kelimeleriyle arama yapıldığında 20 binden fazla sonuçla karşılaşilmektedir. Türkiye’de de bu alandaki çalışmalar son yıllarda ivme kazanmış, bireyselleştirilmiş ölçme ve değerlendirme yöntemlerine olan ilgi artmıştır. Google Akademik üzerinden Türkiye bazlı arama yapıldığında yaklaşık 800 çalışma bulunduğu tespit edilmiştir. Ancak, mevcut araştırmaların geniş bir perspektifte derlenmesi ve sentezlenmesi gerekmektedir. Literatür incelendiğinde, Türkiye’de farklı alanlarda çeşitli amaçlarla yapılan BOBUT araştırmalarının henüz sistematik bir derlemesi yapılmamıştır.

Bu araştırmanın problemi, Türkiye’de BOBUT konulu yapılan çalışmaların hangi alanlarda yoğunlaştığı, hangi yıllarda yapıldığı, hangi teknikleri kullandıkları gibi konularda genel bir bakış açısı sunan sistematik derlemenin eksikliğidir.

1.2. Araştırmanın Amacı ve Önemi

Bu çalışmanın temel amacı Türkiye’de BOBUT konulu araştırmaların sistematik derlemesini yaparak mevcut bilgiyi sentezlemek, süreçteki ilerleyişini takip etmek, alanyazındaki boşlukları belirlemek ve gelecekteki araştırmalara yön göstermektir.

Kâğıt kalem testleri uzun bir süredir ölçme ve değerlendirme süreçlerinin temelini oluştururken, günümüzde bilgisayar tabanlı bireye uyarlanmış test yöntemleri giderek daha fazla önem kazanmaktadır. Ülkemizde, BOBUT konulu yapılan araştırmaların sayısında son yıllarda bir artış gözlemlenmiştir. Ancak, literatürde bugüne kadar yapılan çalışmaları kapsamlı bir şekilde inceleyen ve konuyu tüm yönleriyle ele alan bir sistematik derleme çalışmasına rastlanmamıştır.

Bu çalışmanın amacı, ölçme ve değerlendirme süreçlerinde kalitenin artmasına hizmet edecek BOBUT yöntemini kullanan çalışmaları sistemli bir şekilde derleyerek uygun bir çerçeve sunmaktır. Ayrıca mevcut durumu anlamak, gelecekteki araştırmacılara yol haritası sunmak ve temel bir kaynak oluşturmak için önemli görülmektedir. Eğitim sistemimizin gelişmesine ve eğitimde bireyselleşmenin artmasına yönelik sunulan bir adım olarak da değerlendirilebilir.

1.3. Araştırma Problemi

“Türkiye’de BOBUT konulu yapılan araştırmalarda mevcut durum nedir, BOBUT geliştirme ve uygulama süreçlerinde hangi kapsamlı özellikler ve metodolojik yaklaşımlar benimsenmiştir?” problemini temel alan araştırmada bazı alt problemler de araştırılmıştır. Bunlar:

1. Türkiye’de BOBUT konulu çalışmalar hangi yıllarda yapılmıştır?
2. Çalışmaların yayın türleri nelerdir?
3. Çalışmalar yoğunlukta hangi amaçla yapılmıştır?
4. Çalışmaların ne kadarı Türkçe ve ne kadarı İngilizce yayımlanmıştır?
5. Çalışmalarda hangi örneklem çeşidi yoğunluktadır?
6. Çalışmalarda farklı BOBUT türleri kullanım oranı nedir?
7. Canlı BOBUT uygulaması ve simülasyon verisi kullanım oranı nedir?
8. BOBUT geliştirme ve uygulama sürecinde en sık kullanılan yazılımlar hangileridir?
9. Çalışmaların örneklem büyüklükleri nedir?
10. Çalışmalarda kullanılan madde havuzu genişliği ne kadardır?
11. BOBUT çalışmalarında ölçülen yapıların çeşitliliği nasıldır?
12. Ölçülen yapıların tek boyutlu ve çok boyutlu olma durumları nedir?
13. En yaygın tercih edilen MTK modelleri hangileridir?
14. Testlerde en yaygın kullanılan madde seçim yöntemi, madde kestirim yöntemi, yetenek kestirim yöntemi, testi başlatma kuralı ve test sonlandırma kuralları nelerdir?
15. Çalışmalarda madde kullanım sıklığı kontrolü ve içerik dengeleme yapılma oranları nedir?
16. Çalışmaların genel sonuçları nedir?

1.4. Araştırmanın Kuramsal Temelleri ve İlgili Araştırmalar

1.4.1. Teknoloji ve Ölçme

İnsanlığın varoluşundan beri, bilgi ve becerilerin ölçülmesi ve buna dayalı bir değerlendirme yapılması önemli görülmektedir (Yılmaz, 2015). Tarihin ilerlemesiyle birlikte ölçme ve değerlendirme yöntemleri de değişime uğramıştır. Özellikle

teknolojide yaşanan gelişmeler bu süreçlerde dönüm noktası olmuş, ölçme ve değerlendirme süreçlerine yeni yaklaşımlar sunmuştur (Güneş, 2018).

Teknolojinin kullanılması diğer alanlarda olduğu gibi eğitimde de bir dijital dönüşüm yaşatmış, sunduğu imkânlarla eğitim öğretim sürecinin ayrılmaz bir parçası haline gelmiştir. Günümüzde MEB'in eğitim süreçlerinde teknoloji entegrasyonuna verdiği önem ve bu alandaki hızlı ilerlemeleri dikkat çekmektedir. Öyle ki; e-portfolyo, e-okul, Eğitim Bilişim Ağı, geliştirilen eğitim amaçlı mobil uygulamalar, dijitalleşen yayımlarla teknolojik imkânlardan faydalanılmaktadır. Yaşanan bu hızlı dönüşümle özellikle 2000'li yıllardan itibaren ölçmede teknoloji kullanımında da önemli bir artış gözlenmiş (Suna, 2021) ve bilgisayar tabanlı testler, çevrimiçi sınavlar, adaptif testler gibi yeni nesil ölçme ve değerlendirme araçları ortaya çıkmıştır. Ölçme ve değerlendirmede teknoloji kullanımı, yapılan öğretim etkinliklerinin ne derece etkili olduğuna dair eğitimcilere de hızlı dönütler vererek düzeltme imkânı sunmakta ve mevcut öğretim programının güncellenmesinde de önemli rol oynamaktadır (Çetin vd., 2004). Jimenez ve Boser'e (2021) göre yapay zekâ ve makine öğrenmesi gibi alanlarda yaşanacak değişimlerle gelecekte ölçme süreçlerinde daha adaptif yapıların oluşması da kaçınılmazdır.

Teknolojinin ölçme ve değerlendirme sürecine entegre olmasının bazı sınırlılıkları da bulunmaktadır (Aybek, 2021). Örneğin, ekonomik yük ve yatırım maliyetleri, teknolojiye erişim imkânı, yüz yüze etkileşimde azalma, öğretmenin rolündeki değişim ve teknolojiye dayalı içeriklerin hazırlanmasının zorluğu gibi faktörler öne çıkmaktadır.

1.4.2. Kâğıt Kalem Testlerinden Bilgisayara Geçiş

Ölçülen yeteneğin varlığının göstergesi olarak bir ölçme aracı kullanılması gerekmektedir. Ölçme ve değerlendirme alanında psikoloji bilimi, kâğıt kalem testlerinin geliştirilmesinde önemli bir paya sahiptir (Dirlik, 2021). Eğitim alanında da objektif bir ölçme ve değerlendirme sisteminin oluşturulmasında ilk uygulamalar kâğıt kalem testleriyle yapılmıştır. Bu testler tüm bireylere aynı maddelerin yöneltildiği ve aynı zaman diliminde cevaplanması gereken standartlaştırılmış testlerdir. Merkezi sınavlarda ya da akademik başarının ölçümünde sıklıkla kullanılmış ve hala da

kullanılmaktadır. KTK'ye dayanan bu ölçümler birey merkezli olmayıp puanlama ve uygulama sürecinde farklı zorluklar yaşandığı görülmektedir.

Dijital araçların eğitim öğretim sürecine entegre olmasıyla birlikte ölçme ve değerlendirme süreçleri de dijitalleşme eğilimi göstermiştir. Yukarıda bahsedilen sınırlılıklar ancak bilgisayar tabanlı testlerin geliştirilmesiyle aşılabilmektedir (Aybek, 2014). Bilgisayar tabanlı testler; teknolojik altyapıların gelişmesi, anlık dönütler, bireyselleştirilebilen testler ve erişilebilirlik imkânı ile kâğıt kalem testlerinin oluşturduğu pek çok sınırlılığı ortadan kaldırmıştır. Test uygulamalarının zamanla bilgisayar ortamında yapılma sıklığının artması literatürde BOBUT ve kâğıt kalem testi arasında karşılaştırma çalışmalarının yaygınlaşmasında da etkili olmuştur (Arslan, 2022; Bulut ve Kan, 2012; Çoban, 2020; Demir, 2023; Ita, 2014; Joeng, 2014; Kaptan, 1993; Kezer, 2013; Özbaşı, 2014).

1.4.3. Bireye Uyarlanmış Test Fikrinin Ortaya Çıkışı ve BOBUT'un Gelişim Süreci

Bireylerin yetenek ve bilgi düzeylerine göre maddelerin yöneltildiği, ölçme sürecinin merkezine bireyi alan bireye uyarlanmış test fikrinin ortaya çıkışı 20. yy başlarına dayanmaktadır. 1905 yılında Alfred Binet ve öğrencisi Theodore Simon'un geliştirdiği zekâ testi bireye uyarlanmış testlerin başlangıç noktasıdır (Bejar 1983; Hambleton vd., 1991b; Weiss, 1983). Bu test eğitim almaya uygun olmayanları yani zekâ seviyesi düşük olanları tespit etmek amacıyla geliştirilmiştir. IQ (Intelligence Quotient-Zekâ Katsayısı) olarak adlandırılan bir değerle yaşlılarına göre seviyelerin kıyaslandığı bu test dikkat, hatırlama ve problem çözme gibi birden çok alanı ölçebilmektedir. Bu sonuçlar testi alan bireyin IQ göstergesi olarak kabul edilir ve eğitime alınıp alınamayacağına karar verilir. Yoğun olarak çocukların zekâ ölçümü için kullanılsa da Binet yaptığı iki revizyonla yetişkin gruba da IQ testi geliştirmiş ve toplamda 11 yaş grubu için ölçüm yapılabilir bir zekâ testi oluşturmuştur. Binet'in ölümünün ardından Lewis M. Terman testi güncellemeye devam etmiş ve günümüze kadar kullanılmasını sağlamıştır (Thorndike, 1991).

Binet testinden sonra, testin karmaşıklığı sebebiyle 1970'lere kadar bireye uyarlanmış testlerin geliştirilmesinde sınırlı düzeyde çalışmalar yapılmıştır (Weiss, 2014). Ardından bilgisayar teknolojisinin kullanılmasıyla birlikte bireye uyarlanmış

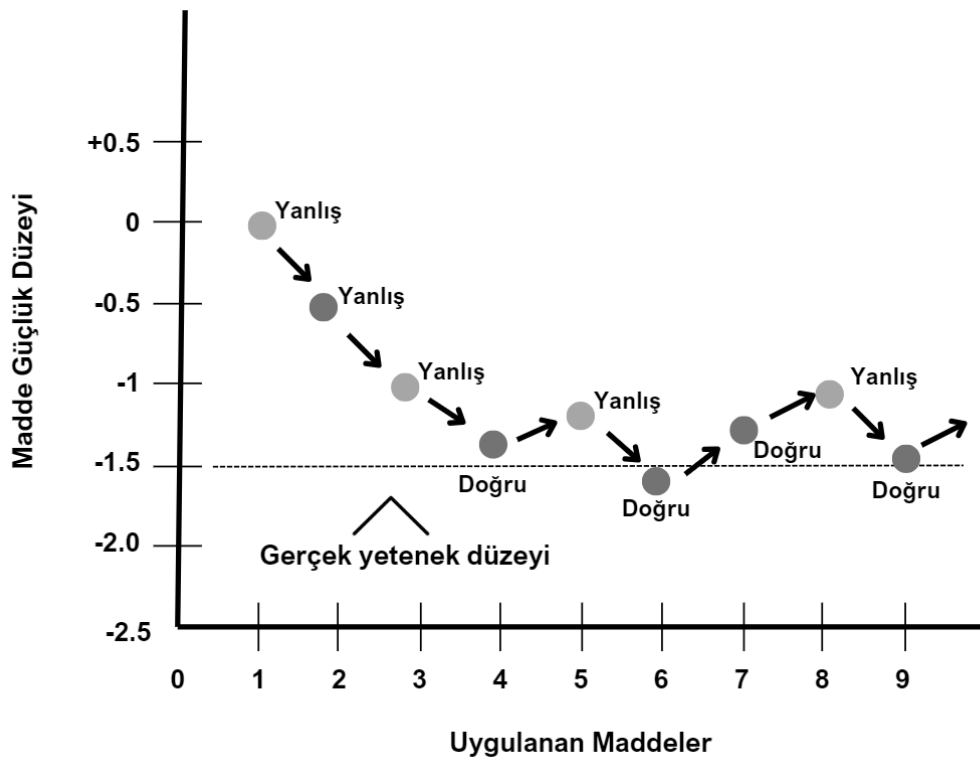
testler bilgisayar ortamına taşınmış, test geliştirme ve uygulama sürecinde yaşanan zorluklar aşılmıştır (Şenel, 2021).

Bilgisayar ortamında uygulanmaya başlanan testler, zamanla literatürde bilgisayar ortamında bireye uyarlanmış testler (BOBUT) (computerized adaptive testing – CAT) olarak anılmaya başlanmıştır. İlk uygulamaları KTK'ye dayanan ve genellikle maddeleri zorluk düzeyine göre kategorileştiren bu testler zamanla madde seçim stratejisi olarak MTK'yi kullanmaya başlamıştır (Weiss, 2014). 1980'lere gelindiğinde artık BOBUT çalışmaları artmış ve ABD Savunma Bakanlığı askeri personel seçme sınavında kullanmıştır (van der Linden ve Glas, 2010). Bu çalışmaların ardından BOBUT'un Amerika'dan tüm dünyaya geniş ölçekli sınavlarda uygulanmaya başlandığı ifade edilebilir.

1.4.4. BOBUT'un Tanımı ve Çalışma Prensibi

BOBUT, testi alan kişinin yeteneğinin bir bilgisayar tarafından tahmin edildiği ve seviyesine uygun maddelerin yöneltildiği bir test biçimidir. Kişilerin maddelere verdiği yanıtlar kaydedilir ve test esnasında düzenli olarak güncellenen bu yetenek düzeyi teta (θ) olarak adlandırılır. Bilgisayar tarafından tahmin edilen θ bulunurken arka planda farklı kestirim algoritmaları kullanılır. Bu algoritmalar sayesinde bireye tahmin edilen θ değerine uygun maddeler yöneltir, bireyler yetenek düzeyine uygun olmayan çok zor ve çok kolay maddelerle karşılaşmamış olur. Bu sayede bireylere kişiselleştirilmiş bir test deneyimi sunulur. Wainer (2000a), BOBUT'un daha hassas, güvenilir sınavlar sunma ve kısa sürede ölçüm yapabilme gibi avantajlara sahip olduğunu ifade etmiştir.

BOBUT'un çalışma prensibi madde seçimiyle başlar. Genellikle orta güçlükte bir madde ya da ön bilgilere dayalı olarak bireye uygun güçlükte bir madde, testi alan kişiye yöneltir. Bilgisayar buradan gelecek cevapla kişinin yeteneği hakkında bilgi edinmeye çalışır. Burada önemli olan hususlardan birisi yöneltilecek maddelerin ayırt ediciliği yüksek ve madde sayısı geniş olan bir havuzdan seçilmesidir. Kişinin verdiği cevaplarla değişen θ değeri, sonraki maddenin seçiminde de önemli rol oynar. Şekil 1.1'de görüleceği üzere bu süreç belirlenen sayıda madde uygulanana ya da belirlenen bir hata değerine erişene kadar devam eder ve bireyin bir yetenek puanı oluşturulur.



Şekil 1.1. İkili Puanlanan Maddelerde BOBUT'un İşleyişi

BOBUT'un uyarlanabilir yapısı bireylerin yetenek puanını en verimli şekilde tespit etmeye olanak sağlar. Bu test sürecinde madde seçim ve yetenek kestirim yöntemleri belirleyici rol oynar. BOBUT alanında birçok araştırmacı ve uzman, bu test tekniğinin ilerlemesi ve gelişmesi noktasında çalışmalar yapmıştır. Özellikle araştırmacı van der Linden ve David Weiss BOBUT'ta kullanılan yöntem ve teknikler konusunda alana oldukça önemli katkılar sağlamıştır (Stone ve Davey, 2011).

1.4.5. Madde Tepki Kuramı (MTK)

BOBUT'un ilerleyişinde ve yaygınlaşmasında etkili olan faktörlerin başında MTK gelmektedir. BOBUT'un işleyişini anlayabilmek adına MTK'yi de incelemek gerekmektedir. Öyle ki farklı testleri alan öğrencilerde karşılaştırma yapabilmek için MTK'nin önemi göz ardı edilemez. BOBUT'un amacı bireylerin yetenek düzeyleri ile bireye sunulan maddelerin güçlük düzeylerini eşleştirerek geçerliği yüksek ölçümler yapmaktır. BOBUT bu amaca hizmet etmek için modern ölçme kuramı olarak da adlandırılan MTK prosedürlerini temel alır.

1.4.5.1. Madde Tepki Kuramı (MTK) nedir?

Testler bireylerin performanslarını ölçmeye yönelik olsa da arka planda testin istenen davranışı ne derece ölçtüğüyle ilgili testi hazırlayan kişiye bilgi de verir (Hambleton vd., 1991b). Bu sebeple geçerli ve güvenilir ölçümler yapabilmek kullanılan test kuramlarının etkinliğine bağlıdır (Embretson ve Reise, 2000). Test kuramları aynı zamanda elde edilen sonuçların yorumlanmasını ve bireyler için ne anlama geldiğini ifade etmemize de yardımcı olur (Lord ve Novick, 1968). Bireylerin aynı sınavı iki kez aldığında puanlarının birbirinden farklı olması ya da ortam koşullarının bireylerin performansını ne derece etkilediği gibi konuları test kuramları ele alır (Reckase, 2009). Test kuramları temel varsayımlar ışığında sonuçları matematiksel olarak modeller ve bunları test kalitesini değerlendirmek için kullanır. Psikometrik ölçme ve değerlendirme alanında iki tür test kuramı vardır: klasik test kuramı (KTK) ve madde tepki kuramı (MTK) (Baker, 2001).

KTK'nin temelleri 1900'lerin başında Charles Spearman ile atılmış, Louis Thurstone, Frederic Lord ve Georg Rasch kuramın gelişmesinde ve uygulanmasında önemli katkılar sağlamıştır. KTK beş temel varsayım kullanarak bireylerin aldığı puan ile gerçek puan arasındaki farkı açıklamaya çalışır (Lord ve Novick, 1968). Bu kuram uzun yıllar ölçme teorilerinin standardını oluşturmuştur. KTK toptan puan hesaplarken tüm maddelerin aynı güçlükte olduğunu ve testi alan bireylerin aynı yeterliliğe sahip olduğunu varsayar. Bu durum heterojen gruplarda ve karmaşık testlerde sonuçların yanıltıcı olmasına sebep olur. 1950'lerde MTK'nin gelişmesiyle beraber farklı yetenek seviyesindeki bireyler ve farklı güçlükteki maddeler üzerinde daha güvenilir ve karşılaştırılabilir sonuçlar elde edilebilmektedir (Birnbaum, 1968; Lord, 1980).

MTK'nin geliştirilmesinde öncü isim Frederic M. Lord'dur. Lord, özellikle MTK modellerinde üç parametrelili lojistik model (3PL) konusunda çalışmalarıyla bilinir (Lord, 1980). Georg Rasch kurama katkı sağlayan diğer önemli araştırmacılardandır ve tek parametrelili lojistik modele (1PL) benzeyen Rasch modelini geliştirmiştir (Rasch, 1960). Rasch aynı zamanda çalışmalarıyla MTK'nin temel varsayımlarını şekillendirmiştir. Bu kişiler dışında da MTK'nin hem teorik temellerini geliştiren hem de pratik uygulamalarını arttıran çok sayıda araştırmacı vardır (Embretson ve Reise, 2000). 1980'lerde kullanımı yaygınlaşan bu kuram,

özellikle eğitim ve psikoloji alanında etkin şekilde kullanılmaktadır (Hambleton vd., 1991b).

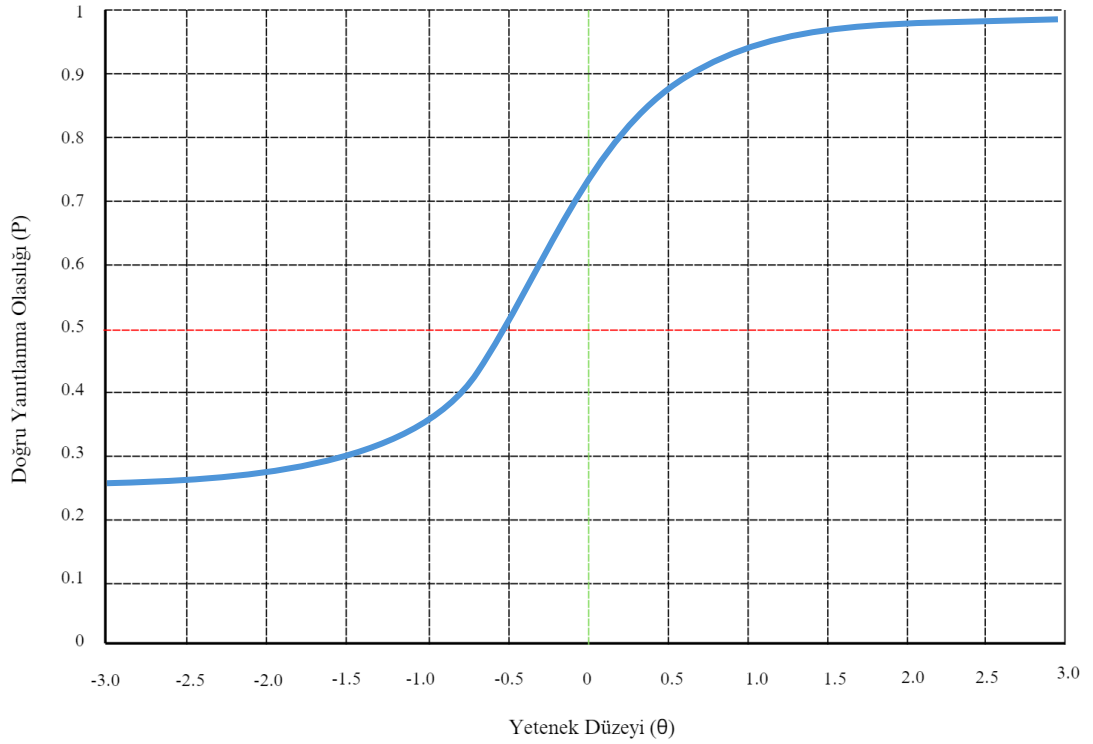
MTK'nin KTK'ye göre çok sayıda avantajı vardır. MTK, test maddelerini bağımsız analiz edebilmesi sayesinde verileri modelleyerek hem maddeleri hem de bireyleri aynı ölçeğe yerleştirebilir (Embretson ve Reise, 2000). KTK madde güçlüğü ile testi alan kişinin yeteneğini birleştirirken, MTK madde özelliklerinin ve testi alan kişinin özelliklerinin ayrı ayrı ayarlanmasına olanak tanır (van der Linden ve Hambleton, 1997). MTK, test sonucunu madde parametrelerinden ayrı tutar bu sayede bireylerin yetenekleri daha doğru şekilde tespit edilebilir. MTK, test performansını belirlerken bireyden aldığı doğru ve yanlış yanıtlarla beraber bu yanıtların olasılık tahminlerini de dikkate alır (Baker, 2001). Bireylerin maddelere verdiği cevaplardan yeterlik düzeyi hakkında ara bilgi elde edilir, ardından gelecek maddelerin daha fazla bilgi veren maddeler olarak seçilmesine imkân sağlar. Böylece birey hakkında daha fazla bilgi edinilmiş olur (Lord, 1980). MTK aynı zamanda BOBUT gibi modern test yöntemlerinin uygulanmasını da kolaylaştırır. Wainer'e (2000a) göre bireylerin yeteneğine uygun maddelerin seçilmesini sağlarken testin gereksiz uzamasını da engeller. Madde parametrelerinin değişmezliği ve yetenek parametrelerinin değişmezliği bu kuramın BOBUT'ta kullanılmasının temel etkenlerindendir (Reckase, 2009).

MTK'nin KTK'ye göre bazı sınırlılıkları da vardır. MTK, daha karmaşık modellere dayanır bu durum ileri düzey istatistik bilgisi gerektirebilir. Bir diğer durum ise küçük örneklem gruplarının madde parametrelerinin işleyişini olumsuz etkileyebilmesidir (Lord ve Novick, 1968). Her ne kadar MTK modellerinin uygulanmasında bazı sınırlılıklar olsa da sahip olduğu uyarlanabilirlik ve gelişmiş analiz yeteneği modern test araçlarının geliştirilmesinde önemli bir yardımcıdır (van der Linden ve Hambleton, 1997; Wainer, 2000a).

1.4.5.2. MTK'nin Temel Kavramları ve Varsayımları

MTK ilk bakışta karmaşık bir yapıya sahiptir, bu yapının işleyişini anlamak için temel kavramlar hakkında bilgi sahibi olmak önemlidir (Embretson ve Reise, 2000; Hambleton ve Rogers, 1991a).

- **Madde:** Bireylerin yetenek ölçümünde kullanılan ve puanlanan her bir soru için kullanılan terimdir.
- **Tepki:** Testi alan kişinin maddelere verdikleri yanıtlardır.
- **Yetenek:** Literatürde yeterlik olarak da geçen ölçülen gizil (örtük) özelliklerdir.
- **Kalibrasyon:** Yetenek ve madde parametrelerinin analizi sürecinde kullanılan işlemlerin tümüdür.
- **Madde Ayırt Ediciliği:** Testin amacını etkileyen bilenle bilmeyenin farkını ortaya koyan özelliktir (Embretson ve Reise, 2000). MTK’de bu değer pratikte 0 ile 2 arasında değer alır, madde ayırt ediciliğin yüksek olması istenen bir özelliktir.
- **Madde Güçlüğü:** Maddenin doğru cevaplanabilme potansiyelidir (Hambleton vd., 1991a). MTK’de -3 ile +3 arasında değer alır.
- **Madde Karakteristik Eğrisi:** Bireyde var olan yetenek ile maddelere doğru yanıt verme olasılığı arasındaki ilişkiyi gösteren grafiktir. Bireyin yeteneği arttıkça maddeleri doğru yanıtlama olasılığının da artması beklenir (Baker, 2001). Şekil 1.2’de örnek bir madde karakteristik eğrisi verilmiştir. Bu eğride bireydeki yetenek düzeyi artarken maddeye doğru cevap verme olasılığının nasıl değiştiği gösterilmektedir. Yatay eksen birey yetenek seviyesi (θ), dikey eksen maddenin doğru yanıtlanma olasılığını (p) göstermektedir.



Şekil 1.2. Madde Karakteristik Eğrisi (3PL Model)

BOBUT, MTK temelli bir yapıya sahip olduğundan temel varsayımların karşılanması BOBUT uygulamalarının güçlü etkinliği açısından önemlidir.

Tek Boyutluluk: Maddeler arasındaki korelasyonun güçlü ve tek bir gizil özelliği ölçtüğü varsayımdır. Psikolojik yapılarda net bir şekilde tek boyuttan bahsedilemeyeceği için maksimum düzeyde tek boyutluluğu sağlaması yeterlidir (Reckase, 2009; Şen vd., 2013; Deng vd., 2008).

Yerel Bağımsızlık: Maddelerin birbirinden bağımsız olarak cevaplanabilmesi ve birbirini etkilememesi olarak ifade edilebilir (Lord ve Novick, 1968). Bu varsayımın ihlal edilmesi maddeler arası beklenmeyen ilişkilerin oluşmasına neden olacağından test sonuçlarının yorumlanmasını zorlaştırabilir.

Madde ve Yetenek Parametrelerinin Değişmezliği: Ölçekte yer alan maddelerin özellikleri testi alan her birey ya da grup için aynı kalmalıdır. Aynı şekilde bireyde ölçülen yetenek düzeyi farklı testlerde de değişmemelidir. Bunun sağlanması için testte model uyumunun mükemmelliğinin sağlanması gerekmektedir.

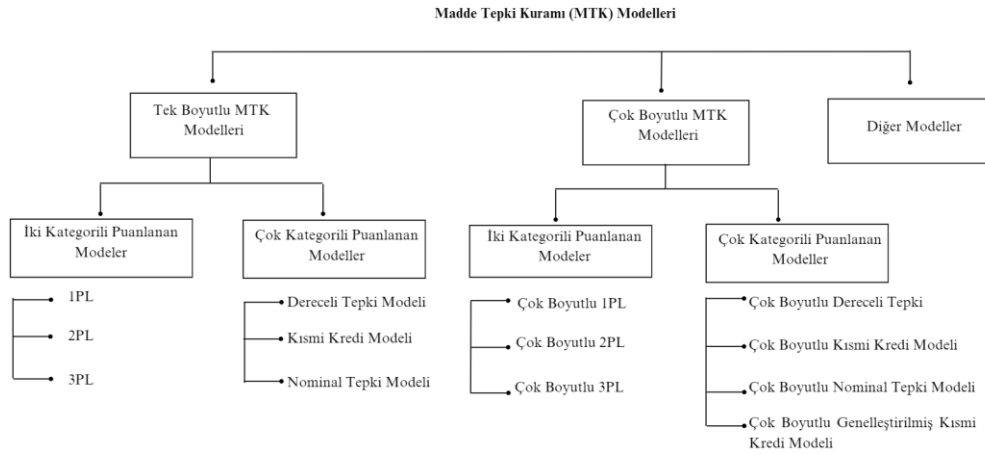
Model Uyumu: Seçilen modelin verilerle uyumlu olmasını ifade eder. Gözlenen değerde maddelere doğru yanıt verenlerin, modelin tahmin ettiği değerle benzer olması model uyumu için güçlü bir gösterge olarak kabul edilir.

Testin Hız Testi Olmaması: MTK için maddelere yanıt verilmemesi ya da yanlış cevaplanması süreden bağımsız olarak bireyde yeteneğin var olmadığının göstergesi olmalıdır (Hambleton ve Swaminathan, 1985).

1.4.5.3. MTK Modelleri

MTK’de ölçülen gizil özellikler θ olarak simgelenir, θ değeri bulunurken gözlenen değişkenlerden yani test maddelerinden faydalanılır. Psikometri alanının gelişimiyle çok fazla MTK modeli geliştirilmiş hala da geliştirilmeye devam edilmektedir, bunlardan bazılarına Şekil 1.3’te yer verilmiştir. MTK modelleri birey yanıtlarıyla madde özelliklerinin nasıl ilişkide olduğunu gösterir (Embretson ve Reise, 2000). Bu kısımda en yaygın kullanılan modellere değinilecektir. İlk sınıflandırma yöntemi ölçtüğü özellik bakımından, tek boyutlu ve çok boyutlu modeldir. Bireyde ölçülen tek bir gizil boyut ise, tek boyutlu model kullanılır. Örneğin matematik yeteneğinin ölçüldüğü bir sınavda tek bir boyuttan bahsedilebilir. Elbette ölçülen yapıların karmaşıklığından dolayı tek bir boyut ifadesi doğru olmayabilir, burada baskın boyutu ölçüyor olması da tek boyutlu olarak kabul edilmektedir (Crocker ve Algina, 2006). Çok boyutlu modelde ise bireyde birden fazla yetenek ölçülebilir. Örneğin dil sınavında bireylerin gramer bilgisi, konuşma ve dinleme becerisi gibi birden fazla boyut ölçülebilir. Her iki modelin de ölçme süreçlerinde uygun koşullarda avantajlarından bahsedilebilir. Diğer sınıflandırma yöntemi ise tek boyutlu modelin alt modelleri olan puanlama yöntemiyle ilgili olarak iki kategorili ve çok kategorili puanlama olmak üzere ikiye ayrılır. İki kategorili puanlamada doğru ve yanlış vardır yani 1 ve 0, doğru madde 1 değerini alırken yanlış maddeler 0 değerini alır. Bu model sıklıkla başarı testlerinde kullanılır ve Rasch modeline dayanır. Rasch modeli MTK uygulamalarında en sık kullanılan modeldir. Rasch modelinin farklı bir versiyonu da 1PL (parametrelili lojistik) modelidir. Bu modelin 2PL, 3PL şeklinde genişletilmiş versiyonları da vardır (Ostini ve Nering, 2006). Burada belirtilen iki kategorili modeller kestirilen madde parametre sayısına göre 1PL (madde güçlüğü) ya da Rasch, 2PL (madde güçlüğü ve ayırt ediciliği) ve 3PL (madde güçlüğü, madde ayırt ediciliği

ve madde tahmin/şans) şeklinde farklı isimler alabilmektedir. Çok kategorili puanlama ise cevapların sadece doğru ve yanlış olarak değerlendirildiği iki kategorili puanlamanın aksine birden çok kategorili cevapların analiz edilmesinde kullanılır. Bu model, bir maddeye verilen farklı cevap kategorilerinin her biri için olasılık tahmin eder ve bu tahminler madde parametrelerine (örneğin, zorluk, ayır edicilik) bağlı olarak değişir. Çok kategorili modellerden bazıları: GRM (Graded Response Model – Dereceli Tepki Modeli) (Samejima, 1969), her bir madde için belirli bir sırayla düzenlenmiş kısmi kredi kategorileri içeren modeldir. PCM (Partial Credit Model – Kısmi Kredi Modeli), (Wright, 1982), her bir madde için belirli bir sıralama olmaksızın kredi kategorileri içeren modeldir. GPCM (Generalized Partial Credit Model – Genelleştirilmiş Kısmi Kredi Modeli) (Patz ve Junker, 1999), PCM'in genelleştirilmiş versiyonu olup, her bir kategori için farklı ayır edicilik parametrelerine izin veren modeldir. NRM (Nominal Response Model – Nominal Tepki Modeli) (Bock, 1972), Cevap kategorilerinin herhangi bir sıralama olmaksızın kullanıldığı modeldir. Bunların dışında MTK modellerinden DINA (Deterministic Inputs, Noisy "And" gate) ve DINO (Deterministic Inputs, Noisy "Or" gate) (Junker ve Sijtsma, 2001) modelleri bilişsel tanı modeli olarak bireylerde bilgi ve becerilerin var olup olmadığını kontrol etmek için kullanılır. Bu modeller, katılımcıların farklı kategorilerdeki cevaplarını analiz ederek, her bir madde için belirli parametreler ve katılımcıların yetenek seviyeleri hakkında bilgi sağlar. Bu sayede, çok kategorili verilerde daha ayrıntılı ve hassas değerlendirmeler yapılabilir. Özellikle anketler, tutum ölçekleri ve yetenek ve performans ölçümü gibi çok kategorili veri setlerinde tercih edilir (van der Linden ve Hambleton, 1997; Samejima, 1969).



Şekil 1.3. MTK Modellerinin Sınıflandırılması (Embretson ve Reise, 2000)

BOBUT madde havuzu, MTK yöntemleriyle güçlük, ayırt edicilik ve tahmin/şans gibi parametrelere göre kalibre edilmiş çok sayıda maddelerden oluşur. Burada en sık kullanılan madde kestirim yöntemleri: marginal maximum likelihood; MML (Bock ve Aithin, 1981), joint maximum likelihood; JML (Birnbbaum, 1968) ve conditional maximum likelihood; CML'dir (Andersen, 1973).

MTK modelleri bireylerin yetenek kestirimi için de bazı yöntemler kullanır. Bu yetenek kestirim yöntemleri BOBUT'un da temel işleyişinin bir parçasıdır. En sık kullanılan modellerden ilki maksimum olabilirlik tahmini (MLE), veri setinde en iyi model parametrelerini bulmak için çalışır. Bu yöntem modellerle verinin yüksek uyumunu sağlamak üzere mevcut verilerle en uyumlu olan parametreleri seçer (Bock ve Aitkin, 1981). Örneğin matematik yeterliklerinin ölçüldüğü çoktan seçmeli bir sınavda, öğrencilerin doğru yanıt verme olasılığı yetenek düzeyine bağlıdır. Bu gizil değişkenin tahmininde MLE yöntemi kullanılabilir. Büyük örneklem gruplarında etkili sonuç verir (Meuleman ve Billiet, 2009). Yetenek kestiriminde sıklıkla kullanılan bir diğer yöntem Bayesci kestirimdir (Bayesian Estimation). Bu yöntem Bayes teoremine dayanır, bireyin yetenek tahmininde ön bilgi ile test sonuçlarını beraber kullanır. Bayesci kestirim kapsamında iki yaygın yöntem bulunmaktadır: Beklenen A Posteriori (EAP) ve Maksimum Olabilirlik A Posteriori (MAP). EAP yöntemi, bireyin yetenek düzeyinin tahmininde tüm olası yetenek düzeylerinin ağırlıklı ortalamasını kullanırken, MAP yöntemi ise yetenek tahminini, olasılık fonksiyonunun maksimum olduğu noktada yapar. Bu yöntemler, özellikle küçük örneklem gruplarında ve ön

bilgiye dayalı yetenek tahminlerinde tercih edilmektedir (Şen vd., 2020; van der Linden ve Glas, 2000).

1.4.5.4. Bireye Uyarlanmış Testler ve MTK

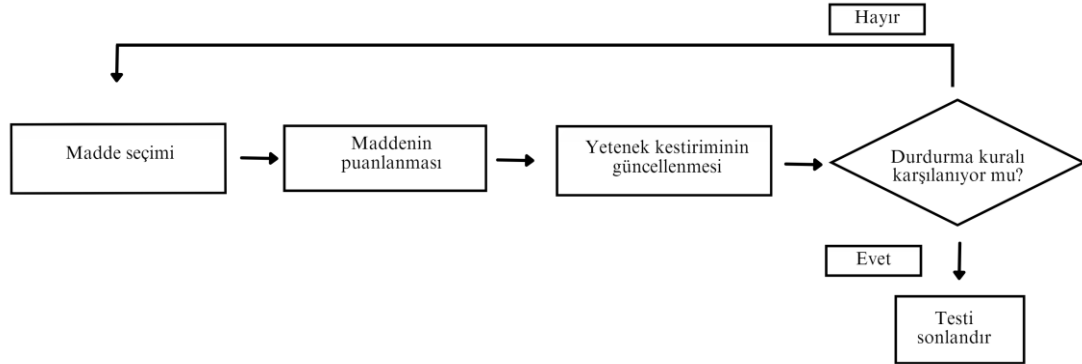
Bireye uyarlanmış testlerin tasarım ve analiz sürecinde MTK'den destek alınır. Bireye uyarlanmış testler MTK tabanlı olmak zorunda değildir ancak BOBUT'un gelişim sürecini hızlandıran en önemli etkenlerden biri MTK'dir. MTK'nin psikometrik özellikler sunması, prensiplerinin BOBUT ile uyumlu olması BOBUT için oldukça elverişlidir (Embretson ve Reise, 2000; Şenel, 2021; van der Linden ve Glas, 2010). Çünkü MTK'nin temel varsayımlarından olan bireyin aldığı teste göre yetenek kestiriminin değişmemesi (madde parametrelerinin değişmezliği) ve madde parametrelerinin de bireylerdeki değişimden etkilenmemesi (yetenek parametrelerinin değişmezliği) en temel gerekçelerdendir. MTK kestirimlerinin teknoloji olmadan yapılabilmesinin zorluğu nedeniyle BOBUT'un teknolojiye paralel uygulama sıklığının arttığı söylenebilir.

BOBUT algoritmaları arka planında MTK prensiplerine dayanmaktadır. Örneğin bireyin yetenek düzeyine göre havuzdan çekilecek olan maddelerde madde güçlük parametresi kullanılır. Yine bu testlerin analizinde MTK analiz tekniklerinden faydalanılır. Güvenilir ve geçerli bir BOBUT geliştirmek için maddelerin güçlüğü ve ayırt ediciliği oldukça önemli rol oynamaktadır.

1.4.6. BOBUT Temel Adımları ve Algoritmalar

Wainer ve arkadaşlarına (2000b) göre, BOBUT tasarımında yanıtlanması gereken temel sorular vardır: Test nasıl başlayacak, nasıl devam edecek ve nasıl sonlanacak? Bu sorular test tasarımında MTK tabanlı algoritmaları oluşturmaya rehberlik eder. Bireyde var olan yeteneğin en doğru şekilde kestirilmesi ve analiz edilmesi için BOBUT'ta takip edilmesi gereken temel adımlar bulunmaktadır. Bu adımlar test uygulama sürecinden en yüksek verimi almak için tasarlanmaktadır. BOBUT farklı yetenekleri ölçebilen ve madde sayısının geniş olduğu bir havuzdan madde seçimiyle giriş kuralına uygun olarak başlar. Ardından Şekil 1.4'te görselleştirilen döngüyle devam eder. Yanıtlayıcının verdiği her cevabın ardından yetenek kestirimi güncellenir. Yetenek kestiriminin ardından test sonlandırma

kuralının gerçekleşip gerçekleşmediği sınanır. Sonlandırma kuralı gerçekleşmediği takdirde madde havuzundan bireyin güncellenen yetenek düzeyine uygun yeni madde seçilir. Test sonlandırma kuralı gerçekleşene kadar bu süreç devam eder.



Şekil 1.4. BOBUT Adımları (van der Linden ve Glas, 2010; Wainer, 2000a)

1.4.6.1. Madde Havuzu Oluşturma

BOBUT'ta ölçülmek istenen yapıyı en iyi şekilde sınayacak farklı güçlüklerde en çok bilgiyi sağlayan maddelerin MTK prensiplerine dayalı olarak oluşturulması gerekmektedir (van der Linden, 2006). Madde havuzunun nitelikli maddelerden oluşması testten alınacak sonuçları da doğrudan etkilemektedir. Madde havuzu oluştururken testin amacının net olarak belirlenmesi ve kapsama uygun maddelerin oluşturulması gerekmektedir. Kolay maddeler düşük yeterlik düzeyindeki bireyler için, zor maddeler ise yüksek yeterlik düzeyindeki bireyler için en doğru yetenek bilgisini verir. Bu sebeple ayırt edicilik gücü yüksek ve kolay-zor dengesi iyi ayarlanmış bir madde havuzu oluşturmak amaçlanmalıdır.

Ölçülen gizil yapıyı en iyi şekilde ortaya koyacak maddenin oluşturulması testin geçerliği açısından da önemlidir. Embretson ve Reise'e (2000) göre, yüksek ayırt ediciliğe sahip, çeşitli güçlük seviyelerinde geniş bir madde havuzu oluşturulması, yetenek tahminini doğrudan etkilemektedir. Test başlatma kuralı, ilk maddenin orta güçlükte olması gerektiğini belirtiyorsa, orta güçlükteki maddelerin sayısının artırılması gerekmektedir. Bununla beraber normal dağılım varsayımıyla da testte orta güçlükteki maddelere daha fazla yer verilmelidir. Madde havuzunun içerik dengesi ve ölçme kalitesi açısından düzenli olarak değerlendirilmesi de yapılmalıdır (Magis vd., 2017).

BOBUT için madde havuzunun genişliğinden söz edilse de net bir madde sayısı vermek güç olabilmektedir (Şahin ve Öztürk, 2019; Magis vd., 2017). Ancak yetenek analizine hangi madde havuzunun daha çok etki edeceğine ilişkin yapılan simülasyon çalışmalarında 200 ve üzeri madde barındıran havuzun yeterli olacağı çıkarımında bulunulabilir (Şahin ve Weiss, 2015; van der Linden ve Glas, 2000). Dar bir madde havuzunda bireylerin aynı maddelerle yeniden karşılaşması madde maruziyetini doğuracağından test güvenilirliğini olumsuz etkileyebilir. Madde havuzunun genişletilmesi için bazı teknikler de ortaya çıkmıştır. Örneğin bilgisayar yardımıyla benzer güçlükteki maddeler çoğaltılabilmektedir. Oluşturulan madde havuzunun ve maddelerin parametre değerlerinin belirlenebilmesi için son olarak deneme uygulamasına da ihtiyaç vardır.

1.4.6.2. Test Başlatma Kuralı ve İlk Madde Seçimi

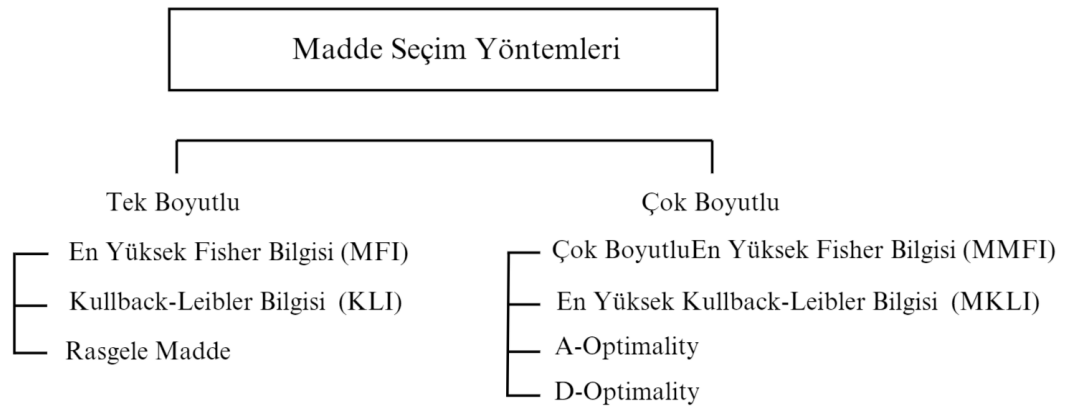
Bu kural bireye yöneltilecek ilk maddeyi belirlemek için kullanılır. Magis'e (2017) göre, sıklıkla kullanılan yöntem ortalama güçlükte bir maddeyle teste başlamaktır, bu kural nötr teta başlangıcı olarak da bilinir. Birey hakkında herhangi bir ön bilgi olmadığı durumlarda tercih edilir (Embretson ve Reise, 2000). Bu yöntemin tercih edileceği durumlarda havuzda orta güçlükte madde sayısı azalacağından testin güvenilirliğinin etkilenmemesi için madde havuzu tasarımının ve dengesinin çok iyi planlanması gerekmektedir.

İlk madde seçiminde kullanılan bir diğer yöntemde ise birey hakkında ön bilgi varsa, bu bilgi yetenek düzeyine uygun maddenin seçiminde kullanılabilir (Embretson ve Reise, 2000). Örneğin diploma notu, daha önce girdiği ön test sonuçları bireyin yetenek seviyesiyle ilişkilendirilebilir. Kalendergil (2009) makalesinde üniversite giriş sınavında uygulanacak BOBUT için lise diploma notunun yetenek kestiriminde ön bilgi verebileceğini ifade etmektedir.

1.4.6.3. Madde Seçim Yöntemleri

Bireyin yetenek düzeyine en uygun maddenin seçilmesi testin amacına ulaşabilmesi için oldukça belirleyici bir etkiye sahiptir (Wainer vd., 2000b). BOBUT'un bireye uyarlanması adımı yetenek kestiriminde gerçekleşir. Yeteneğin kestirimi ve ardından uygun maddenin seçimi, test sonlandırma kuralı gerçekleşene kadar devam eder. Bireye sunulan maddeler havuzdan daha önce yöneltilmemiş

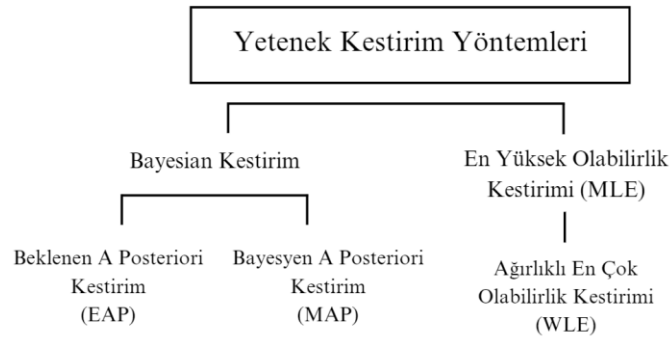
maddelerden seçilmelidir. Şekil 1.5’te görüleceği üzere BOBUT testlerinde kullanılan madde seçim yöntemleri testin yapısına göre tek boyutlu ve çok boyutlu olarak iki ana kategoriye ayrılır. Tek boyutlu yapılarda sıklıkla tercih edilen maksimum Fisher bilgisi (MFI) (Lord, 1980), yetenek kestiriminde kullanılan maksimum olabilirlik (ML) yöntemini esas alır (van der Linden, 2006). MFI algoritması birey hakkında en fazla bilgiyi verecek maddeyi seçmeye çalışır yani seçtiği maddeyle standart hatayı düşürmeyi hedefler. Maddeden alınan yanıtı göre θ düzeyinin kestirilmesi için madde bilgi fonksiyonunu kullanır. Diğer yaygın kullanılan madde seçim yöntemi Kullback-Leibler bilgisi (KLI) (Kullback ve Leibler, 1951) iki olasılık dağılımının birbirinden ne kadar farklı olduğunu ölçen bir yöntemdir. Bu yöntem de MFI gibi birey hakkında en çok bilgiyi veren maddeyi seçmek için çalışır. Bu yöntemin Posterior Weighted Kullback-Leibler (PWKL) gibi bireylerin yetenek dağılımına göre madde seçimi yapan versiyonları da bulunmaktadır. Bireyin yetenek dağılımına göre madde seçimi yapan bir diğer yöntem Maximum Posterior Weighted Information (MPWI)’dır. Rastgele madde seçimi ise, belirli bir bilgi kriterine dayanmadan, test maddelerinin rastgele seçilmesi yöntemidir. Testin başlangıç aşamalarında veya çeşitliliği sağlamak için kullanılır. Çok boyutlu yapılar için tercih edilen madde seçim yöntemlerinden olan MMFI yöntemi, MFI’nın bir versiyonudur ve birden fazla yetenek boyutunda her bir maddenin sağlayacağı toplam bilgi miktarını maksimize etmeyi amaçlar. Çok boyutlu yapılarda, tüm yetenek boyutlarında dengeli bir bilgi sağlamak için kullanılır. MKLI yöntemi, KLI’dan türetilmiş bir yöntem olup kişinin mevcut yetenek tahminleri ve olası yetenek düzeyleri arasındaki ayrımı en iyi yapan maddeyi seçer. Çok boyutlu yapılarda çeşitli boyutlarda en bilgilendirici maddelerin seçilmesini sağlar. A-Optimality yöntemi (Chang ve Ying, 1996), test sırasında bilgi matrisinin toplam miktarını (iz) minimumda tutarak bireyin yetenek seviyesini en doğru şekilde tahmin etmek için en uygun maddeyi seçer. Çok boyutlu yapılarda, bilgi matrisi izini minimize ederek tahmin doğruluğunu artırmak için kullanılır. D-Optimality yöntemi (van der Linden ve Reise, 1998) ise çok boyutlu yetenek uzayında, madde seçiminde bilgi matrisi determinantını maksimize etmeye çalışır. Çok boyutlu yapılarda, tüm yetenek boyutlarında yüksek doğrulukla ölçüm yapmak için kullanılır.



Şekil 1.5. Madde Seçim Yöntemleri (van der Linden ve Pashley, 2000; Wainer, 2000a)

1.4.6.4. Yetenek Kestirim Yöntemleri

Bireyde nihai yetenek kestiriminin yapılabilmesi BOBUT'un temel amacıdır. Bu amaç dâhilinde uygulanan testi başlatma, sonraki madde için yetenek kestiriminin güncellenmesi ve sonlandırma döngüsü yetenek kestirimi tamamlanıncaya kadar devam eder. van der Linden ve Glas'a (2010) göre yetenek kestiriminde kullanılan farklı yöntemler vardır. Bunlardan en yaygın kullanılanlarına Şekil 1.6'da yer verilmiştir.



Şekil 1.6. Yetenek Kestirim Yöntemleri (van der Linden ve Glas, 2010)

En yaygın kullanılan yetenek kestirim yöntemlerinden biri maksimum olabilirlik (MLE) yöntemidir (Lord, 1980). Bu yöntemde, yeteneğin tahmin edilebilmesi için bireyin en az bir doğru ve en az bir yanlış cevap vermesi gerekmektedir. Aksi halde tüm testin doğru ya da yanlış cevaplanması durumunda yetenek kestirimi yapılamaz. Bu sınırlılığı aşmak için, test sırasında bireye hem yanlış

cevaplayabileceği kadar zor hem de doğru cevaplayabileceği kadar kolay maddeler sunulmalıdır (Şenel ve Kutlu, 2018).

Bir diğer yaygın yetenek kestirim yöntemi olan Bayesci kestirim yönteminde (Mislevy, 1986) mükemmel yanıt örüntüsü aranmaz. Bu yöntem yetenek kestirimini bireyin verdiği ilk yanıtla beraber başlatacağı için test başlangıcında kullanılabilir. Ancak MLE'nin güvenilirlik değerine ulaşmak için daha fazla madde sayısına ihtiyaç duyulur (Mislevy, 1986; Mislevy ve Stocking, 1989; van der Linden ve Glas, 2010).

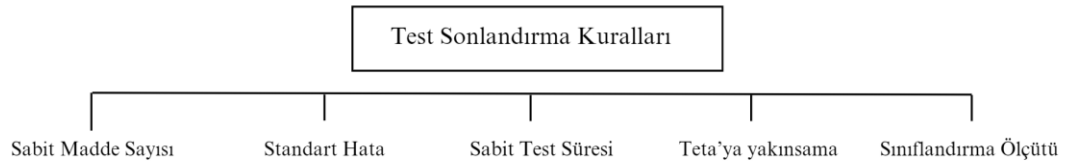
Beklenen A Posteriori (EAP) kestirimi (Bock ve Mislevy, 1982), Bayesci yönteme dayalı olarak kişinin yanıtlarını ve önceden belirlenmiş yetenek dağılımını kullanarak mevcut yetenek düzeyini tahmin eder. Küçük örneklem gruplarında avantaj sağlarken, sonuçlar başlangıçta belirlenen önsel dağılımdan etkilenir. Yanlış bir dağılım kullanıldığında sonuçlar da buna bağlı olarak yanlış hesaplanabilir (Bock ve Mislevy, 1982).

Bayesian A posteriori (MAP) kestirimi de EAP gibi önsel dağılımı kullanan bir yöntemdir. Bayesci yönteme dayanır. EAP, bireyin yetenek seviyesini belirlerken dağılımın ortalama değerlerini baz alır. MAP ise en yüksek değeri bulur ve bu değeri bireyin olası yetenek seviyesi olarak belirler. Ayrıca, MAP önsel dağılımdan daha fazla etkilenir (Chen ve Ankenmann, 2004). Ağırlıklı olabilirlik tahmini (WLE) geliştiricisi Terence A. Warm'dır (1989). MLE yöntemine benzer ancak uç değerlerdeki hataları azaltmak için kestirim yaparken ağırlık kullanır (van der Linden ve Glas, 2010).

Yetenek kestiriminde; Fisher'in puanlama yöntemi (Fisher, 1925), ağırlıklı maksimum olabilirlik (WML) (Warm, 1989), tam bilgi ile maksimum olabilirlik tahmini (MLEF) (Bock ve Aitkin, 1981), ağırlıklı en küçük kareler (WLS) (Rao, 1973) ve Newton-Raphson algoritması (Press vd., 1992) gibi yaygın olmayan farklı yöntemler de mevcuttur. Bunlar için ilgili kaynaklar incelenebilir.

1.4.6.5. Test Sonlandırma Kuralı

BOBUT çalışma prensibine göre test önceden belirlenen sonlandırma kuralı gerçekleşene kadar devam eder. Sonlandırma kuralının seçimi yapılırken testin amacı, madde havuzunun özelliği gibi durumlar göz önüne alınmalıdır (Segall, 2004). Test sonlandırmak için kullanılan en yaygın yöntemler Şekil 1.7'de verilmiştir.



Şekil 1.7. Test Sonlandırma Kuralları (Wainer vd., 2000b)

Sabit madde sayısı yöntemi (Wainer vd., 2000b) önceden belirlenmiş madde sayısına erişilinceye kadar test devam eder. Bu sonlandırma kuralında bireyler yetenekleri hakkında pek bilgi vermeyen maddelerle karşılaşabilmekte ve BOBUT uygulamalarının etkili olmasını sınırlayabilmektedir (Choi vd., 2010). Bu yöntem tek başına uygulanabileceği gibi diğer yöntemlerle beraber kullanıldığında bu sınırlılıklar en aza indirilebilir (Şenel, 2021).

Standart hata yönteminde (Weiss, 1983) ise testi alan her kişi için güvenilirlik bilgisini, yetenek kestiriminin standart hatası olarak kabul eder. Bu standart hata kabul edilebilir bir seviyenin altına geldiğinde testin sonlandırılması istenebilir. Ölçüm bu değer altına inmedikçe testin madde seç – uygula – kestirim yap döngüsü devam eder. Bu yöntem BOBUT'ta kullanılan en yaygın test sonlandırma yöntemidir (Weiss ve Kingsbury, 1984).

Sabit test süresi yönteminde (Wainer vd., 2000b) bireyler test için belirlenen süreye erişilince test sona erer. Yani bireyin yetenek kestirimi o sürede cevapladığı maddelerle yapılır. Bu yöntemde bireyler yetenekleri düzeyinde yeterli miktarda maddeye erişemeyebilir bu durum yetenek kestirimini olumsuz etkiler.

Teta'ya yakınsama yönteminde (Kinsbury ve Weiss, 1983) ise test boyunca kestirilen yetenek düzeyindeki yani θ 'daki değişim testin sonlarına doğru azalmaktadır (Weiss ve Kingsbury, 1984). Bu değişim değeri baz alınarak yetenek tahmini belirli bir kararlılığa ulaştığında testi sonlandırma kriteri olarak uygulanır.

Sınıflandırma kriteriyle sonlandırma yönteminde, birey önceden belirlenmiş objektif bir kriteri aştığında test sonlandırılır. Kişinin belirli bir yeterlik düzeyine sahip olup olmadığını belirlemek için kullanılır. Örneğin seviye belirleme sınavında bu kriter kullanılarak birey hakkında geçti ya da kaldı şeklinde karar alınabilir. Bu yöntem zaman yönetimi açısından avantaj sağlar, testin gereksiz uzamasının önüne geçer (van der Linden ve Glas, 2010; Weiss ve Kingbury, 1984).

1.4.7. Farklı BOBUT Türleri

Çeşitli ihtiyaçları karşılamak ve farklı test senaryolarını karşılaştırmak için farklı BOBUT türleri geliştirilmiştir. Bu kısma kadar tam uyarlamalı BOBUT türünden bahsedilmiş olsa da birden fazla yeteneğin kestirimini aynı anda yapmak, bireylerin bilişsel becerilerini detaylı şekilde değerlendirmek gibi amaçlarla geliştirilen BOBUT türleri de mevcuttur (Wainer vd., 2000b). Çalışmanın bu kısmında BOBUT'un esnekliğini arttırarak farklı alanlara uygulanmasını sağlayan BOBUT türlerine değinilmiştir.

1.4.7.1. Çok Aşamalı BOBUT (Multistage Testing-ca-MST)

Bu test türünde bir maddeye yanıt verilmesinin ardında diğer bir maddenin seçildiği BOBUT uygulamasının aksine, bireylerin belli madde setini yanıtladıktan sonra havuzdan yeni madde setini cevaplaması gerekmektedir (Cook ve Eignor, 1991; Wainer ve Kiely, 1987). En yaygın kullanılan modeli iki aşamalı testlerdir (two stage tests). Bu yöntemde tüm katılımcılara 10 ile 20 maddelik bir test uygulanır, puanlanır ve sonraki madde seti için maddelerin zorluk seviyesine karar verilir. Katılımcıların nihai puanı iki aşamanın sonunda tüm maddelere verdikleri yanıtlardan oluşturulur. İki aşamalı testler bilgisayar olmadan da uygulanabildiği için geleneksel yöntemlerin yerine ilk tercih edilebilen uyarlamalı testler olmuştur (Lord, 1980). Ancak bu uyarlama madde bazlı değil madde seti şeklindedir. Bireyler modül adı verilen madde setlerini yanıtladıktan sonra yeniden yetenek uyarlaması yapılarak yeni madde setlerini yanıtlamaktadır. Bireyin performansı ölçüt puanının altındaysa daha kolay bir modül, ölçüt puanının üstündeyse daha zor bir modül uygulanır. İki aşamalı testlerde, bireyler farklı güçlük seviyesindeki maddelere yanıt verdikleri için doğru cevaplar üzerinden puanlama yapılamamakta ve bu durum puanlamanın adil olmasını engelleyebilmekteydi. Bu sorunu çözmek için, ikiden fazla modülün uygulandığı çok aşamalı bireysel testler, MTK tabanlı yapısı sayesinde, bireylerin yetenek seviyesine uygun maddeler yönlendirmektedir. Çok aşamalı bireysel testler içerik dengesini sağlamada oldukça başarılı olduğu için günümüzde yeniden popüler hale gelmiştir (Magis vd, 2018).

1.4.7.2. Bilişsel Taniya Dayalı BOBUT (Cognitive Diagnostic Computerized Adaptive Testing - CD-CAT)

Bilişsel tanı modeline (cognitive diagnostic model – CDM) dayanan bu test türü, MTK modellerinin aksine testi alan katılımcıları kategoriler halinde sınıflandırmak için tasarlanmıştır. Bu test türü bireylerin hangi konuda yeterli olduklarını ve hangi konuda zorlandıklarını tespit etmek için de kullanılabilir. Örneğin bireylere uygulanan matematik testiyle belirli bir problem çözme becerisi ölçülebilir ve bireyin bu yeterliliğe hakim olup olmadığı değerlendirilir. Bilişsel taniya dayalı BOBUT'un amacı, sınava girenleri bir madde setine verdiği yanıtlara göre hangi becerilerde ustalaşıp ustalaşmadığını tahmin etmektir. Burada Q matrisi kullanılır ve beceriyi doğru ölçebilecek madde havuzu oluşturmak kritik öneme sahiptir. MTK modeline benzer yapısıyla doğru kalibre edilmiş bir madde havuzu oluşturulmalıdır, aksi takdirde test geçerliği sağlanamayabilir (Embretson, 1991; Leighton ve Gierl, 2007). Bilişsel taniya dayalı BOBUT eğitim çıktılarının bireysel olarak değerlendirilmesine ve öğrenme süreçlerinin daha iyi anlaşılmasına katkı sağlayabilir.

1.4.7.3. Çok Boyutlu BOBUT (Multidimensional Computerized Adaptive Testing - MCAT)

Çok boyutlu BOBUT'ta amaç bireyi aynı anda tüm boyutlarda kontrollü bir şekilde ölçmektir. Aynı anda birden fazla yetenek tahmini yürütülür, sonuçlar bireylerin madde setine verdiği yanıt örüntüleriyle ilişkilendirilir. Bu her madde için çeşitli parametreler özellikle madde ayırt edicilik parametresi tahmin edilerek yapılır (Weiss ve Kingsbury, 1984; van der Linden ve Hambleton, 1997). Örneğin Türkçe testi ile bireylerin hem dil bilgisi hem de kelime bilgisi çok boyutlu BOBUT ile ölçülebilir. Testin durdurulmasında her bir boyut için standart hata eşik değeri kullanılır. Boyutlardan biri bu eşiğe geldiğinde madde bankası o boyuta dair madde sunmaz ve diğer iki boyutta standart hata eşik değerine ulaşana kadar devam eder.

1.4.7.4. Sınıflandırmaya Dayalı BOBUT (Computerized Adaptive Classification Testing – CACT)

Sınıflandırmaya dayalı BOBUT belli bir özelliğin bireylerde var olup olmama durumunu belli ölçütler dahilinde kontrol eden ve sınıflandırma yapan bir BOBUT türüdür (Thompson ve Weiss, 2009). Genel bilinen BOBUT ile aynı mantıkta çalışarak

bireyin yetenek seviyesine göre testi bireye uyarlar. Örneğin belli bir barajın belirlendiği seçme amacıyla uygulanan testler ya da başarılı/ başarısız durumuna karar vermek için yapılan sınavlar gibi.

1.4.8. BOBUT'un Avantajları ve Sınırlılıkları

BOBUT'un birçok alanda avantajı olduğu yadsınamaz bir gerçektir. Geniş madde havuzundan bireye özgü bir test uygulanabilir. Bu esneklik sayesinde kişiler diledikleri zaman sınava girebilirler. Ülke çapında uygulanan sınavlarda belirtilen saatte sınav salonunda bulunulmadığında adaylar sınavı kaçırmış sayılır. Bu gibi kritik sınavlar bireyin uygunluk durumuna göre planlanabilir. Zamansal sınav esnekliği BOBUT'un en önemli avantajlarından biridir (Linacre, 2000; Şenel, 2017).

Kâğıt kalem testlerinde yetenek düzeyi düşük olan birey testteki tüm maddeleri yanlış cevaplayabilir; bu durumda bireyin yeteneğinin olduğu alanlarda bilgi sahibi olunamaz. Diğer taraftan, testteki tüm maddeleri doğru cevaplayan bireyin yeteneğinin olmadığı alanlarda bilgi elde edilemez. Bunun gibi uç durumlarda testten elde edilecek bilgi ciddi miktarda azalır. BOBUT, algoritması sayesinde testteki her maddeyi bireyin yeteneklerine paralel olarak uyguladığı için daha güvenilir ölçümler yapar (Wainer, 2000a). Bireyler yalnızca kendi yetenek seviyelerine uygun maddeleri yanıtladıklarından, tesadüfi hataların azaltılması sağlanır ve bireyler yeteneklerini aşan maddelerle karşılaşmadıkları için motivasyonlarını kaybetmezler. Kâğıt kalem testleri yeterlik düzeyi ortalama olan bireyler dikkate alınarak hazırlandığı için, bireyler yetenek düzeylerinin altında ve üstünde maddelerle karşılaştığında test süresi gereksiz yere uzamış olur. BOBUT gereksiz maddeleri sunmadığı için zaman kaybettirmeden bireyin yetenek düzeyi hakkında daha doğru kestirimler yapar. Aynı zamanda bireylerin gereksiz maddelerle karşılaşmaması stres düzeyini azaltabilir ve tesadüfi hataları engeller. BOBUT'un kâğıt kalem testlerine göre %50-%80 oranında daha kısa sürdüğü görülmektedir (Kezer ve Koç, 2014).

BOBUT'un sunduğu en önemli avantajlardan biri de test esnasında testi alan bireylerin puanlarının gerçek zamanlı olarak hesaplanabilmesidir. Bu durum, testi yanıtladıkları anda bireylerin kendi performansları hakkında hızlı geri bildirim almalarını sağlar (Wainer vd., 2000b). BOBUT karşılaştırılabilir puanlar da sağlar. Farklı test formlarından elde edilen sonuçların karşılaştırılabilmesi için test eşitleme

yapılması gerekmektedir. Büyük örneklem grup üzerinde yapılması gereken bu test eşitleme süreci oldukça zahmetli bir süreçtir (Kan, 2011). BOBUT, test eşitlemeye gerek duymadan mevcutta farklı maddeleri alan bireylerin puan düzeyinde aynı ölçek üzerinde karşılaştırmasını yapabilir.

BOBUT aynı zamanda test güvenliğini de sağlar. Farklı maddelerle karşılaşan bireylerde kopya eğiliminin önüne geçilmiş olur. Test materyallerinin saklanması konusunda da BOBUT daha avantajlıdır. Dijital olarak saklanan veriler, basılı materyallerle kıyaslandığında daha erişilemez konumdadır. BOBUT'un, testi alan kişiler için psikolojik yönden avantajları da vardır. Yeterlik seviyesine yakın maddelerle karşılaşan bireylerin başarıma motivasyonu canlı tutulur ve cesaretleri zedelenmez (Wainer vd., 2000b). Bireylerin istedikleri zaman sınava girebileceklerini bilmeleri, sınav kaygısının azalmasına pozitif etki eder. Engel durumu olan bireyler sınavlarda işaretleyici, okuyucu gibi işlemlerde zorlandıklarından sınav görevlilerine ihtiyaç duyarlar (Milli Eğitim Bakanlığı, 2016). Okunan maddenin anlaşılmadığı durumlarda yeniden okunmasını ya da yavaş okunmasını talep etmekte çekinebilmektedirler (Şenel ve Kutlu, 2018). BOBUT uygulamasında metin okuyucu, sesli klavye gibi teknolojiler kullanılabildiğinden yardımcı sınav görevlisine ihtiyaç duymadan sınavlarını yapabilirler.

BOBUT'un bilgisayar tabanlı olması sayesinde veriler kolaylıkla saklanabilmektedir. Bireylerin maddeyi yanıtlama süreleri kayıt altına alındığından verilen cevabın tesadüfi mi yoksa bilgiyle mi verildiği kestirilebilir (Wise ve Kingbury, 2016). Testin puanlaması teknoloji sayesinde oldukça kolaydır. Çoktan seçmeli testlerde testi alan bireyler maddeyi cevapladıkları anda geri bildirim alabilirler. Diğer formdaki testler için de geri bildirim sunmayı hedefleyen teknolojiler zamanla artmaktadır (Uysal ve Doğan, 2021). Optik formlara işaretlenen sınavlarla karşılaştırıldığında, hatalı kodlama ya da kaydırma gibi olumsuz durumlar en aza indirilmiş olur.

Yukarıda bahsedilen avantajlarının yanında BOBUT'un bazı sınırlılıkları da vardır. BOBUT'un bahsedilen güçlü yönlerinin ortaya çıkması geniş madde havuzu oluşturulmasına bağlıdır (Verschoor vd., 2019). MTK temelli güvenilir, kaliteli bir madde havuzunun oluşturulması zaman, kaynak ve emek gerektirebilir. Ayrıca

teknoloji gerekliliği sebebiyle ciddi bir maliyet doğurabilir ve teste erişim her bireyde mümkün olmadığında eşdeğer ölçmelerin yapılması zorlaşır. Son olarak BOBUT uygulamalarının ilk kez yapılacağı bir alanda bireylere verilecek uyum eğitimlerinin planlanması ve yeniliğe gösterilecek kültürel direnç de sınırlılıkları arasında değerlendirilebilir (Segall, 2004).

1.4.9. BOBUT'un Uygulama Alanları

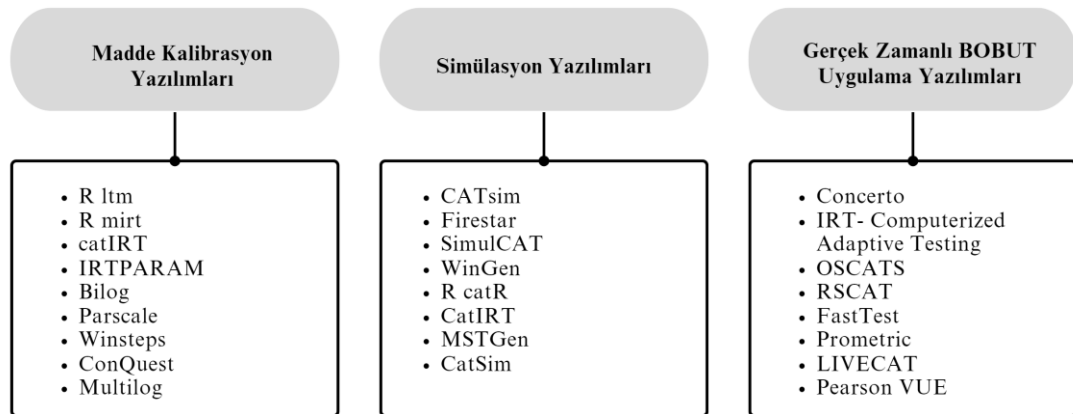
Günümüzde BOBUT, bireylerin yeteneğini verimli bir şekilde ölçebilen bir araç olarak çeşitli standart testlerde kullanılmaktadır. Bunlardan en yaygın olanları: GRE (Graduate Record Examination), lisansüstü eğitim almak isteyen bireyler için uygulanan temel yeterlikleri ölçen bir sınavdır. GMAT (Graduate Management Admission Test), bireylerin matematik, İngilizce gibi temel ön becerilerini ölçer. TOEFL IBT (Test of English as a Foreign Language, Internet-Based Test), anadili İngilizce olan ülkelerde ve eğitim dili İngilizce olan kurumlarda eğitim almak isteyen bireylerin İngilizce seviyelerinin tespitinde kullanılır. PTE Academic (Pearson Test of English Academic), bu sınavda TOEFL IBT gibi İngiltere'de çalışmak ve eğitim almak isteyen bireyler için dil becerilerinin tespitinde kullanılır. SAT (Scholastic Aptitude Test), Amerika'da lisans eğitimine kabul için yeterlik ölçen bir sınavdır. NCLEX-RN (National Council Licensure Examination for Registered Nurses), Amerika'da kullanılan bu sınav hemşirelik mesleği için işe alım süreçlerinde kullanılır. MCAT (Medical College Admission Test), tıp fakültesi kabulünde uygulanan bir sınavdır. O/A Levels İngiltere için okullardaki eğitimlerin sınındığı mezuniyet ortalamasına benzer değerlerde kabul edilen bir sınavdır (Kalender, 2009; IACAT, 2022).

Bunların dışında, BOBUT'un eğitim alanının yanı sıra özellikle sağlık ve psikoloji alanında da kullanımının yaygın olduğu görülmektedir (Gausden vd., 2018; Gibbons vd., 2016). BOBUT ölçmede deyim yerindeyse çığır açan bir gelişme olarak görülmektedir. Kullanım alanları ve yaygınlaşmasının da önümüzdeki yıllarda yükselen bir hızla artması beklenmektedir.

1.4.10. BOBUT Yazılımları

Bireye uyarlanmış test geliştirme ve uygulama sürecinde belli yazılımlardan faydalanılmaktadır. Testte yer alan maddelerin uygun zorluk seviyelerine sahip olması yapılacak istatistiksel analizler için önemlidir. Testin geçerli ve güvenilir bir ölçme

yapabilmesi için madde parametrelerinin testten önce belirlenmesi gerekir, bu işleme madde kalibrasyonu denir. MTK temelli geliştirilen BOBUT'un madde kalibrasyonu için kullanılan bazı yazılımlar mevcuttur. Tasarlanan bir BOBUT için uygulama öncesi testin geçerliğinin sınanması, maddelerin ölçülen özellik bakımından uygunluğunun denetlenmesi ve test modelinin oluşturulması gerekmektedir. Bu noktada simülasyon yazılımlarıyla BOBUT test edilebilir. Bu yazılımlar gerçek test ortamına benzer ölçümler yapılmasına olanak tanır. Choi (2009), en uygun BOBUT ayarlamalarının yapılması ve testten en yüksek verimi alabilmek için simülasyon çalışmalarından destek alınması gerekliliğini vurgulamıştır. Simülasyon çalışmalarında da belli çeşitlilikler söz konusudur; gerçek veriye dayalı yapılan post-hoc simülasyonları, verinin simüle edildiği Monte Carlo simülasyonu ve gerçek veri ile simüle verinin beraber kullanıldığı karma (hibrit) simülasyon yöntemleri kullanılmaktadır. Son olarak BOBUT yazılımlarıyla gerçek uygulamalar yapılabilir. Yukarıda bahsedilen yazılımlara örnekler Şekil 1.8'de sunulmuştur.



Şekil 1.8. BOBUT Uygulamalarında Kullanılan Yazılımlar

Madde kalibrasyon yazılımları testte yer alan maddelerin parametrelerinin tahmin edilmesini sağlar ve testin güvenilirliğini arttırmak için geliştirilmiştir. IRTPARAM (Park, 2011) tek ve çok parametrelili lojistik modellerin parametre tahminlerinde kullanılır. Bilog (Mislevy ve Bock 1982), eğitim ve psikolojik BOBUT uygulamalarının test geliştirme ve madde parametrelerinin analizinde kullanılır. Parscale (Hambleton ve Swaminathan 1983), özellikle derecelendirilmiş yanıtlarda kullanılan test biçimi için uygundur, psikometrik analizler yaparken tercih edilir.

Winsteps (Linacre, 1998), sosyal bilimler araştırmalarında tercih edilir, maddelerin detaylı analizinde ve test yapısının genel incelemesinde kullanılır. ConQuest (Wu vd., 1997), geniş ölçekli grup değerlendirmelerinde ve eğitim araştırmalarında kullanılır. Farklılaştırılmış madde fonksiyonu (differential item functioning; DIF) analizi yapabilir ve çoklu grupların değerlendirilmesinde tercih edilir. Multilog (Thissen, 1986), psikometrik testlerde sıklıkla kullanılır, kategorik yanıtlarda kolaylık sağlar. Bunların yanında R programlama dilinde geliştirilen MTK paketleri de test maddelerinin analiz edilmesi ve kalibre edilmesinde kullanılabilmektedir. R paketleri açık kaynak kodlu olması ve istatistiksel hesaplamalar için uygun olması sebebiyle sıklıkla tercih edilmektedir. R ltm (Rizopoulos, 2006) paketi tek boyutlu modellerin bazılarında kullanılabilmektedir. R mirt (Chalmers, 2012) paketi hem tek boyutlu hem de çok boyutlu modellerde uygulanabilir. CatIRT (Bischof, 2009), MTK temelli BOBUT simülasyonlarında ve verilerin analizinde kullanılır. Uyarlanabilir algoritmalar ve madde seçim yöntemleri sunar.

Simülasyon yazılımlarında R temelli catR (Magis ve Raiche, 2012) yazılımı BOBUT senaryoları için oldukça fazla seçenek barındırdığından diğer simülasyon yazılımlarına göre daha sıklıkla tercih edilmektedir. Aybek (2021), CatIRT (Nydick ve Wiberg, 2012) için carR ve mirt paketlerini kullanarak kodlama yapmadan da arayüz üzerinden BOBUT çalışmaları yapılabileceğini belirtmektedir. CATsim (Meyer, 2010), farklı test stratejilerinin simülasyonunda kullanılır, test performansını analiz edebilir. Firestar (He, 2011), BOBUT için MTK temelli simülasyonlar sağlar. SimulCAT (UCLA Center for the Study of Evaluation, 1980), özellikle eğitim amaçlı BOBUT uygulamalarında sistem simülasyonları sağlar. WinGen (Kolen ve Brennan, 1995), MTK tabanlı veri üretmek ve test geliştirme süreçlerinde kullanmak için tercih edilir. MSTGen (Magis vd., 2010), çok aşamalı test simülasyonlarında, testin tasarımı ve performansının incelenmesinde kullanılır. CatSim (Bischof, 2009), BOBUT algoritmalarının etkililiğini ve test performansını değerlendirmede kolaylık sağlar.

Gerçek zamanlı BOBUT uygulama yazılımlarında açık kaynak kodlu yazılımlar ücretsiz olması sebebiyle öne çıkmaktadır. Bunlardan bazıları, Cambridge Üniversitesi tarafından geliştirilen Concerto (University of Cambridge, 2008), 1PL ve 3PL modeller için kullanılan IRT – Computerized Adaptive Testing ve diğeri OSCATS'tur (Muthen ve Asparouhov, 2012). CONCERTO, CATR ile entegre

çalışarak, testlerin yönetimi ve analizi için gerekli altyapıyı sağlar. OSCATS kütüphane olarak GObject kullanır ve nesne yönelimli C’de oluşturulmuştur. RSCAT (Zhang, 2010), R dilinde oluşturulmuştur, test sürecinde anlık madde seçimi yapabilir. FastTest (Assessment Systems Corporation, 2004), test geliştirme sürecinde kullanıcı dostu bir arayüz sunar, bireylerin yanıtlarına göre anında uyarlanabilir bir yapısı vardır. Prometric (Prometric Inc., 1994), geniş ölçekli BOBUT uygulamalarında tercih edilir. Profesyonel sınavlarda dünya çapında birçok kurum tarafında kullanılmaktadır. LIVECAT (Pearson VUE, 2002), canlı BOBUT uygulamalarında tercih edilir, bireylerin performansına uygun olarak anlık uyarlanabilir. Pearson VUE (Pearson Education, 1994), geniş çaplı test uygulamalarında güvenli bir test ortamı sağlar.



2. ÖNCEKİ ÇALIŞMALAR

BOBUT'un ölçme ve değerlendirme alanındaki gelişimi her geçen gün artarak devam etmektedir. Buna paralel olarak bu konuda yapılan akademik çalışmalara da eğilim artmaktadır. BOBUT konulu araştırmalar yurt dışında yoğunlukta olsa da son dönemlerde Türkiye'de yapılan çalışmaların sayısının arttığı da ifade edilebilir. Ancak alan yazında sistematik derleme çalışmaları oldukça sınırlıdır. Bu kapsamda ilgili çalışmalar üç başlıkta incelenmiştir. Literatür incelemesinin ilk bölümünde yurt içinde BOBUT ve kâğıt kalem testlerinin karşılaştırıldığı çalışmalarına, ikinci bölümünde yurt dışında yapılan karşılaştırma çalışmalarına ve üçüncü bölümünde BOBUT konulu yapılmış sistematik derleme, bibliyometrik analiz ve meta-analiz çalışmalarına yer verilmiştir.

2.1. BOBUT Konulu Yurt İçinde Yapılan Çalışmalar

BOBUT üzerine Türkiye'de yapılan araştırmalar, genellikle klasik kâğıt kalem testleri ile BOBUT arasındaki karşılaştırmalar üzerine odaklanmıştır. Bu bölümde de BOBUT'un etkililiğine genel bir bakış sunmak için gerçek zamanlı BOBUT uygulamasının yapıldığı ve kâğıt kalem testleriyle BOBUT'un karşılaştırıldığı çalışmalara yer verilmiştir. Bu çalışmaların büyük çoğunluğu, BOBUT'un madde sayısı ve zaman kullanımı açısından avantajlı olduğunu, ayrıca yetenek kestiriminde güvenilir sonuçlar verdiğini ortaya koymuştur.

İlk çalışma, Kaptan (1993) tarafından gerçekleştirilmiş olup, sayısal yetenek kestirimi açısından BOBUT ve klasik testleri karşılaştırılmıştır. Bu çalışma, BOBUT'un madde sayısı ve zaman kullanımı açısından avantajlı olduğunu ortaya koymuştur. Benzer şekilde, Cömert (2008) tarafından yapılan yüksek lisans tezi, BOBUT'un zamandan tasarruf sağladığını ve gereksiz maddelerle karşılaşmayı engellediğini vurgulamıştır.

Kalender (2011) ve Bulut ve Kan (2012) gibi araştırmacılar, BOBUT ile kâğıt kalem testlerini çeşitli yetenek kestirim yöntemleri ve test sonlandırma kuralları açısından karşılaştırmış ve BOBUT'un daha az madde ile güvenilir sonuçlar verdiğini raporlamıştır. Kezer (2013) de benzer bulgular elde etmiş ve BOBUT'un madde sayısından ciddi tasarruf sağladığını belirtmiştir.

Özbaşı (2014), bilgisayar okuryazarlık testi üzerinden BOBUT ve kâğıt kalem testlerini karşılaştırmış ve BOBUT'un güvenilirlik açısından avantajlı olduğunu göstermiştir. Ayrıca, Özbaşı (2016) tarafından yapılan bir diğer araştırma, öğrencilerin BOBUT'a yönelik olumlu görüşlerini ortaya koymuştur.

Son yıllarda yapılan çalışmalardan Arslan (2022) ve Kaya (2022), COVID-19 pandemisiyle birlikte çevrimiçi sınavların önem kazandığını ve BOBUT'un bu bağlamda avantajlı olduğunu vurgulamıştır. Özdemir ve Gelbal (2022) ise çok aşamalı BOBUT (ca-MST) ile kâğıt kalem testlerini karşılaştırmış ve ca-MST'nin test uzunluğunu azaltırken daha hassas yetenek ölçümleri sağladığını belirtmiştir.

Gökcan (2023) ve Demir (2023), BOBUT'un yetenek kestiriminde yüksek düzeyde ilişki ve geçerlilik sağladığını raporlamışlardır. Gökcan, İngilizce kelime bilgisi testinde BOBUT'un az madde kullanarak yetenek kestirimi yapabildiğini gösterirken; Demir, mesleki olgunluk envanteri üzerinden BOBUT'un geçerliliğini test etmiş ve olumlu sonuçlar elde etmiştir.

Bu genel bulgular, BOBUT'un klasik kâğıt kalem testlerine kıyasla çeşitli avantajlar sunduğunu ve güvenilir bir ölçme yöntemi olarak kullanılabileceğini göstermektedir. 2010 yılına kadar, BOBUT konulu araştırmalar sınırlı sayıda olup, genellikle gerçek zamanlı BOBUT uygulamalarına odaklanmıştır. Bu dönemde gerçekleştirilen çalışmalar, testlerin geçerliliği ve uygulama yöntemleri üzerine sınırlı değerlendirmeler sunmuş ve bilgi birikimini belirli bir seviyede tutmuştur. Ancak, 2010 sonrasında BOBUT ile ilgili yapılan çalışmaların sayısında belirgin bir artış gözlemlenmiştir. Bu dönemdeki çalışmalar, ağırlıklı olarak simülasyon tabanlı yaklaşımlar üzerine yoğunlaşmış ve BOBUT'un çeşitli uygulama ve madde havuzu stratejilerinin etkinliğini değerlendirme açısından yeni perspektifler sunarak alandaki literatürü önemli ölçüde zenginleştirmiştir.

2.2. BOBUT Konulu Yurt Dışında Yapılan Çalışmalar

BOBUT üzerine yurt dışında yapılmış çok sayıda çalışma bulunmaktadır. ABD, Kanada, Almanya, İngiltere, Avustralya ve Japonya gibi ülkeler, bu alandaki önde gelen ülkelerdir. Uluslararası literatürde, teknolojik yenilikler ve makine öğrenmesi gibi modern yaklaşımlar BOBUT sistemlerine entegre edilmektedir. Ancak,

bu çalışmaların tamamını burada sunmak mümkün olmadığından, BOBUT'un genel durumunu yansıtan ve karşılaştırma sağlayan çalışmalar kısaca özetlenmiştir.

Yurt dışında BOBUT üzerine yapılan çalışmalar, Türkiye'deki araştırma sonuçlarına paraleldir. Çalışmalar BOBUT'un çeşitli avantajlarını ve performanslarını inceleyen önemli bulgular sunmaktadır. Kingsbury ve Houser (1988), ilkökul sınav sistemi bağlamında BOBUT ile iki farklı kâğıt kalem testini karşılaştırmış, BOBUT'un madde sayısı ve zaman kullanımı açısından avantajlı olduğunu belirtmiştir. Sonraki çalışmalarında Kingsbury ve Hauser (2004), BOBUT'un kâğıt kalem testlerine göre yetenek ölçümünde daha ayrıntılı bilgi sunduğunu ve daha yüksek ölçüm güvenirliliği sağladığını vurgulamıştır. Moshinsky ve Kazin (2005) İsrail'de özel gereksinimleri olan öğrenciler için geliştirdikleri BOBUT yazılımının, kâğıt kalem testleriyle kıyaslandığında daha güvenilir sonuçlar verdiğini belirtmiştir. Ita (2014), BOBUT ve kâğıt kalem testlerinin öğrenci performansı üzerindeki etkilerini karşılaştırmış, her iki test arasında anlamlı bir performans farkı bulamamış ve öğrencilerin kâğıt kalem testlerini tercih ettiğini ortaya koymuştur. Güney Kore'de Joeng (2014) tarafından yapılan çalışmada, teknoloji kullanımının yüksek olduğu bir ortamda BOBUT'tan beklenen yüksek puanların aksine, öğrencilerin kâğıt kalem testinden daha yüksek puan aldıkları görülmüştür. Ling (2017) ise BOBUT'un öğrencilerin motivasyonuna etkilerini araştırmış, BOBUT uygulamasının daha düşük kaygı ve yüksek katılım sağladığını vurgularken test performanslarında anlamlı bir farklılık oluşturmadığını belirtmiştir. Sangmeister (2017) araştırmasında, mesleki eğitim stajyerlerinin yeterliliklerini test eden BOBUT'un hem içerik hem de ölçme formatı açısından geçerli ve güvenilir olduğunu, motivasyon üzerinde ise etkili olmadığını ortaya koymuştur. Bu genel değerlendirmeler, BOBUT'un kâğıt kalem testlerine kıyasla sağladığı çeşitli avantajları ve uygulama farklılıklarını yansıtmaktadır. Literatürde bu tür karşılaştırma çalışmalarının yanı sıra, BOBUT'un diğer yönlerini ele alan birçok araştırma da mevcuttur (Chen vd., 2020; Embretson ve Reise, 2000; Hambleton vd., 1991b; Lord, 1980; Sahin ve Gelbal, 2020; Tapan ve Kılıç, 2020; van der Linden ve Glas, 2010; Weiss ve Kingsbury, 1984).

2.3. BOBUT Konulu Yapılan Önceki Derlemeler, Bibliyometrik Analizler ve Meta-Analiz Çalışmaları

BOBUT üzerine yapılan önceki derlemeler, bibliyometrik analizler ve meta-analizler, BOBUT sistemlerinin etkinliğini, kullanım alanlarını ve metodolojik gelişmelerini kapsamlı bir şekilde inceleyerek alandaki bilgi birikimini artırmıştır. Bu çalışmalar, BOBUT'un gelişim sürecini ve mevcut literatürdeki eğilimleri anlamak için temel bir bakış açısı sunmaktadır. Uluslararası literatürde bu alanda yapılan çalışmaların sayısı fazlayken, Türkiye'de bu konu üzerine yapılmış sınırlı sayıda çalışma bulunmaktadır. Aşağıda, bu çalışmalardan elde edilen bulgular kısaca özetlenmiştir.

Georgiadou ve arkadaşlarının (2007) çalışması, 1983-2005 yılları arasındaki BOBUT sistemlerini ele alarak, madde kullanım sıklığı kontrol stratejilerini beş ana kategoriye ayırmış ve her bir stratejinin avantajları ile dezavantajlarını detaylandırmıştır. Bu araştırma, BOBUT sistemlerinin gelişim sürecine dair geniş bir perspektif sunmaktadır. Walker ve arkadaşlarının (2010) sistematik derlemesi ise, tıbbi veri tabanlarında yapılan 32 çalışmayı inceleyerek, BOBUT'un sağlık alanında geleneksel yöntemlere göre daha kesin tahminler sağladığını ve kişiselleştirilmiş test deneyimi sunduğunu ortaya koymuştur.

Wang ve Ding (2010), Çin'deki BOBUT uygulamalarını genel olarak değerlendirerek, çoğu araştırmanın simülasyon verileri kullandığını ve BOBUT'un sağlık alanında eğitimden daha yaygın olarak kullanıldığını belirtmiştir. Ayrıca, kâğıt kalem testleri ile BOBUT'un benzer ölçümler sağladığını, ancak uygulama ve geçerlilik açısından farklılıklar gösterdiğini ifade etmiştir. Sepenassato ve arkadaşları (2015) ise, 2014 yılına kadar yapılmış 182 makaleyi derinlemesine inceleyerek BOBUT'un teknik ve metodolojik yönlerini kapsamlı bir şekilde ele almış, en çok kullanılan modeller ve yazılımlar hakkında detaylı bilgiler sunmuştur.

Türkiye'de yapılan bibliyometrik analiz ve meta-analiz çalışmaları, BOBUT konusundaki çalışmaların sayısının zamanla arttığını göstermiştir. Yurtçu ve Güzeller (2021), dünya genelinde 1946-2016 yılları arasındaki 637 makaleyi inceleyerek, BOBUT çalışmalarında özellikle sağlık alanında belirgin bir yoğunlaşma gözlemlemiş ve 2010 yılından itibaren çalışmalardaki artışı vurgulamıştır. Nalbantoğlu Yılmaz

(2021) tarafından yapılan meta-analiz çalışmasında ise, Türkiye’de 1993-2020 yılları arasında BOBUT ve kâğıt kalem testlerini karşılaştıran 37 çalışmanın bulguları değerlendirilmiş, BOBUT’un kâğıt kalem testlerine iyi bir alternatif olduğu sonucuna varılmıştır.

Frey ve arkadaşları (2024) ise, dünya genelinde 27 çalışmayı meta-analiz yöntemiyle inceleyerek, BOBUT’un motivasyon ve duygusal etkilerini değerlendirmiş ve olumsuz duygular üzerindeki etkisinin anlamlı olduğunu vurguarken motivasyon ve olumlu duygular üzerindeki etkisinin anlamlı olmadığını belirtmiştir.

Yukarıdaki çalışmalar incelendiğinde ülkemizde yapılan BOBUT konulu çalışmalar için bir sistematik derleme eksikliği vardır. Hammad ve Hallinger’e (2017) göre sistematik derleme bir alanda yapılan tüm çalışmaların belli kriterler gözetilerek detaylı olarak incelenmesi ve erişilen bulguların sentezlenmesidir. Bu araştırmayla Türkiye’de BOBUT uygulaması yapan tüm çalışmalara erişerek geniş açıdan konuyu değerlendirmek ve sistematik derleme eksikliğinin giderilmesi amaçlanmaktadır.

3. GEREÇ VE YÖNTEM

3.1. Araştırmanın Deseni

Bu çalışmada BOBUT konusunda Türkiye’de yapılan araştırmaları sistemli bir şekilde bir araya getirip mevcut durumu incelemek amacıyla sistematik derleme yöntemi kullanılmıştır. Sistematik derleme bir araştırma konusuna yönelik oluşan soruları yanıtlamak için ilgili konuda yapılan tüm araştırmaları kapsamlı bir şekilde organize ederek elde edilen bulguların bir araya getirilme sürecidir (Bellibaş ve Gümüş, 2018; Hammad ve Hallinger, 2017; Higgins ve Green, 2011). Sosyal bilimler alanında bir araştırma konusunun gelişiminin incelenmesinde literatür incelemesi oldukça yaygın kullanılan bir araştırma yöntemidir (Norris ve Ortega, 2006). Bu çalışmada içerik analiz tekniği kullanılmıştır. Araştırmanın amacı göz önüne alındığında içerik analizi tekniklerinden olan betimsel içerik analizi en uygun yöntem olarak belirlenmiştir. Bu analiz türü, bir alanda yapılan araştırmaları amaç, yöntem, örneklem büyüklüğü, araştırma yılı ve ülkesi gibi kriterler doğrultusunda sistematik ve objektif olarak inceleyerek araştırmaların yoğunlaştığı alanları tespit etmeyi amaçlar.

Araştırmanın objektifliğini ve tekrarlanabilirliğini sağlamak için PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) yönerge ve kontrol listesi rehber alınmıştır. PRISMA, sistematik derleme çalışmalarının şeffaflığını ve güvenilirliğini arttırmayı amaçlamaktadır (Moher, vd., 2009). Bu çerçevede yapılan çalışma aşağıdaki aşamalardan oluşmuştur:

3.2. Veri Toplama Süreci

İlgili çalışmaların tespiti amacıyla kapsamlı bir literatür incelemesi yapılmıştır. Literatür taramasında konuya özgü anahtar kelimeler belirlenmiştir. Bunlar:

- Bilgisayar Ortamında Bireye Uyarlanmış Test
- Bilgisayarlı Bireyselleştirilmiş Test
- BOBUT
- Computerized Adaptive Testing
- CAT
- Bireye Uyarlanmış Test

- Bireyselleştirilmiş Test
- Bilgisayar Ortamında Test

Bu anahtar kelimeler AND ve OR gibi operatörler kullanılarak optimize edilmiş ve araştırma alanı daraltılmıştır. Bu anahtar kelimelerle 5 farklı veri tabanı ve bir arama motorunda arama yapılmıştır. Bunlar:

- Google Akademik
- YÖKTez
- DergiPark
- TRDizin
- Web of Science
- ERIC

Çalışmanın literatür taraması 25.10.2023 – 11.03.2024 tarih aralığında yapılarak toplam 739 çalışmaya erişilmiştir. Kayıtlar Microsoft Excel programında tutularak incelenen makaleler veri kaynağına göre ayrı klasörlerde saklanmıştır. Literatür tarama süreci aşağıdaki adımlardan oluşmuştur.

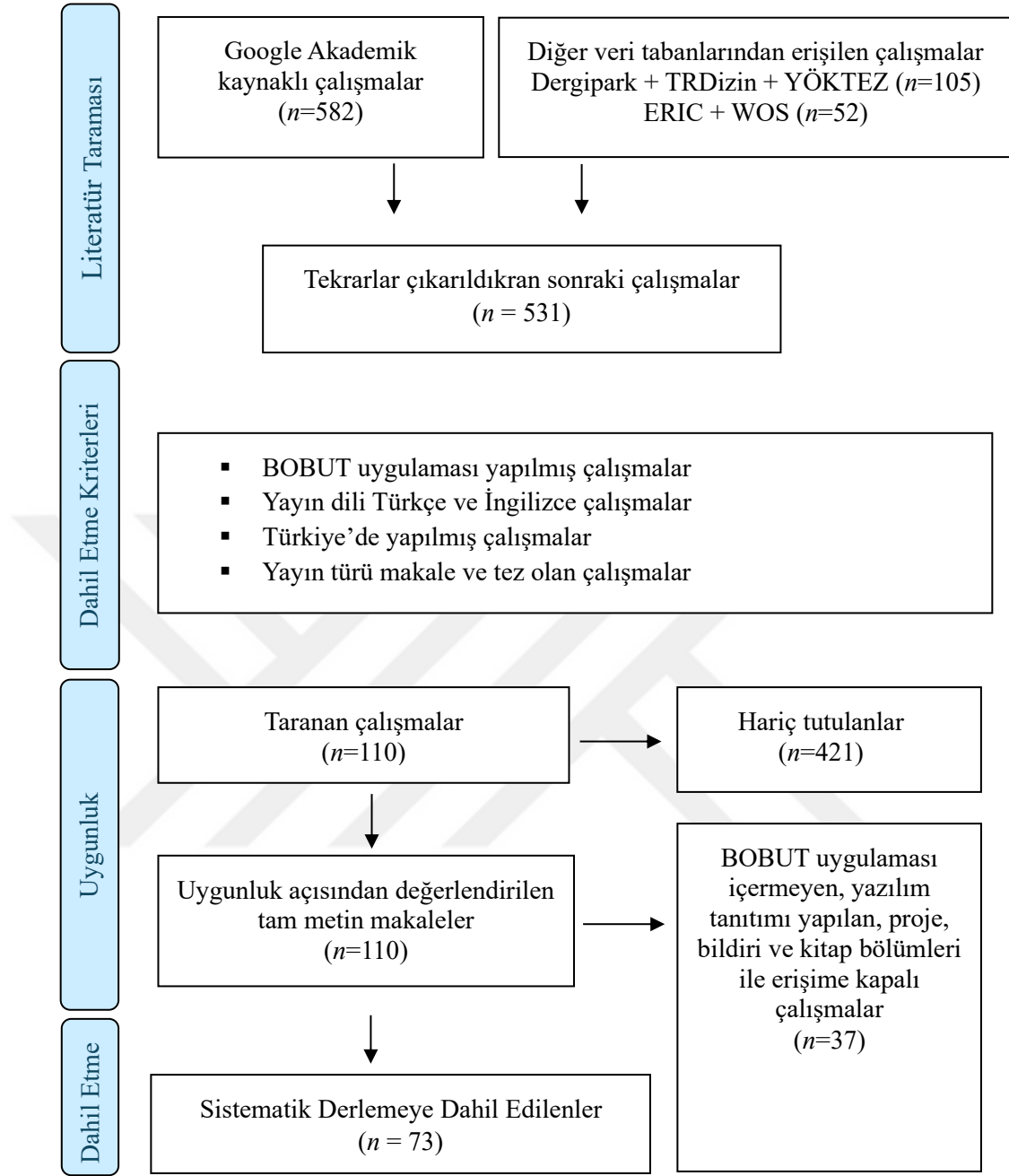
1. Tarama: Bu taramayla BOBUT konusunda yapılan tüm çalışmalara çeşitli veri tabanlarından anahtar kelimeler kullanılarak erişilmiştir.
2. Eleme: Erişilen araştırmalar başlık, içerik, ülke, çalışıldığı alan gibi konularda incelenmiş belli kriterlere göre uygun olmayanlar çıkarılmıştır.
3. Uygunluk: Son durumda kalan araştırmalar tam metin olarak incelenmiş dâhil etme kriterine uygun olup olmama durumlarına karar verilmiştir.
4. Dâhil Etme: Nihai olarak incelemeye dâhil olan çalışmaların bulguları özetlenerek genel değerlendirme yapılmıştır.

Bu adımlar dahilinde incelenen çalışmalar yüzde ve frekans bilgileriyle Çizelge 3.1’de özetlenmektedir.

Çizelge 3.1. Literatür Taraması Süreci

Aşama	Açıklama	f	Toplam
Literatür Taraması	Google Akademik kaynaklı	582	739
	Diğer veri tabanları	157	
Hariç Tutulanlar	Tekrarlı çalışma	208	666
	Kapsam dışı	421	
	BOBUT uygulaması içermeyen	37	
Analize Dahil Edilenler		73	73

Çizelge 3.1’de görüleceği üzere tüm veri kaynaklarından 739 çalışmaya erişilmiştir. Bu sayı anahtar kelimelere göre arama yapıldığından konu dışı olan tüm çalışmaları ve farklı kaynaklardan erişilen tekrarlanan çalışmaları da kapsamaktadır. İlk olarak çalışmaların kaydedildiği veri dosyasında inceleme yapılmış ve 208 çalışmanın tekrarlandığı tespit edilerek analizden çıkarılmıştır. Geriye kalan 531 çalışma başlık ve özet incelemesi yapılarak farklı konuları ele alan ya da araştırmanın odağından uzaklaşarak kapsam dışına çıkan 421 çalışma derlemeye dahil edilmemiştir. Bu aşama araştırmanın ana problemini ele almak adına önemlidir. Kalan 110 çalışma detaylı tam metin olarak okunmuş 37’si BOBUT uygulaması içermediğinden, YÖKTez’de erişime kapalı tez ya da ücretli erişim istendiğinden, kitap bölümü, bildiri özetleri ya da proje konulu olduğundan çalışma dışında tutulmuştur. Son durumda çalışmada 73 çalışma analiz edilmiştir. Literatür tarama süreci ve iş akışı Şekil 3.1’de PRISMA diyagramıyla görselleştirilmiştir.



Şekil 3.1. PRISMA Akış Süreci

Araştırmanın literatür taraması sürecinde erişilebilen altı kaynaktan tarama yapılmıştır. Belirlenen kriterlere göre uygun bulunan 73 çalışmanın kaynaklara göre dağılımı Çizelge 3.2’de verilmiştir.

Çizelge 3.2. Dahil Edilen Çalışmaların Kaynakları

	Örneklem	
	f	%
Google Akademik	30	41
YÖKTez	26	36
DergiPark	6	8
TRDizin	6	8
Web of Science	5	7
Toplam	73	100

Çizelge 3.2'ye göre Google Akademik en çok çalışmanın erişildiği kaynak olmuştur öyle ki çalışmaların 30'una (%41) buradan erişilmiştir. Türkiye'de yapılan tez çalışmalarının tarandığı YÖKTez'de araştırmanın kriterlerine uygun 26 (%36) çalışma bulunmuştur. Türkiye'de akademik çalışmaların yayınlandığı DergiPark ve TRDizin de araştırma sürecinde kullanılan önemli kaynaklardır ve bu veritabanlarından 12 (%16) çalışmaya ulaşılmıştır. Uluslararası kaynak taramasında kullanılan Eric kaynağında araştırma kriterlerini karşılayan kaynağa erişilemezken bir diğer saygın kaynak Web of Science'ta 5 (%7) çalışma tespit edilmiştir.

3.3. Dâhil Etme ve Hariç Tutma Kriteri

Literatür taraması sonucu elde edilen çalışmaları incelemek için belirli dâhil etme ve hariç tutma kriterleri oluşturulmuştur. Araştırmanın amacına yönelik belirlenen dâhil etme kriterleri:

- BOBUT uygulaması yapılmış çalışmalar
- Yayın dili Türkçe ve İngilizce çalışmalar
- Türkiye'de yapılmış çalışmalar
- Yayın türü makale ve tez olan çalışmaları kapsamaktadır.

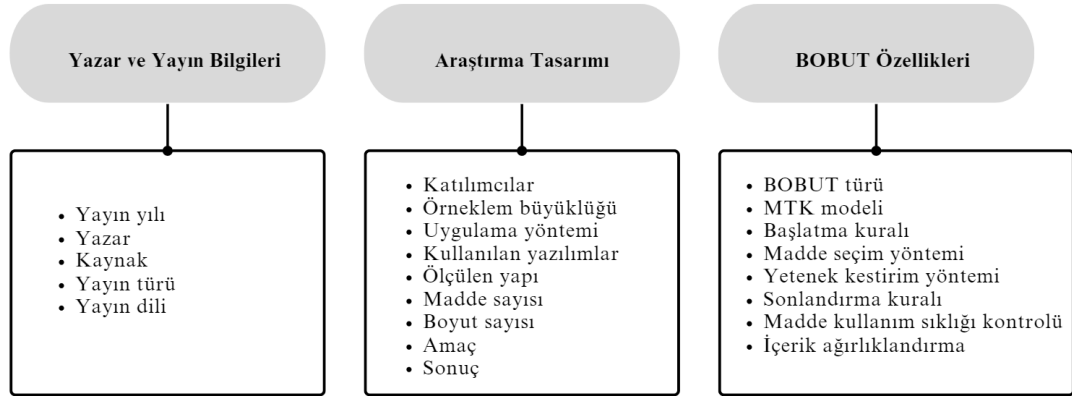
Araştırmanın amacına hizmet etmeyen çalışmalar hariç tutma kriterleri kullanılarak çalışmaya dâhil edilmemiştir. Bu hariç tutma kriterleri:

- Tam metin erişimi sağlanamayan çalışmalar
- Yazılım tanıtımı ya da genel BOBUT anlatımı yapan çalışmalar

- Türkiye dışında yapılan çalışmalar
- Tez ve makale dışında olan yayın türleri
- Tekrarlanan çalışmalar
- BOBUT uygulaması içermeyen çalışmalar olarak belirlenmiştir.

3.4. Veri Dosyasının Oluşturulması ve Kodlama Güvenirliği

Belirlenen anahtar kelimeler doğrultusunda yapılan literatür taramasıyla uygun bulunan çalışmalar pdf dosya formatında yayın yılı ve yazar ismi eklenerek bilgisayarda oluşturulan klasörlere kaydedilmiştir. Ana klasör altında alt klasörler oluşturularak erişilen kaynaklara göre bu klasörler isimlendirilmiş ve çalışma hangi veri tabanından elde edildiye o klasörde saklanmıştır. Ardından bu çalışmalar dahil etme ve hariç tutma kriterlerine göre ayıklanarak dahil edilenler kodlama dosyası aracılığıyla Microsof Excel (Microsoft Corporation, 2024) programında analiz edilmiştir. İlk adımda tablo hazırlama ve belirlenen değişkenlerin kontrolü konusunda ölçme ve değerlendirme alan uzmanı görüşü alınmıştır. Kodlama dosyasında yer almasına karar verilen değişkenler üst başlıklar halinde sınıflandırılarak Şekil 3.2’de verilmiştir.



Şekil 3.2. Kodlama Değişkenleri

Her bir değişken için oluşturulma nedeni ve kodlama süreci detayları aşağıda belirtilmiştir:

“Türkiye’de BOBUT kullanan çalışmalar hangi yıllarda yapılmıştır?” alt problemine yönelik Türkiye’de yapılan BOBUT çalışmalarının zaman içindeki

gelişimini takip etmek amacıyla yayın yılı değişkeni oluşturulmuştur. Yapılan çalışmaların yayın tarihleri yıl bilgisi şeklinde girilmiştir.

“Çalışmaların yayın türleri nelerdir?” alt problemine yönelik çalışmanın bilimsel geçerliğini sağlamak ve yayınlandığı platformların dağılımlarını incelemek amacıyla yayın türü değişkeni oluşturulmuştur. Burada araştırma kapsamını daraltmak için sadece tez ve makale türüne yer verilmiştir. Yayın türleri kodlanırken, tekrarlı çalışmalarda tezden makaleye çevrilmiş yayınlara rastlanmış ve ilk kaynak olarak tezler kaynak olarak çalışmaya dahil edilmiştir.

“Çalışmalar yoğunlukta hangi amaçla yapılmıştır?” alt problemine yönelik çalışmanın ana amacını belirtmek ve özet bilgi sunmak için çalışmanın amacı değişkeni oluşturulmuştur. Bu değişken kodlanırken çalışmaların özet kısmından faydalanılmış ve kısaca bilgi verilmiştir.

“Çalışmaların ne kadarı Türkçe ve ne kadarı İngilizce yayımlanmıştır?” alt problemine yönelik yayın dili değişkeni oluşturulmuştur. Bu değişken ile çalışmaların dil bağlamında incelemesi yapılmış ve veriler her bir çalışma için veri dosyasına kodlanmıştır.

“Çalışmalarda hangi örneklem çeşidi yoğunluktadır?” alt problemine yönelik BOBUT uygulamalarının Türkiye’de yapılan çalışmalarda hedef kitlesini ve çalışma alanını belirlemek için katılımcılar değişkeni oluşturulmuştur. İncelenen yayınlarda veriler kişi ya da simülasyon verisi olacak şekilde gruplandırılmıştır.

“Çalışmalarda farklı BOBUT türleri kullanım oranı nedir?” alt problemine yönelik BOBUT türü değişkeni oluşturulmuştur. Kullanım amacı doğrultusunda değişen BOBUT türleri (ca-MST, CD-CAT, MCAT) kısaltmalar kullanılarak kodlanmıştır.

“Canlı BOBUT uygulaması ve simülasyon verisi kullanım oranı nedir?” alt problemine yönelik çalışmaların yürütüldüğü test ortamlarını tanımlamak için uygulama yöntemi değişkeni oluşturulmuştur. Simülasyon ya da canlı BOBUT olarak uygulanan çalışmalar kategorize edilerek dosyaya işlenmiştir.

“BOBUT geliştirme ve uygulama sürecinde en sık kullanılan yazılımlar hangileridir?” alt problemine yönelik yaygın kullanılan yazılımları ve analiz

platformlarını tespit etmek için kullanılan yazılımlar değişkeni oluşturulmuştur. Yazılımlar kalibrasyon, analiz, BOBUT uygulama yazılımları olarak ayrılmadan aynı sütunda beraber kodlanmıştır.

“Çalışmaların örneklem büyüklükleri ne kadar?” alt problemine yönelik çalışmaların istatistiksel gücünü görmek amacıyla örneklem büyüklüğü değişkeni oluşturulmuştur. Burada madde bankası geliştirme sürecinde katılan örneklem grubu dikkate alınmamış sadece BOBUT uygulamasına katılan örneklem büyüklükleri simülasyon ve canlı BOBUT olacak şekilde ayrı ayrı veri dosyasına işlenmiştir.

“Çalışmalarda kullanılan madde havuzu genişliği ne kadardır?” alt problemine yönelik madde sayısı değişkeni oluşturulmuştur. BOBUT uygulamalarında madde havuzunun genişliği testin işlevi açısından önemli görülmektedir. İncelenen çalışmalarda simülasyon ve canlı BOBUT uygulamasında yer alan madde sayıları ayrı ayrı kodlanmıştır.

“BOBUT çalışmalarında ölçülen yapıların çeşitliliği nasıldır?” alt problemine yönelik çalışmaların yoğunlaştığı alanları tespit etmek amacıyla ölçülen yapı değişkeni oluşturulmuştur. Yapılar detaylı konulara yer vermeden matematik, İngilizce, mesleki bilgi şeklinde genel olarak sınıflandırılarak kodlanmıştır.

“Ölçülen yapıların tek boyutlu ve çok boyutlu olma durumları nedir?” alt problemine boyut sayısını belirlemeye yönelik boyut sayısı değişkeni oluşturulmuştur. Birden fazla olan boyutlar çok boyutlu şekilde veri dosyasına kodlanmıştır.

“En yaygın tercih edilen MTK modelleri hangileridir?” alt problemine yönelik madde seçimi ve test modelini belirlemek amacıyla MTK modeli değişkeni oluşturulmuştur. Veriler incelenirken en sık kullanılan modeller analizlerde değerlendirilmiştir.

“Testlerde en yaygın kullanılan madde seçim yöntemi, madde kestirim yöntemi, yetenek kestirim yöntemi, testi başlatma kuralı ve test sonlandırma kuralları nelerdir?” alt problemine yönelik test algoritmalarında en sık kullanılan yöntemleri tespit etmek amacıyla madde seçim yöntemi, test başlatma kuralı, yetenek kestirim yöntemi ve durdurma kuralı değişkenleri oluşturulmuştur. Veriler kodlanırken

çalışmalarda farklı isimlendirilen yöntemler tespit edilirken zorluk yaşanmıştır. Birden fazla yöntem kullanan çalışmalar arada virgül olacak şekilde yan yana kodlanmıştır.

“Çalışmalarda madde kullanım sıklığı kontrolü ve içerik dengeleme yapılma oranları nedir?” alt problemine yönelik madde kullanım sıklığı ve içerik dengeleme değişkeni oluşturulmuştur. Evet ve hayır şeklinde yapılan kodlamada yöntemlerin isimlerine yer verilmemiştir.

“Çalışmaların genel sonuçları nedir?” alt problemine yönelik BOBUT uygulamalarının sonuçlarını özetlemek adına çalışmanın sonucu değişkeni oluşturulmuştur. Yayınların sonuç ve tartışma kısmından özet ve kısa bilgilere yer verilmiştir.

Veri analiz sürecinde güvenilir bulgular elde etmek, çalışmanın geçerliğini ve güvenilirliğini sağlamak için büyük öneme sahiptir (Lincoln ve Guba, 1985). Bu araştırmada güvenilirliği sağlamak için Miles ve Huberman (1994) tarafından geliştirilen formül kullanılmıştır. Bu formül güvenilirliği sağlamak için kodlayıcılar arasında oluşan görüş birliği ve görüş ayrılığı oranını hesaplar. Bu sayede kodlayıcıların uyumu ve kodlamanın kararlılığı konusunda fikir birliğine varılır. Bu araştırmada da öncelikle dahil edilen tüm çalışmalar ölçme ve değerlendirme alanında uzman akademisyen tarafından detaylı olarak incelenmiş ve araştırmacı tarafından analiz edilen veriler uzman tarafından yeniden analiz edilmiştir. Bu kapsamda 69 çalışmada görüş birliğine varılırken kullanılan BOBUT türü ve uygulama alanı konusunda 4 çalışmada görüş ayrılığı olmuştur. Görüş ayrılığına sebep olan dört çalışma hakkında yeniden görüşülmüş ve bu çalışmalar dahil etme kriterlerine uygun bulunarak araştırmaya dahil edilmiştir. Bu kapsamda araştırmacı ve uzman akademisyen arasında uyum seviyesi %93 olarak hesaplanmıştır. Miles ve Huberman’a (1994) göre uyum seviyesinin %70 üzerinde olması güvenilirğin bir bulgusu olarak görülmektedir.

3.5. Verilerin Analizi

Dâhil edilen çalışmalar sistemli olarak incelenmiş ve bu bilgiler Excel programında veri toplama formuna kodlanmıştır. Kodlama işleminde yayın yılı, yazar bilgileri, yayın türü, yayın dili gibi temel bilgilerle beraber çalışmalarda kullanılan test

madde sayıları, örneklem büyüklüğü, testi uygulama yöntemi, kullanılan yazılımlar, ölçek boyut yapısı, tercih edilen MTK modeli, madde seçim ve yetenek kestirim yöntemleri, durdurma ve başlatma kuralları gibi bilgiler de toplanmıştır.

Baltacı'ya (2019) göre nitel veri üzerinde detay gerektirmeyen araştırmalarda betimsel içerik analiz yöntemi kullanılabilir. Bu çalışmada da toplanan veriler betimsel içerik analiz yöntemiyle analiz edilmiştir. Araştırmaların benzerlik ve farklılıkları tespit edilerek bu bulgular sentezlenmiştir. Creswell'e (2013) göre nitel araştırmalarda bulguların objektif bir şekilde yorumlanması ve araştırma güvenilirliğinin sağlanması için alanında uzman kişilerden görüş alınmalıdır. Bu derleme çalışmasında da ölçme ve değerlendirme alanında üç uzmandan görüş alınmıştır. Bu kişiler çalışmanın bulgularını incelemiş analiz sürecine yönelik geri bildirimde bulunmuşlardır.

Tüm bu işlemler sistematik derleme sürecine uygun olarak güvenilir sonuçlar elde etmek için dikkatlice uygulanmıştır. Ayrıca kodlama sürecinde çalışmanın güvenilirliğini arttırmak adına tekrarlı kontroller yapılarak gerekli yerler düzeltilmiştir.

4. BULGULAR

Bu bölümde araştırmanın alt problemlerine yönelik elde edilen bulgulara yer verilmiştir. Araştırma temel amacı olan “Türkiye’de BOBUT konulu yapılan araştırmalarda mevcut durum nedir, BOBUT geliştirme ve uygulama süreçlerinde hangi kapsamlı özellikler ve metodolojik yaklaşımlar benimsenmiştir?” sorusuna cevaben araştırmaya dahil edilen çalışmalara ait; yılı, yazarı, kaynağı, yayın türü, yayın dili, çalışmanın adı ve çalışmanın bulgularına ait bilgiler Çizelge 4.1’de verilmiştir. Ayrıca incelenen 73 çalışmanın künyesi Ek-1’de sunulmuştur.



Çizelge 4.1. Araştırma Dahilinde Analiz Edilen Çalışmalara Ait Özellikler

	Yazar	Kaynak	Yayın Türü	Yayın Dili	Çalışmanın Adı	Çalışmanın Bulguları
1	Kaptan (1993)	Google Akademik	Tez	Türkçe	Yetenek kestiriminde adaptive (bireyselleştirilmiş) test uygulaması ile geleneksel kâğıt-kalem testi uygulamasının karşılaştırılması	BOBUT, kâğıt-kalem testlerine göre %55 zaman ve %70 madde tasarrufu sağlarken, öğrenci yetenek kestirimleri ÖSYS puanlarıyla anlamlı ilişki göstermiştir.
2	Yaşar (1999)	YÖKTez	Tez	Türkçe	Bireyselleştirilmiş testler üzerine bir çalışma	KTK ve MTK testlerinin güvenilirlik düzeyleri arasında anlamlı fark bulunmazken, yetenek ölçümleri arasındaki ilişki düşük bulunmuştur.
3	Elhan vd. (2008)	Google Akademik	Makale	İngilizce	An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain	BOBUT'un bel ağrısı olan hastalarda engelliliği ölçmek için potansiyel bir yöntem olduğu, iki boyutta yüksek korelasyon gösterdiği ve cevaplanan maddeleri önemli ölçüde azalttığı belirlenmiştir
4	Öztuna (2008)	YÖKTez	Tez	Türkçe	Kas-iskelet sistemi sorunlarının özürüllük değerlendirilmesinde bilgisayar uyarlamalı test yönteminin uygulanması	BOBUT ile engel düzeylerinin klasik uygulamaya göre daha az madde ile daha kısa sürede belirlenebileceği tespit edilmiştir.
5	Kalender (2011)	Google Akademik	Tez	İngilizce	Effects of different computerized adaptive testing strategies on recovery of ability	BOBUT'un ÖSS fen alt testinde, kâğıt-kalem testine kıyasla daha az soruyla daha yüksek güvenilirlikte yetenek kestirimleri sağladığı bulunmuştur.
6	Kaskatı (2011)	YÖKTez	Tez	Türkçe	Rasch modelleri kullanarak romatoid artirit hastaları özürüllük değerlendirmesi için bilgisayar uyarlamalı test yönteminin geliştirilmesi	Hastaların özürüllük düzeyini daha doğru ve etkin bir şekilde ölçmek için BOBUT yönteminin uygulanabilir olduğunu raporlanmıştır.
7	Aybek (2012)	Google Akademik	Tez	Türkçe	Kâğıt-kalem formu ve bilgisayar ortamında uygulanan genel yetenek testinin psikometrik özelliklerinin karşılaştırılması	Genel yetenek testte kâğıt-kalem ve BOBUT'un büyük ölçüde eşdeğer olduğu, bilgisayar aşinalığı ve cinsiyete göre fark göstermediği, ancak bilgisayara yönelik tutuma göre fark gösterdiği belirlenmiştir.
8	Özyurt vd. (2012)	WOS	Makale	İngilizce	Integrating computerized adaptive testing into UZWEBMAT: Implementation of individualized assessment module in an e-learning system	UZWEBMAT ile öğrencilerin bireysel farklılıkları daha iyi belirlenerek daha etkili bir bireysel değerlendirme sağlanmıştır.
9	Kalender (2012)	Google Akademik	Makale	İngilizce	Yükseköğretime öğrenci seçmek için bilgisayar ortamında bireyselleştirilmiş testler	BOBUT formatının, soru sayısını %80 azaltarak yüksek güvenilirlikte yetenek kestirimleri sağladığı belirlenmiştir.
10	Bulut ve Kan (2012)	Google Akademik	Makale	İngilizce	Application of computerized adaptive testing to entrance examination for graduate studies in Turkey.	BOBUT, EGS'nin kâğıt-kalem formatına göre %70 daha az madde ile daha doğru yetenek tahminleri sağlamış ve EAP, diğer tahmin yöntemlerinden daha iyi performans göstermiştir.
11	Gökçe (2012)	Google Akademik	Tez	İngilizce	Comparison of linear and adaptive versions of the Turkish pupil monitoring system (PMS) mathematics assessment	ÖİS matematik değerlendirmesi için en uygun algoritmanın WML olduğu ve hem madde kullanım sıklığı hem de içerik kontrol stratejilerinin kullanılması gerektiği belirlenmiştir.
12	Koşan (2013)	YÖKTez	Tez	Türkçe	Tıp eğitiminde gelişim sınavı soru bankası oluşturulması ve benzetim verileri ile bilgisayar uyarlamalı test uygulaması	Tıp eğitiminde gelişim sınavı için BOBUT uygulamaları, soru bankası ile daha az madde kullanarak yüksek güvenilirlikte yetenek düzeyleri kestirebildiği raporlanmıştır.
13	Kezer (2013)	YÖKTez	Tez	Türkçe	Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması	Çevrimiçi BOBUT testler, kâğıt kalem testlerine göre madde sayısında büyük tasarruf sağlamış ve farklı stratejilerin test performansını etkilediği belirlenmiştir.

14	Eroğlu (2013)	YÖKTez	Tez	Türkçe	Bireyselleştirilmiş bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının ölçme kesinliği ve test uzunluğu açısından karşılaştırılması	20 madde sabit uzunlukta RMSE ve yanlışlık değerlerinin düşük olduğu, madde havuzu büyüdükçe ölçme kesinliği için elde edilen RMSE ve yanlışlık değerlerinin genel olarak daha düşük olduğu bulunmuştur.
15	Sulak (2013)	YÖKTez	Tez	Türkçe	Bireyselleştirilmiş bilgisayarlı test uygulamalarında kullanılan madde seçme yöntemlerinin karşılaştırılması	Ortalama Maksimum Bayesyen Optimizasyonu yöntemi testin başında daha düşük standart hata ve daha iyi madde havuzu dengesi sağlamıştır.
16	Özbaşı (2014)	Google Akademik	Tez	Türkçe	Bilgisayar okuryazarlığı testinin bilgisayar ortamında bireye uyarlanmış test olarak uygulanabilirliğine ilişkin bir araştırma	BİLOKUR testinin, BOBUT uygulamasıyla %30 madde tasarrufuyla daha güvenilir kestirimler sağladığı belirlenmiştir
17	Öztürk (2014)	YÖKTez	Tez	Türkçe	Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde kullanım sıklığı kontrol yöntemlerinin incelenmesi	Azalarak Kaybolma yöntemi, madde havuzu kullanımı ile güvenilirliği en iyi sağlamıştır. Simpson-Hetter yöntemi ikinci en iyi güvenilirliği sağlamıştır. Madde kullanım kontrolü ölçme kesinliğini etkilememiştir.
18	Özyurt (2014)	TRDizin	Makale	İngilizce	The classification of the probability unit ability levels of the eleventh grade Turkish students by cluster analysis	UZWEBMAT-CAT'in Türk müfredatında güvenilir ve hassas testler üretebildiği gösterilmiştir.
19	Şahin ve Weiss (2015)	WOS	Makale	İngilizce	Effects of calibration sample size and item bank size on ability estimation in computerized adaptive testing	BOBUT uygulamalarında, 150 katılımcıyla yapılan kalibrasyon yeterlidir ve 200 maddelik madde bankası çoğu amaç için yeterli bilgi sağlamıştır.
20	Ayhan (2015)	Google Akademik	Tez	İngilizce	Comparability of scores from CAT and paper and pencil implementations of student selection examination to higher education	BOBUT'un daha az madde ile kağıt kalem formatından daha güvenilir ölçüm sağladığı ve alternatif yöntem olabileceği gösterilmiştir.
21	Özdemir (2015)	YÖKTez	Tez	Türkçe	Madde düzeyinde boyutluluk modellerinin bilgisayar ortamında bireyselleştirilmiş test yöntemleri üzerindeki etkisinin incelenmesi	A-Optimality madde seçiminde, Bayesyen MAP ve Fisher Puanlama benzer sonuçlar vermiş, içerik ağırlıklandırma madde sayısını artırıp güvenilirliği azaltmış, ancak RMSD ve standart hataları düşürmüştür.
22	Kaplan vd. (2015)	Google Akademik	Makale	İngilizce	New item selection methods for cognitive diagnosis computerized adaptive testing	MPWKL ve GDI indeksleri, PWKL'ye göre daha kısa testlerle daha yüksek doğru yetenek sınıflandırma oranları sağlamıştır. GDI indeksinin ise en kısa uygulama süresine sahip olduğu bulunmuştur
23	Gelibolu (2015)	YÖKTez	Tez	Türkçe	Temel bilgisayar eğitimi alanında bilgisayarlı uyarlanabilir sınavların ölçme özelliklerinin diğer sınav türleriyle karşılaştırılması	BOBUT'un diğer sınav türlerine göre daha az madde ile aynı veya daha iyi ölçüm yaptığı, zaman tasarrufu sağladığı ve sınav kaygısını azalttığı belirtilmiştir.
24	Özdemir (2016)	WOS	Makale	İngilizce	Comparison of different unidimensional-cat algorithms measuring students' language abilities: post-hoc simulation study	BOBUT, kağıt-kalem formatına göre daha yüksek güvenilirlik ve daha az madde ile yetenek tahminleri sağlamıştır. En iyi sonuçlar, FP-KL madde seçimi ve EAP kestirim yöntemi ile elde edilmiştir
25	Aybek (2016)	YÖKTez	Tez	Türkçe	Kendini değerlendirme envanterinin Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT) olarak uygulanabilirliğinin araştırılması	KDE-BOBUT'un geçerliliği ve güvenilirliği sağlanmış, KDE'nin BOBUT olarak uygulanabilirliği belirlenmiştir. Simülasyon ve canlı uygulama sonuçları tutarlı olduğu bulunmuştur.
26	Sarı (2016)	YÖKTez	Tez	İngilizce	Examining content control in adaptive tests: Computerized adaptive testing vs. computerized multistage testing	BOBUT ve ca-MST'nin içerik kontrolü karşılaştırıldığında, BOBUT'un daha esnek olduğu, ancak ca-MST'nin ölçüm doğruluğunda bazı avantajlar sunduğu tespit edilmiştir
27	Kaçar (2016)	Google Akademik	Tez	Türkçe	Bireye uyarlanmış bilgisayarlı test ortamında maksimum fisher madde seçme yönteminin incelenmesi	Madde havuzu arttıkça tahminin standart hatası düştüğü tespit edilmiştir.

28	Kaplan (2016)	Google Akademik	Makale	Türkçe	Nitelik sayısının madde seçme algoritmalarının performansı üzerindeki etkisi	Nitelik sayısının artması, ortalama test uzunluklarını artırmıştır. EBK değerlerinin artırılması, iki algoritmada da test uzunluklarını artırırken, kullanılan modellerin test uzunluklarına etkisi gözlenmemiştir.
29	Şahin ve Özbaşı (2017)	Google Akademik	Makale	İngilizce	Effects of content balancing and item selection method on ability estimation in computerized adaptive tests	İçerik dengelemede, düşük yetenek düzeylerinde ($\theta < 0$) LWI, yüksek yetenek düzeylerinde ($\theta > 0$) ise FMI daha başarılı bulunmuştur.
30	Şimşek (2017)	Google Akademik	Tez	Türkçe	Becerilere güven mesleki ilgi envanterinin uyarlanması ve bilgisayarlı bireyselleştirilmiş test uygulamasının geliştirilmesi	Mesleki ilgi envanterinin BOBUT ve kâğıt-kalem formu arasında yüksek eşdeğerlik bulunmuş, BOBUT'un kullanışlılığını artırmak için akıllı telefon uygulamaları önerilmiştir
31	Kalender ve Berberoğlu (2017)	Google Akademik	Makale	İngilizce	Can computerized adaptive testing work in students' admission to higher education programs in Turkey?	BOBUT'un Türkiye'de üniversiteye giriş için uygulanabilir bir alternatif olduğu ve kâğıt kalem testleri ile benzer sonuçları daha az maddeyle sağladığı gösterilmiştir.
32	Şahin (2017)	YÖKTez	Tez	Türkçe	Gerçek ve üretilmiş veri setlerinde çok boyutlu bireyselleştirilmiş bilgisayarlı test uygulama sonuçlarının incelenmesi	En iyi performans, A-optimality madde seçimi, MAP yetenek kestirimi ve madde-içi boyutluluk modeli ile elde edilmiştir.
33	Gür (2017)	YÖKTez	Tez	Türkçe	Bireye uyarlanmış testlerde madde kullanım sıklığı kontrol yöntemlerinin farklı koşullarda ölçme duyarlılığına ve test güvenliğine etkisi	Aşamalı Düşürme, Sympson Hetter'e göre madde havuzunu daha dengeli ve güvenli kullanmıştır. Kontrol olmadığında güvenlik düşmüş, ancak Aşamalı Düşürme daha iyi sonuçlar vermiştir.
34	Şenel (2017)	Google Akademik	Tez	Türkçe	Bilgisayar ortamında bireye uyarlanmış testlerin görme engelli öğrencilere uygunluğunun incelenmesi	CAT, kâğıt kalem testinden daha güvenilir ve hızlıdır, görme engelli öğrenciler tarafından olumlu bulunmuştur. İlk kez katılanlarda bilgisayar okuryazarlığı eksikliği nedeniyle endişeler oluşmuştur.
35	Ayan (2018)	Google Akademik	Tez	Türkçe	Bilişsel tanı modelinde geleneksel ve bilgisayarlı sınıflamalı test uygulamalarının psikometrik özelliklerinin karşılaştırılması	Hiyerarşik Bilişsel Tanı Modeli ile elde edilen örtük sınıf kestirimleri ile S-BOBUT sonuçları arasındaki tutarlık düzeyi belirlenmiştir.
36	Gündeğer ve Doğan (2018)	TRDizin	Makale	Türkçe	Bireyselleştirilmiş bilgisayarlı sınıflama testi kriterlerinin test etkililiği ve ölçme kesinliği açısından karşılaştırılması	Güven Aralığı (GA) ve Genelleştirilmiş Olabilirlik Oranı (GOO) sınıflama kriterlerinin test etkililiği ve ölçme kesinliği açısından en başarılı olduğu belirlenmiştir.
37	Yigit vd. (2018)	Google Akademik	Makale	İngilizce	Computerized adaptive testing for cognitively based multiple-choice data	JSD yöntemi, GDI yöntemine göre daha iyi sınıflandırma doğruluğu sağlamış ve bu, daha kısa test sürelerinde bile elde edilmiştir, bu da test süresini azaltabileceğini göstermektedir.
38	Mao Xiuzhen vd. (2018)	Google Akademik	Makale	İngilizce	Investigating the effect of exposure-control strategies on item selection methods in MCAT	V2 madde seçim yöntemi, diğerlerine göre daha iyi sonuç vermiştir. RT yöntemi en güvenilir olanı, RPG ve MPI ise onu takip ettiği belirtilmiştir.
39	Gür ve Karabay (2018)	Google Akademik	Makale	Türkçe	Motorlu taşıtlar sürücü kursiyerleri sınavının simülatif bilgisayar ortamında bireye uyarlanmış test uygulaması	BOBUT uygulaması, kâğıt kalem testiyle benzer yetenek ölçümü sağladığı ve daha az maddeyle kısa sürede testi tamamlayabildiği belirlenmiştir.
40	Demir (2018)	YÖKTez	Tez	Türkçe	Çok kategorili bireyselleştirilmiş bilgisayarlı test uygulamalarının farklı madde seçim yöntemlerinde sonlandırma kuralları açısından incelenmesi	MFI ve SH<0.30, çok kategorili modeller için en uygun madde seçim yöntemi olarak belirlenmiştir. BOBUT ve kâğıt kalem uygulamaları arasında pozitif anlamlı bir ilişki bulunmuştur.
41	Gökçe ve Glas (2018)	DergiPark	Makale	İngilizce	Can TIMSS mathematics assessments be implemented as a computerized adaptive test?	TIMSS 8. sınıf matematik değerlendirmeleri için BOBUT'un adımları: 6 rastgele madde seçimi, EAP yöntemiyle yetenek kestirimi, 20 madde uzunluk ve madde kullanım sıklığının kontrolü olduğu belirlenmiştir.

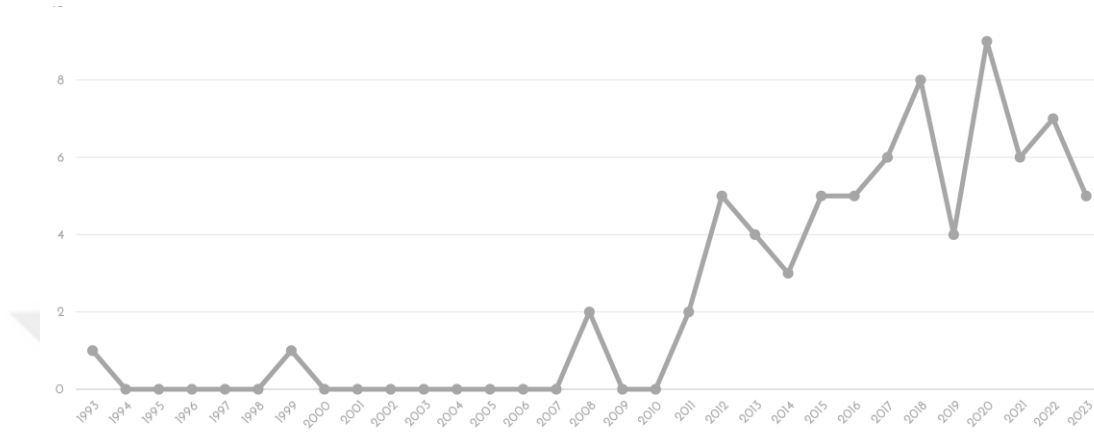
42	Şenel ve Kutlu (2018)	Google Akademik	Makale	Türkçe	Görme engelli öğrencilere yönelik bilgisayar ortamında bireye uyarlanmış test tasarımı	BOBUT'un görme engelli öğrenciler için geçerli ve güvenilir test sonuçları sağlayabileceği ve bu tür testlerin görme engelli öğrencilerin motivasyonunu ve özgüvenini artırabileceğini belirlenmiştir.
43	Erdem Kara (2019)	YÖKTez	Tez	Türkçe	Değişen madde fonksiyonu gösteren madde oranının bireyselleştirilmiş bilgisayarlı ve çok aşamalı testler üzerindeki etkisi	Çalışma, görme engelli öğrenciler için bireye uyarlanmış testlerin geçerli ve güvenilir olduğunu, öğrencilerin motivasyonunu ve özgüvenini artırabileceğini ortaya koymuştur.
44	Sarı (2019)	TRDizin	Makale	İngilizce	Investigating consequences of using item pre-knowledge in computerized multistage testing	Madde ön bilgisinin kullanıldığı durumlarda yetenek kestirimlerinin ciddi şekilde etkilendiğini ve test güvenliğinin artırılması gerektiğini göstermiştir.
45	Öztürk ve Şahin (2019)	TRDizin	Makale	İngilizce	Effects of item pool characteristics on ability estimate and item pool utilization: a simulation study	Madde havuzu büyüdükçe ölçme kesinliğinin de arttığı, madde kullanım sıklığı ve içerik dengeleme yöntemleriyle kullanılmayan madde sayısı azaldığı belirlenmiştir.
46	Aşiret (2019)	YÖKTez	Tez	Türkçe	Bilişsel tanıya dayalı bilgisayar ortamında bireye uyarlanmış testlerde madde seçim yöntemlerinin incelenmesi	Test uzunluğu arttıkça, JSD ve MPWKL algoritmalarının örüntü ve nitelik koruma oranlarının en yüksek olduğu belirlenmiştir.
47	Bostan (2020)	TRDizin	Makale	Türkçe	Using fuzzy logic in computerized adaptive tests	Uzman bilgisi ve bulanık mantık hesaplamalarının kullanıldığı uyarlamalı testlerin uzaktan eğitimde etkin olarak kullanılabileceği belirlenmiştir.
48	Çoban (2020)	YÖKTez	Tez	Türkçe	Bilgisayar temelli bireyselleştirilmiş test yaklaşımlarının Türkiye'deki merkezi dil sınavlarında uygulanabilirliğinin araştırılması	BOBUT ölçme kesinliği açısından öne çıkmış, BST-KP ise sınıflama doğruluğu açısından en iyisi olmuş. ÇABT ise madde havuzu kullanımı ve test güvenliği açısından en başarılı bulunmuştur.
49	Özgür (2020)	YÖKTez	Tez	Türkçe	Rasch modeline uygun yanıt desenlerinde bilgisayar uyarlamalı test yöntemi ile makine öğrenmesi yöntemlerinin tahmin performanslarının benzetim çalışması ile karşılaştırılması	Makine öğrenmesi yöntemlerinin BOBUT yöntemine alternatif olabileceği raporlanmıştır.
50	Ertuna (2020)	YÖKTez	Tez	İngilizce	The effect of item selection and parameter estimation methods to the accuracy of pretest item parameters on online calibration in CAT	OEM yöntemi bazı durumlarda daha doğru parametre tahminleri sağlamıştır. Örneklem büyüklüğü arttıkça parametre doğruluğu iyileşmiştir.
51	Kaplan ve Torre (2020)	WOS	Makale	İngilizce	A Blocked-CAT procedure for CD-CAT	Katılımcılara madde gözden geçirme imkanı tanıyan yeni CD-CAT prosedürünün, test kaygısını azaltırken sınıflandırma doğruluğunu etkilemediği belirlenmiştir.
52	Kürşad (2020)	Google Akademik	Tez	Türkçe	Madde parametre sapmasının bireye uyarlanmış testlerden elde edilen yetenek kestirimlerine etkisinin farklı koşullar altında incelenmesi	Madde parametre sapmasının ölçme kesinliği, test bilgi fonksiyonu ve test etkililiği üzerinde olumsuz etkiler yarattığı gösterilmiştir.
53	Şahin (2020)	DergiPark	Makale	İngilizce	Analyzing different module characteristics in computer adaptive multistage testing	Ölçme hassasiyeti, modül sayısı ve uzunluğu arttıkça yükselmiş, yüksek madde ayırt ediciliğine sahip durumlarda hataların azaldığı tespit edilmiştir.
54	Çıkrıkçı vd. (2020)	Google Akademik	Makale	İngilizce	Development of a computerized adaptive version of the Turkish driving licence exam	ECODL'nin CAT olarak uygulanabilir olduğu ve adayların teorik sürüş bilgi yeterliliklerini güvenilir şekilde ölçebildiği belirlenmiştir. Adaylar, bilgisayar tabanlı testi daha pratik bulmuşlardır.
55	Tat (2020)	YÖKTez	Tez	Türkçe	Açıklayıcı madde tepki modellerinin bilgisayar ortamında bireye uyarlanmış testlerde kullanımı	Açıklayıcı madde tepki modellerinin madde havuzları ve yetenek kestirimleri üzerinde etkili olduğu bulunmuştur. Modeller arasında manidar farklar da tespit edilmiştir.

56	Raborn ve Sarı (2021)	WOS	Makale	İngilizce	Mixed adaptive multistage testing: A new approach	Çalışma, ma-MST'nin geleneksel MST'ye göre daha düşük ortalama yanlışlık, daha düşük RMSE ve daha yüksek korelasyon sağladığını göstermiştir.
57	Kezer (2021)	Google Akademik	Makale	İngilizce	The effect of item pools of different strengths on the test results of computerized-adaptive testing	Yetenek seviyesiyle uyumlu madde zorluk seviyelerine sahip havuzların daha az maddeyle daha doğru tahminler yaptığı görüldüğü ifade edilmiştir.
58	Demir ve Atar (2021)	TRDizin	Makale	İngilizce	Investigation of classification accuracy, test length and measurement precision at computerized adaptive classification tests	Çalışmanın sonuçlarına göre, MFI-EB, MFI-CB'den daha faydalı bulunmuş ve MMM içerik dengeleme açısından daha başarılı olarak değerlendirilmiştir.
59	Alkan (2021)	Google Akademik	Tez	Türkçe	Mesleki alan ilgi envanterinin bilgisayar ortamında bireye uyarlanmış formunun geliştirilmesi	Mesleki Alan İlgi Envanteri'nin BOBUT uygulamasının daha az maddeyle yüksek güvenilirlik sağladığı ve post-hoc simülasyonlar ile canlı uygulamalar arasında anlamlı ilişkiler bulunduğu gözlemlenmiştir.
60	Arslan (2021a)	YÖKTez	Tez	Türkçe	Bilgisayar uyarlamalı testlerde madde yanıt sürelerinin kullanımı	BOBUT ve ÇD modellerinin performansları karşılaştırılmış, her iki modelin de yetenek düzeyi kestiriminde başarılı olduğu bulunmuştur. Ancak, ÇD modelinin daha kapsamlı bilgi sağladığı gözlenmiştir.
61	Arslan vd. (2021b)	DergiPark	Makale	Türkçe	Uzaktan eğitimde bilgisayar uyarlamalı test ve kâğıt-kalem testi ile yetenek düzeylerinin belirlenmesi: çok merkezli gerçek veri uygulaması	BOBUT'un kâğıt-kalem testine göre daha hızlı ve az madde ile güvenilir sonuçlar sağladığı ve soru sayısını %80 oranında azalttığı bulunmuştur.
62	Balta ve Uçar (2022)	DergiPark	Makale	İngilizce	Bilgisayar ortamında bireye uyarlanmış test uygulamalarında ölçme kesinliğinin ve test uzunluğunun farklı koşullar altında incelenmesi	Araştırmada, farklı durdurma kurallarının test uzunluğunu ve ölçme kesinliğini etkilediği, en düşük yanlışlık ve RMSE değerlerinin SH<0.30 durdurma kuralıyla elde edildiği sonucuna varılmıştır.
63	Kaya (2022)	Google Akademik	Tez	İngilizce	A comparability and classification analysis of computerized adaptive and conventional paper-based versions of an English language proficiency reading subtest	BOBUT ve kâğıt kalem uygulamalarının yetenek kestirimlerinde benzer olduğu belirtilmiştir. Tek bir kesme puanının kullanılmasının daha iyi sınıflandırma performansına yol açtığı tespit edilmiştir.
64	Ertaş Polat (2022)	YÖKTez	Tez	Türkçe	Çok aşamalı bireye uyarlanmış testlerde farklı koşullardan elde edilen yetenek kestirimlerinin karşılaştırılması	Farklı zorluk seviyelerinden oluşturulan testlerde, EAP ve statik yönlendirme yöntemleriyle yapılan uzun modüllü tasarımın en düşük hata oranlarına sahip olduğu bulunmuştur.
65	Aracı ve Tan (2022)	Google Akademik	Makale	İngilizce	Multidimensional computerized adaptive testing simulations in R	Çok boyutlu BOBUT'un kabul gören BOBUT'lara göre daha etkili olduğu ve ölçüm doğruluğunu artırdığı bulunmuştur.
66	Şen (2022)	YÖKTez	Tez	Türkçe	Bilgisayarda bireyselleştirilmiş test uygulamalarında maddeyi yeniden cevaplayabilmenin ölçme hatasına ve yanlışlığına etkisinin incelenmesi	Öncül bilgilerin, BOBUT algoritmasını daha kullanışlı kıldığı tespit edilmiştir.
67	Aşiret ve Sumbül (2022)	Google Akademik	Makale	Türkçe	BT-BOBUT uygulamalarında madde seçim algoritmalarının performanslarının ölçme doğruluğu açısından incelenmesi	Test uzunluğu arttıkça ölçme doğruluğunun arttığı, nitelik sayısı arttıkça ölçme doğruluğunun azaldığı bulunmuştur. JSD en yüksek ölçme doğruluğuna sahipken, PWKL algoritması en düşük doğruluğa sahiptir.
68	Gören vd. (2022)	DergiPark	Makale	İngilizce	The effect of aberrant responses on ability estimation in computer adaptive tests	4PL modeli, sapkın yanıtların varlığında yetenek kestiriminde daha yüksek doğruluk ve etkinlik sağlamıştır. Değişken uzunluklu testlerde genel olarak daha iyi performans göstermiştir.
69	Gökcan (2023)	YÖKTez	Tez	Türkçe	İngilizce kelime bilgisi testinin (VST) bireyselleştirilmiş bilgisayarlı test olarak uygulanabilirliğinin incelenmesi	Kâğıt-kalem testi ile BOBUT uygulamasında kestirilen yetenek puanları arasında yüksek ilişki bulunmuştur. BOBUT daha az madde kullanarak İngilizce kelime bilgisini güvenilir bir şekilde kestirmiştir.

70	Taymur (2023)	YÖKTez	Tez	Türkçe	BOBUT uygulamalarında farklı koşullardan elde edilen ölçme kesinliği kestirim değerlerinin karşılaştırılması	1000 madde olan havuzun daha iyi sonuçlar verdiği madde seçme yöntemleri arasında MFI ve yetenek kestirim yöntemleri arasında ise EAP yöntemlerinin daha etkili olduğu belirlenmiştir.
71	Yiğiter (2023a)	Google Akademik	Tez	Türkçe	Sabit ve anında bireyselleştirilmiş çok aşamalı testlerin karşılaştırılması	A-BÇAT'nin hem ölçme kesinliği hem de madde güvenliği açılarından S-BÇAT'tan daha iyi sonuçlar ürettiği belirtilmiştir.
72	Yiğiter ve N doğan (2023b)	DergiPark	Makale	İngilizce	Comparison of different computerized adaptive testing approaches with shadow test under different test length and ability estimation method conditions	CAT, 2 Aşamalı O-MST ve 3 Aşamalı O-MST'nin etkinlik açısından oldukça benzer olduğu, LOFT'un ise daha az etkili olduğu ifade edilmiştir.
73	Kürşad (2023)	Google Akademik	Makale	İngilizce	The effects of different item selection methods on test information and test efficiency in computer adaptive testing	MFI yöntemi, test bilgi fonksiyonu (TIF) açısından en yüksek performansı gösterdiği, test verimliliği açısından, rastgele seçim yöntemi dışındaki tüm yöntemlerin benzer olduğu belirlenmiştir.

4.1. Çalışmaların Yıllara Göre Dağılımı

Araştırmanın ilk alt problemi olan “Türkiye’de BOBUT konulu çalışmalar hangi yıllarda yapılmıştır?” sorusuna yönelik çalışmaların yıllara göre değişimi Şekil 4.1’de ve dağılımı da Çizelge 4.2’de sunulmuştur.



Şekil 4.1. BOBUT Çalışmalarının Yıllara Göre Değişimi

Şekil 4.1, 1993 ile 2024 yılları arasında BOBUT çalışmalarının süreçteki ilerleyişini ve yıllara göre değişimini göstermektedir. Şekle göre 2010 yılından itibaren BOBUT çalışmalarına eğilimin belirgin şekilde arttığı söylenebilir.

Çizelge 4.2. İncelenen Çalışmaların Yıllara Göre Dağılımı

	f	%
1993-2000	2	2.7
2001-2005	0	0
2006-2010	2	2.7
2011-2015	19	26
2016-2020	32	43.8
2021-2024	18	24.7
Toplam	73	100

Çizelge 4.2’de yıllara göre BOBUT alanında yapılan çalışmaların dağılımı gruplandırılmış veriler halinde sunulmuştur. Görüleceği üzere BOBUT alanında Türkiye’de ilk çalışma 1993 yılında yapılmış ve 2000 yılına kadar sadece 2 (%2.7) çalışma yapılmıştır. 2000-2005 yıl aralığında hiç çalışma yapılmamış ve 2006-2010

yıllarında ise yine 2 (%2.7) çalışma yapılmıştır. 2011 yılı itibariyle belirgin bir artışla BOBUT alanında yapılan çalışma sayısı 19'a (%26) yükselmiştir. BOBUT alanında en yoğun çalışmaların olduğu dönem %43.8 oranıyla 2016-2020 yıllarında olup toplam 32 çalışma yapıldığı görülmektedir. Literatür taramasının 2024 Mart ayına kadar olduğu göze alındığında son iki yılda 18 (%24.7) çalışmanın yapılmış olması çalışma sayısının artarak devam ettiğini göstermektedir. Çizelge 4.2'deki verilere göre BOBUT konulu çalışmaların zamanla arttığı ve özellikle 2010 yılından sonra daha fazla yoğunlaştığı ifade edilebilir.

4.2. Çalışmaların Yayın Türlerine Göre Dağılımı

Araştırmanın bu bölümünde ikinci alt problem olan “Çalışmaların yayın türleri nelerdir?” sorusuna yönelik bulgulara yer verilmiştir. 1993 ile 2024 yılları arasında yapılan çalışmaların yayın türlerinin dağılımı Çizelge 4.3'te sunulmuştur.

Çizelge 4.3. Çalışmaların Yayın Türlerine Göre Dağılımı

	f	%
Tez	40	54.8
Makale	33	45.2
Toplam	73	100

Çizelge 4.3'e bakıldığında araştırma kapsamında incelenen çalışmaların yayın türleri frekans ve yüzde olarak verildiği görülmektedir. Çizelgeye göre BOBUT konulu yapılan çalışmaların 40'ının (%54.8) tezlerden, 33'ünün (%45.2) makalelerden oluştuğu görülmektedir. Toplam 73 çalışmanın değerlendirildiği araştırmada tez sayısının makale sayına göre daha fazla olduğu tespit edilmiştir.

4.3. Çalışmaların Yapılma Amacına Göre Dağılımı

Araştırmanın üçüncü alt problemi ışığında “Çalışmalar yoğunlukta hangi amaçla yapılmıştır?” sorusuna yanıt aranmış ve bulgular Çizelge 4.4'te sunulmuştur.

Çizelge 4.4. Çalışmaların Amacına Göre Dağılımı

	f	%
Gerçek Uygulama Yapma	26	35.6
Test Stratejilerinin ve Parametrelerin İncelenmesi	24	32.8
Farklı BOBUT Türlerinin Denenmesi	17	23.3
Özel Alan ve Konular için BOBUT Geliştirme	6	8.2
Toplam	73	100

Çizelge 4.4 çalışmaların yapılma amacına göre benzer konularda sınıflandırılmasını içermektedir. Çizelgedeki verilere göre %35.6 (26) oranıyla en çok çalışmanın gerçek BOBUT uygulaması gerçekleştirmek amacıyla yapıldığı görülmektedir. Test stratejisi ve parametrelerinin incelenmesi kategorisi bu alanda farklı yöntemlerin (başlatma kuralı, madde seçim, yetenek kestirim, durdurma kuralı) ve parametrelerin (örneklem büyüklüğü, madde havuzu genişliği, madde parametreleri) denendiği çalışmaları kapsamaktadır. Bu amaçla yapılan çalışmaların sayısı 24 olup ikinci en çok çalışılan amaçla tüm çalışmaların %32.8'ini oluşturmaktadır. Çok aşamalı BOBUT, bilişsel tanıya dayalı BOBUT ve çok boyutlu BOBUT gibi farklı BOBUT türlerinin denendiği çalışmalar ağırlık olarak %23.3 (17) ile üçüncü sırada gelmektedir. Özel gereksinimli bireyler ya da belli amaca yönelik BOBUT geliştirme çalışmalarının bir araya getirildiği özel alan ve konular için BOBUT geliştirilme kategorisi 6 çalışmayla %8.2 oranında bulunmaktadır.

4.4. Çalışmaların Yayın Diline Göre Dağılımı

Araştırmanın dördüncü alt problemi olan “Çalışmaların ne kadarı Türkçe ve ne kadarı İngilizce yayımlanmıştır?” sorusuna yanıt olarak bulgular Çizelge 4.5’te verilmiştir.

Çizelge 4.5. Çalışmaların Yayın Diline Göre Dağılımı

	f	%
Türkçe	41	56.2
İngilizce	32	43.8
Toplam	73	100

Çizelge 4.5'te çalışmaların hangi dillerde yayımlandığı yüzde ve frekans bilgileriyle verilmiştir. Çalışmaların Türkçe olarak yayımlanma oranı %56.2 iken İngilizce yayımlananı %43.8 olduğu görülmektedir. İncelenen 73 çalışmanın 41'i Türkçe, 32'si İngilizce olup Türkçe çalışmaların İngilizce çalışmalara göre nispeten fazla olduğu görülmektedir.

4.5. Çalışmaların Örneklem Gruplara Göre Çeşitliliği

Araştırmanın beşinci alt problemi olan “Çalışmalarda hangi örneklem çeşidi yoğunluktadır?” sorusuna yönelik çalışmaların örneklem grup dağılımı Çizelge 4.6'da sunulmuştur.

Çizelge 4.6. Çalışmalarda Kullanılan Örneklem Gruplarının Dağılımı

	f	%
Simülasyon Verisi	39	51.3
Öğrenci	26	34.2
Gerçek Veriye Dayalı Simülasyon	9	11.8
Hasta	2	2.6

Çizelge 4.6 incelendiğinde çalışmalarda en yoğun kullanılan örneklem türünün %51.3 (39) ile simülasyondan üretilmiş veri olduğu görülmektedir. Ardından Canlı BOBUT uygulamalarında tercih edilen öğrenciler %34.2 (26) oranıyla ikinci ağırlıkta örneklem gurubunu oluşturmaktadır. Gerçek veri örnekleminden oluşan simülasyon çalışmaları %11.8 (9) oranında olurken en az çalışılan örneklem grubu %2.6 (2) oranıyla hastalardan oluşmaktadır. Bu çizelgede bazı çalışmalar hem simülasyon hem gerçek uygulama üzerinden yapıldığından analiz edilen örneklem grup miktarı çalışma sayısından fazla olmaktadır.

4.6. Çalışmalarda Kullanılan Farklı BOBUT Türleri

Araştırmanın bu bölümünde altıncı alt problem olan “Çalışmalarda farklı BOBUT türleri kullanım oranı nedir?” sorusuna yönelik bulgulara yer verilmiştir. Çalışmalarda kullanılan BOBUT türlerinin dağılımı Çizelge 4.7'de sunulmuştur.

Çizelge 4.7. Çalışmalarda Kullanılan Farklı BOBUT Türlerinin Dağılımı

	f	%
BOBUT	55	70
Çok Aşamalı BOBUT	6	8
Sınıflandırmaya Dayalı BOBUT	5	6
Çok Boyutlu BOBUT	4	5
Bilişsel Tanıya Dayalı BOBUT	4	5
Diğer	5	6

Çizelge 4.7'ye bakıldığında en yoğun çalışmanın %70 (55) oranında BOBUT'un kendisi üzerinde yapıldığı görülmektedir. Farklı türler arasında ise en fazla kullanılan %8 (6) oranıyla çok aşamalı BOBUT olurken onu %6 (5) oranında kullanımla sınıflandırmaya dayalı BOBUT takip etmektedir. Yine az sayıda çalışmanın olduğu diğer BOBUT türleri %5 (4) oranla çok boyutlu BOBUT ile aynı oranda bilişsel tanıya dayalı BOBUT olmuştur. Diğer kategorisi karma BOBUT gibi yaygın kullanılmayan ve deneme amaçlı yapılan çalışmaları içermektedir. Bazı çalışmalar BOBUT'un farklı türlerini karşılaştırmasını içerdiğinden çizelgede görülen BOBUT türleri sayısı çalışma sayısının toplamından fazladır.

4.7. Çalışmalarda Canlı BOBUT Uygulaması ve Simülasyon Oranları

Araştırmanın yedinci alt problemi ışığında “Canlı BOBUT uygulaması ve simülasyon verisi kullanım oranı nedir?” sorusuna yanıt aranmış ve bulgular Çizelge 4.8'de sunulmuştur.

Çizelge 4.8. Çalışmaların Canlı BOBUT ve Simülasyonda Yapılma Oranları

	f	%
Canlı BOBUT Uygulaması	26	27.3
Simülasyon Uygulaması	69	72.6
Monte Carlo Simülasyonu	33	47.8
Post-hoc Simülasyonu	21	30.4
Türü Belirtilmemiş	15	21.7

Çizelge 4.8'e bakıldığında uygulama türleri canlı BOBUT ve simülasyon çalışmaları olarak iki gruba ayrılırken simülasyonda yapılan çalışmalar kendi içinde üç grupta incelenmiştir. Bazı çalışmalar hem simülasyon hem de gerçek uygulamaya dayalı olduğundan uygulama yöntemi incelenen çalışma sayısından fazladır. Veriler incelendiğinde canlı BOBUT uygulama sayısı 26 olurken oranı %27.3'tür. Simülasyonda yapılan toplam 69 çalışmaya erişmiş ve bu tüm çalışmaların %72.6'sını kapsamaktadır. Simülasyon çalışmalarının türlerine bakıldığı Monte Carlo simülasyonları %47.8 (33) ile en yaygın kullanılan simülasyon çeşidi olurken Post-hoc simülasyon kullanım oranı %30.4 (21) olduğu görülmektedir. Simülasyonda yapıldığı belirtilen ancak hangi tür simülasyon çalışması olduğu belirtilmeyen çalışmaların oranı ise %21.7'dir (15).

4.8. Çalışmalarda Kullanılan BOBUT Uygulama, Simülasyon ve Madde Kalibrasyon Yazılımları

Araştırmanın sekizinci alt problemi olan “BOBUT geliştirme ve uygulamasında sürecinde en sık kullanılan yazılımlar hangileridir?” sorusuna yanıt olarak sonuçlar kategoriler halinde Çizelge 4.9'da verilmiştir.

Çizelge 4.9. Çalışmalarda Kullanılan Yazılımlar

	Yazılımın Adı	f
BOBUT Uygulama Yazılımları	Concerto	6
	Firestar	5
	SmartCAT	4
	UZWEBMAT-CAT	2
	R (caMST)	1
	Firestar-D	1
	MICROCAT	1
Simülasyon Uygulaması	SimulCAT	7
	R (catR)	6
	R (catSIM)	4
	R (mstR)	3
	Simulate_CAT	3
	MSTGen	1
Madde Kalibrasyonu	BILOG-MG	7
	R (catIrt)	5
	R (RUMMss)	4
	R (mirt)	4
	MULTILOG	2
	Parscale	2
	Diğer	6
Bilgi Verilmemiş		9

Çizelge 4.9’da görüldüğü üzere yazılımlar BOBUT uygulama yazılımları, simülasyon yazılımları ve madde kalibrasyon yazılımları olmak üzere üç kategoride incelenmiştir. İlk kategori olan BOBUT uygulama yazılımlarına bakıldığında Concerto 6 çalışmada kullanılarak en sık kullanılan yazılım olarak öne çıkmaktadır. Firestar 5 çalışmada kullanılarak ikinci en yaygın yazılım olurken SmartCAT 4 çalışmada tercih edilmiştir. UZWEBMAT-CAT, R (caMST), Firestar-D ve MICROCAT daha az sıklıkta tercih edilmiştir. İkinci kategori olan simülasyon

yazılımları incelendiğinde SimulCAT toplam 7 çalışmada tercih edilerek en sık kullanılan simülasyon yazılımı olduğu görülmektedir. R programı temelli catR, catSIM, mstR paketleri de simülasyonda sıklıkla tercih edilmiştir. Simülasyon yazılımlarında MstGEN sadece bir çalışmada kullanılarak en az kullanılan simülasyon yazılımı olduğu görülmektedir. Üçüncü kategori olan madde kalibrasyon yazılımlarında ise 7 çalışmada kullanılan BILOG-MG en sık tercih edilen madde kalibrasyon yazılımı olmuştur. Bunların yanında R programlama dilinde geliştirilen MTK paketlerinden catIrt, RUMMss, mirt test maddelerinin analiz edilmesi ve kalibre edilmesinde sıklıkla tercih edilmiştir. MULTILOG, Parscale gibi yazılımların madde kalibrasyonunda kullanımı diğerlerine oranla daha azdır. İncelenen çalışmaların dokuzunda kullanılan yazılım bilgisi verilmediği için analize dahil edilmemiştir.

4.9. Çalışmaların Örneklem Büyüklüklerinin Dağılımı

Araştırmanın dokuzuncu alt problemi olan “Çalışmaların örneklem büyüklükleri nedir?” sorusuna yönelik örneklem büyüklüklerinin gruplandırılmış dağılımı ayrı ayrı değerlendirilmiş ve sonuçlar canlı BOBUT uygulamaları için Çizelge 4.10’da, simülasyon uygulamaları için Çizelge 4.11’de sunulmuştur.

Çizelge 4.10. Canlı BOBUT Uygulamalarında Kullanılan Örneklem Büyüklüğünün Dağılımı

Örneklem Büyüklüğü	f	$\bar{x}$	M	σ	%
25-100	14	56.6	56	22.8	53.8
101-200	7	144.1	142	18.3	26.9
201-300	2	275	275	5	7.7
301-400	0	0	0	0	0
401-500	2	474	474	17	7.7
501-600	0	0	0	0	0
601-700	1	608	608	0	3.8

f: Frekans, $\bar{x}$: Ortalama, M: Medyan, σ : Standart Sapma

Çizelge 4.10’da canlı BOBUT uygulamalarında kullanılan örneklem büyüklükleri gruplandırılmış olarak sunulmuştur. Veriler incelendiğinde en yaygın örneklem grubu 14 çalışmada kullanılan %53.8 oranıyla 25-100 aralığında yer

almaktadır. Örneklem büyüklüğü olarak 101-200 aralığı 7 çalışmada kullanılmış olup 26.9 oranındadır. 201-600 aralığında 5 örneklem büyüklüğü kullanılmış olduğu görülmekle birlikte canlı BOBUT uygulamalarının sıklıkla küçük örneklem gruplarında yoğunlaştığı ifade edilebilir.

Çizelge 4.11. Simülasyon Uygulamalarında Kullanılan Örneklem Büyüklüğünün Dağılımı

Örneklem Büyüklüğü	f	$\bar{x}$	M	σ	%
250-1000	28	831.7	1000	275.3	40.6
1001-2000	13	1590.2	1440	335.3	18.8
2001-3000	7	2834.8	3000	277.9	10.1
3001-4000	2	3529.0	3529	471.0	2.9
4001-5000	8	5000.0	5000	0	11.6
5001-10000	6	8666.6	10000	1885.6	8.7
10001.15000	3	13535.0	13605	1225.7	4.4
Bilgi Verilmemiş	2	0	0	0	2.9

f: Frekans, $\bar{x}$: Ortalama, M: Medyan, σ : Standart Sapma

Çizelge 4.11’de görüldüğü üzere simülasyon yazılımlarında kullanılan örneklem büyüklüğü geniş bir yelpazede dağılım göstermektedir. Simülasyon uygulamalarında en yaygın tercih edilen örneklem büyüklüğü %40.6 oranıyla 28 çalışmada kullanılan 250-1000 aralığıdır. 1001-2000 aralığı ise %18.8 oranıyla 13 çalışmada tercih edilerek ikinci en sık kullanılan örneklem grubudur. 2001-3000 aralığındaki örneklem büyüklükleri çalışmaların %10.1’inde (7) kullanılmıştır. 3001 ile 15000 aralığı toplam 26 çalışmada tercih edilmiştir. İncelenen çalışmalardan 2’sinde örneklem büyüklükleri belirtilmemiştir.

4.10. Çalışmaların Madde Havuzu Büyüklüklerinin Dağılımı

Araştırmanın bu bölümünde onuncu alt problem olan “Çalışmalarda kullanılan madde havuzu genişliği ne kadardır?” sorusuna yönelik bulgulara yer verilmiştir. BOBUT uygulamalarında kullanılan madde havuzu büyüklükleri dağılımı Çizelge 4.12’de sunulmuştur.

Çizelge 4.12. Çalışmalarda Kullanılan Madde Havuzu Büyüklükleri

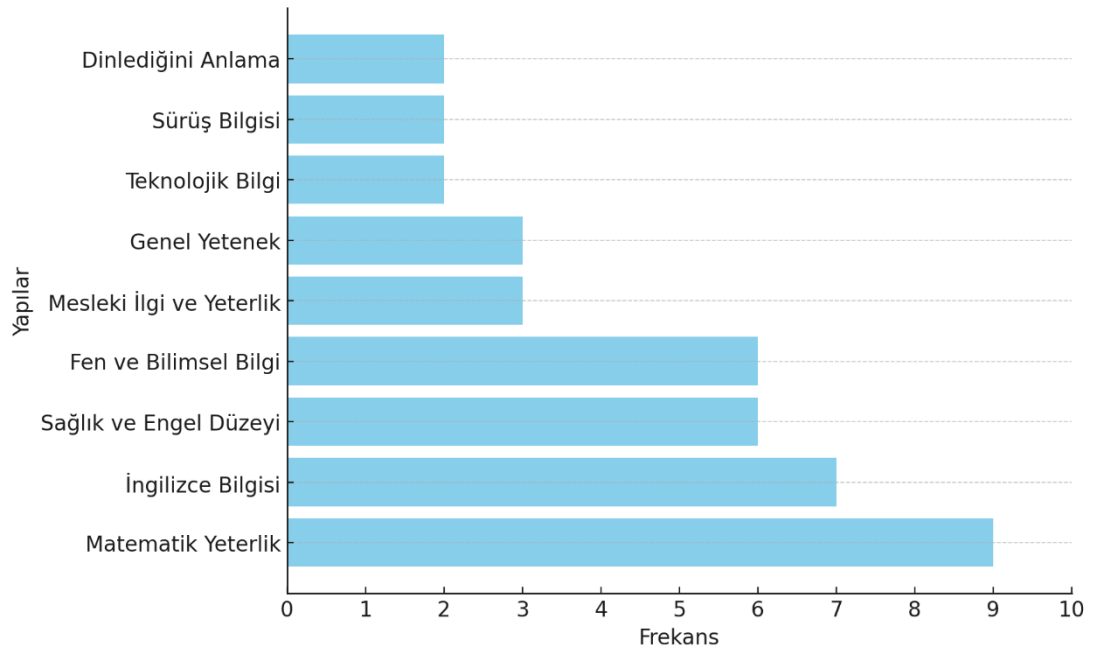
Madde Sayısı	f	$\bar{x}$	M	σ	%
10-100	22	57.0	50	25.1	21.8
101-200	17	148.5	142	33.5	16.8
201-300	15	249.2	244	21.8	14.9
301-400	4	178.3	397	33.8	4
401-500	16	491.3	500	16.7	15.8
501-600	7	578.1	600	35.7	6.9
601-750	3	644.7	628	29.4	3
750-1000	10	898.2	926	101.5	9.9
1001-3000	7	1420.0	1240	476.2	6.9

f: Frekans, $\bar{x}$: Ortalama, M: Medyan, σ : Standart Sapma

Çizelge 4.12’ye bakıldığında BOBUT uygulamalarında kullanılan madde havuzu çeşitli dağılımlar göstermektedir. 10-100 aralığı 22 çalışmada kullanılmış olup 21.8 oranıyla en sık kullanılan madde havuzu büyüklüğü olduğu ifade edilebilir. İkinci en yaygın kullanılan madde havuzu 101-200 aralığı olup 17 (16.8) çalışmada kullanılmıştır. 401-500 aralığında 16 (15.8) çalışma, 201-300 aralığında 15 (%14.9) çalışma, 750-1000 aralığında 10 (%9.9), 501-600 aralığında 7 (%6.9) çalışma yer almaktadır. 301-400 ile 601-710 aralığı sırasıyla 4 (%4) ve 3 (%3) çalışma bulunurken en az tercih edilen aralık olmuşlardır. Veriler incelendiğinde BOBUT çalışmalarında kullanılan madde havuzu büyüklüğün 10-100 aralığında yoğunlaştığı ve dar madde havuzlarının daha fazla tercih edildiği görülmüştür.

4.11. Çalışmalarda Ölçülen Yapıların Çeşitliliği

Araştırmanın on birinci alt problemi ışığında “BOBUT çalışmalarında ölçülen yapıların çeşitliliği nasıldır?” sorusuna yanıt aranmış ve yapılar ortak konularda bir araya getirilerek Şekil 4.2’de görselleştirilmiştir.



Şekil 4.2. BOBUT Uygulamalarında Ölçülen Yapıların Çeşitliliği

Şekil 4.2’de görüldüğü üzere BOBUT uygulamalarında ölçülen yapılar belli alanlarda yoğunlaşmaktadır. Matematik yeteneği 9 çalışmada tekrarlanarak en sık ölçülen yapı olmuştur. İngilizce yeterliği de 7 çalışmada yer alarak en sık ölçülen ikinci yapı durumundadır. Sağlık ve engel düzeyi toplam 6 çalışmada ölçülürken, mesleki ilgi ve yeterlik, genel yetenek düzeyi gibi yapılar üçer çalışmada ölçülmüştür. Araştırmalarda teknolojik bilgi, sürüş bilgisi ve dinlediğini anlama durumu da BOBUT’ta ölçülen yapılar arasındadır. Çalışmaların 32’sinde ölçülen yapılar hakkında herhangi bir bilgi verilmemiştir.

4.12. Çalışmalarda Ölçülen Yapıların Boyut Durumları

Araştırmanın on ikinci alt problemi olan “Ölçülen yapıların tek boyutlu ve çok boyutlu olma durumları nedir?” sorusuna yanıt olarak bulgular Çizelge 4.13’te verilmiştir.

Çizelge 4.13. Çalışmalarda Ölçülen Boyut Durumları

	f	%
Tek Boyutlu	32	42.2
Çok Boyutlu	14	18.7
Bilgi Verilmemiş	29	38.7

Çizelge 4.13'te veriler tek ve çok boyutlu olarak gruplandırılmıştır. Ölçülen yapının iki ve daha fazla boyut içermesi çok boyut olarak adlandırılmıştır. Çizelgeye bakıldığında tek boyutlu yapıların testlerde belirgin şekilde öne çıktığı ve 32 kez ölçülerek %42.2 oranında yer verildiği görülmektedir. Çok boyutlu yapıların oranı %18.7 olurken 14 çalışmada tercih edilmiştir. Araştırmaların %38.7'sini içeren 29 çalışmada ölçülen yapının durumu hakkında bilgi verilmemiştir.

4.13. Çalışmalarda Kullanılan En Yaygın MTK Modellerinin Dağılımı

Araştırmanın on üçüncü alt problemi olan “En yaygın tercih edilen MTK modelleri hangileridir?” sorusuna yönelik MTK modellerinin dağılımı Çizelge 4.14'te verilmiştir.

Çizelge 4.14. Çalışmalarda Kullanılan MTK Modellerinin Dağılımı

	f	%
1PL	6	7.1
2PL	12	14.3
3PL	33	39.3
DINA	6	7.1
DINO	4	4.8
GRM	4	4.8
Diğer	11	13.1
Bilgi Verilmemiş	8	9.5

1PL: 1-Parameter Logistic Model, 2PL: 2-Parameter Logistic Model, 3PL: 3-Parameter Logistic Model, DINA: Deterministic Inputs, Noisy "And" gate, DINO: Deterministic Inputs, Noisy "Or" gate, GRM: Graded Response Model

Çizelge 4.14'te görüldüğü üzere en sık kullanılan MTK modeli %39.3 oranıyla 33 çalışmada kullanılan 3PL modelidir. İkinci en sık kullanılan MTK modeli %14.3 oranıyla 12 çalışmada kullanılan 2PL modelidir. 1PL ve DINA modeli %7.1 oranla altışar çalışmada tercih edilirken DINO ve GRM modeli dörder (%4.8) çalışmada kullanılmıştır. Diğer kategorisinde toplanan kısmı kredi ve yenelleştirilmiş kısmı kredi modeli gibi modeller 11 çalışmada %13.1 oranında tercih edilmiştir. Çalışmaların 8'inde (%9.1) kullanılan MTK modeli hakkında herhangi bir bilgi verilmemiştir.

4.14. Çalışmalarda En Yaygın Kullanılan Yöntemler

Araştırmanın bu bölümünde on dördüncü alt problem olan “Testlerde en yaygın kullanılan madde seçim yöntemi, madde kestirim yöntemi, yetenek kestirim yöntemi, testi başlatma kuralı ve test sonlandırma kuralları nelerdir?” sorusuna yanıt aranmıştır. Yöntemler ayrı ayrı başlıklar halinde incelenmiş ve sonuçlar tablolastırılmıştır. En yaygın kullanılan madde seçim yöntemleri Çizelge 4.15’te, madde kestirim yöntemleri Çizelge 4.16’da, yetenek kestirim yöntemleri Çizelge 4.17’de, testi başlatma kuralları Çizelge 4.18’de ve testi sonlandırma kuralları Çizelge 4.19’da verilmiştir.

4.14.1. En Yaygın Kullanılan Madde Seçim Yöntemleri

Çizelge 4.15’te en yaygın kullanılan madde seçim yöntemleri yüzde ve frekans bilgileriyle verilmiştir.

Çizelge 4.15. Çalışmalarda En Yaygın Kullanılan Madde Seçim Yöntemleri

	f	%
MFI	44	30.1
KLI	9	6.2
GDI	6	4.1
PWKL	5	3.4
MPWI	4	2.7
MFI-CB	4	2.7
MFI-EB	4	2.7
Diğer	60	41.1
Bilgi Verilmemiş	10	6.8

MFI: Maximum Fisher Information, KLI: Kullback-Leibler Information, GDI: Generalized Discrimination Index, PWKL: Posterior-Weighted Kullback-Leibler, MPWI: Maximum Posterior Weighted Information, MFI-CB: Maximum Fisher Information cut-point based, MFI-EB: Maximum Fisher Information estimated-ability based

Çizelge 4.15’e göre MFI yöntemi %30.1 oranıyla 44 kez kullanılarak en fazla kullanılan madde seçme yöntemi olduğu görülmekte ve belirgin şekilde öne çıkmaktadır. İkinci en sık kullanılan madde seçme yöntemi %6.2 oranıyla 9 kez

kullanılan KLI yöntemidir. GDI yöntemi 6 (%4.1) kez, PWKL yöntemi 5 (%3.4) kez, MPWI, MFI-CB ve MFI-EB yöntemleri dörder (%2.7) kez kullanılmıştır. 60 (%41.1) çalışmada tercih edilen madde seçim yöntemleri; rastgele madde seçimi ve maksimum bilgi tabakalama gibi çok yaygın olmayan yöntemlerin bulunduğu 'diğer' kategorisinde toplanmıştır. Çalışmaların 10'unda (%6.8) kullanılan madde seçim yöntemi hakkında herhangi bir bilgi verilmemiştir. Ayrıca çalışmalarda birden çok yöntem kullanılabildiğinden toplam sayı incelenen çalışma sayısından fazladır.

4.14.2. En Yaygın Kullanılan Madde Kestirim Yöntemleri

Çizelge 4.16'da en yaygın kullanılan madde seçim yöntemleri yüzde ve frekans bilgileriyle verilmiştir.

Çizelge 4.16. Çalışmalarda En Yaygın Kullanılan Madde Kestirim Yöntemleri

	f	%
ML	3	4.1
MMLE	2	2.7
JML	1	1.4
Bilgi Verilmemiş	67	91.8

ML: Maximum Likelihood, MMLE: Marginal Maximum Likelihood Estimation, JML: Joint Maximum Likelihood Estimation

Çizelge 4.16 incelendiğinde çalışmaların 67'sinde (%91.8) madde kestirim yöntemi belirtilmemiştir. Bilgi verilen 6 çalışmada ise ML yöntemi 3 kez (%4.1), MMLE yöntemi 2 kez (%2.7) ve JML yöntemi 1 kez (%1.4) kullanılmıştır.

4.14.3. En Yaygın Kullanılan Yetenek Kestirim Yöntemleri

Çizelge 4.17'de en yaygın kullanılan yetenek kestirim yöntemleri yüzde ve frekans bilgileriyle verilmiştir.

Çizelge 4.17. Çalışmalarda En Yaygın Kullanılan Yetenek Kestirim Yöntemleri

	f	%
EAP	48	42.5
MLE	31	27.4
MAP	13	11.5
WLE	4	3.5
Diğer	14	12.4
Bilgi Verilmemiş	3	2.7

EAP: Expected A Posteriori, MLE: Maximum Likelihood Estimation, MAP: Maximum A Posteriori, WLE: Weighted Likelihood Estimation

Çizelge 4.17'ye göre EAP yöntemi çalışmalarda 48 kez kullanılarak %42.5 oranıyla en sık tercih edilen yetenek kestirim yöntemi olmuştur. Bir diğer yaygın kullanım oranına sahip MLE yöntemi ise 31 (%27.4) kez kullanılmıştır. Ön bilgiye dayalı yetenek kestiriminde tercih edilen MAP yöntemi ise 13 (%11.5) çalışmada kullanılan üçüncü en sık kullanılan kestirim yöntemidir. WLE yöntemi de %3.6 oranıyla toplam 4 kez çalışmalarda kullanılmıştır. Bu yöntemlerden türetilen bazı yetenek kestirim yöntemleri de çalışmalarda az da olsa tercih edilmiş ve bunlar %12.4 oranıyla diğer kategorisinde ele alınmıştır. İncelenen çalışmalardan 3'ünde (%2.7) kullanılan yetenek kestirimleri hakkında herhangi bir bilgi vermemiştir.

4.14.4. En Yaygın Kullanılan Test Başlatma Kuralları

Çizelge 4.18'de en yaygın kullanılan test başlatma kuralları yüzde ve frekans bilgileriyle verilmiştir.

Çizelge 4.18. Çalışmalarda En Yaygın Kullanılan Test Başlatma Kuralları

	f	%
Nötr Teta Başlangıcı ($\theta=0$)	52	67.5
Rastgele Madde	6	7.8
Önsel Bilgi	1	1.3
Diğer	7	9.1
Bilgi Verilmemiş	11	14.3

Çizelge 4.18'e bakıldığında en sık kullanılan test başlatma kuralı %67.5 oranında 52 kez kullanılan ve $\theta=0$ olarak da ifade edilen nötr teta başlangıç kuralıdır. Çalışmaların 6'sı (%7.8) rastgele madde seçim kuralıyla, 1 (%1.3) çalışmanın da önsel bilgiyle teste başladığı görülürken diğer az bilinen teknikler diğer kategorisinde olup toplam 7 (%9.1) çalışmada kullanılmıştır. Çalışmaların 11'i (%14.3) kullanılan test başlatma kuralı hakkında herhangi bir bilgi vermemiştir.

4.14.5. En Yaygın Kullanılan Test Sonlandırma Kuralları

Çizelge 4.19'da en yaygın kullanılan test sonlandırma kuralları yüzde ve frekans bilgileriyle verilmiştir.

Çizelge 4.19. Çalışmalarda En Yaygın Kullanılan Test Sonlandırma Kuralları

	f	%
Standart Hata	42	40
Sabit Test Uzunluğu	39	37.1
Teta'ya Yakınsama	2	1.9
Sınıflandırma Ölçütü	2	1.9
Diğer	10	9.5
Bilgi Verilmemiş	10	9.5

Çizelge 4.19'da çalışmalarda kullanılan çeşitli test sonlandırma kurallarının dağılımı görülmektedir. Verilere göre standart hata eşik değeri 42 çalışmada kullanılmış ve %40 oranında en çok kullanılan test sonlandırma kuralı olmuştur. Sabit uzunluğa dahayı test sonlandırma kuralı ise %37.1 (39) oranıyla ikinci en sık kullanılan sonlandırma kuralıdır. Teta'ya yakınsama ve sınıflandırma ölçütü kuralı ise ikişer (%1.9) çalışmada kullanılmıştır. Diğerlerine göre pek yaygın kullanılmayan sonlandırma kuralları diğer kategorisinde bir araya getirilmiş ve 10 (%9.5) kez kullanılmıştır. İncelenen çalışmaların 10'unda (%9.5) sonlandırma kuralıyla ilgili herhangi bir bilgi sunulmamıştır.

4.15. Çalışmalarda Madde Kullanım Sıklığı ve İçerik Dengeleme Yapılma Oranları

Araştırmanın on beşinci alt problemi olan “Çalışmalarda madde kullanım sıklığı kontrolü ve içerik dengeleme yapılma oranları nedir?” sorusuna yönelik çalışmalarda madde kullanım sıklığı kontrolü ve içerik dengeleme yapılma durumları Çizelge 4.20’de yüzde ve frekans bilgileriyle verilmiştir.

Çizelge 4.20. Çalışmalarda Yapılan Madde Kullanım Sıklığı ve İçerik Dengeleme Oranları

		f	%
Madde Kullanım Sıklığı Kontrolü	Evet	26	32.9
	Hayır	10	12.7
	Bilgi Verilmemiş	37	47.4
İçerik Dengeleme	Evet	9	13.8
	Hayır	12	18.5
	Bilgi Verilmemiş	52	67.7

Çizelge 4.20’de görüldüğü üzere çalışmaların 37’inde (%47.4) madde kullanım sıklığının kontrol edilip edilmediği 52’sinde de (%67.7) içerik dengelemenin yapılıp yapılmadığı hakkında bilgi verilmemiştir. İncelenen çalışmalardan 26’sı (%32.9) madde kullanım sıklığı kontrolü yaparken 10 (%12.4) çalışma madde kullanım sıklığı kontrolü yapmamıştır. İçerik dengeleme yapan çalışma sayısının 9 (%13.8) olduğu görülürken 12 (%18.5) çalışmada içerik dengeleme yapılmamıştır. Veriler incelendiğinde genel olarak çalışmalarda madde kullanım sıklığı kontrolünün içerik dengeleme oranı yapan çalışma sayısına göre de fazla olduğu görülmektedir.

4.16. Çalışmaların Genel Sonuçlarına Göre Durumu

Araştırmanın on altıncı alt problemi ışığında “Çalışmaların genel sonuçları nedir?” sorusuna yanıt aranmış ve sonuçlar ortak noktalarda bir araya getirilerek Şekil 4.3’te görselleştirilmiştir



Şekil 4.3. Çalışmaların Genel Sonuçları

Şekil 4.3'e göre çalışmaların genel sonuçları incelendiğinde 18 çalışma BOBUT'un daha az maddeyle ölçüm yaptığını dolayısıyla zamandan da büyük oranda tasarruf sağladığı belirtilmektedir. BOBUT'tan alınan test sonuçlarının daha hassas ve güvenilir olduğunu belirten çalışma sayısı 23'tür. Ayrıca test sonuçlarına yönelik yapılan karşılaştırmaların 16'sında BOBUT'un yetenek ölçümünde daha doğru sonuçlar verdiği belirtilmiştir. İncelenen araştırmaların 32'sinde farklı algoritma ve yöntemlerin performansı da gözlemlenmiş avantaj ve dezavantajları değerlendirilmiştir. Bu kapsamda madde kalibrasyonu ve test uzunluğu konusunda 15 çalışma, geniş madde havuzunun ve test maddesinin arttırılmasının ölçme kesinliğini yüksek ve pozitif yönlü etkilediğini belirtmiştir.

5. TARTIŞMA

Araştırmanın birinci alt problemine yönelik elde edilen bulgulara bakıldığında Türkiye’de BOBUT konusunda ilk çalışmanın 1993 yılında yapıldığı ve özellikle 2010 yılından sonra bu alanda yapılan çalışmaların belirgin şekilde arttığı görülmektedir. Tarihsel süreçte BOBUT’un yaygınlaşmasında teknolojik gelişmeler önemli bir yer tutmaktadır. Ülkemizde de 2010 yılında fırsatları artırma ve teknolojiyi iyileştirme hareketi (FATİH) projesinin temelleri atılmış ve eğitimde teknoloji kullanımı o tarihten itibaren yaygınlaşmıştır. Öte yandan Tull ve Ayebe (2020)’in de belirttiği üzere pandemi dönemi uzaktan eğitim ve online değerlendirme süreçlerinde ilerleyişi zorunlu hale getirmiştir. Bu kapsamda BOBUT çalışmalarının 2010 yılından itibaren hızlı bir artış göstermesi hem FATİH projesi hem de pandemi döneminde oluşan ölçme değerlendirmeye alanında yeni bir yöntem ihtiyacıyla ilişkilendirilebilir. Ayrıca bu BOBUT çalışmalarının artışı bu dönemdeki akademik destek programlarıyla da ilişkilendirilebilir. Özellikle, Öğretim Üyesi Yetiştirme Programı (ÖYP), yüksek lisans ve doktora eğitimlerini destekleyerek akademik personelin yetişmesini teşvik etmektedir. ÖYP programı aracılığıyla doktora yapan araştırmacıların sayısındaki artış, BOBUT alanında nitelikli çalışmaların artmasına ve literatürde önemli bir yer edinmesine katkı sağlamış olabilir. Yurtçu ve Güzeller (2021)’in bibliyometrik analiz sonuçlarına paralel olarak son 10 yılda BOBUT alanında yapılan çalışma sayısının artması ölçme ve değerlendirme alanında teknolojik gelişmelerin hızla benimsendiğinin ve BOBUT’un bireyselleştirilmiş değerlendirmede önemli bir araç olarak görüldüğünün göstergesidir.

Araştırmanın ikinci alt problemine yönelik bulgular incelendiğinde BOBUT alanında yapılan çalışmalarda %54.8 oranıyla tezler ağırlıktadır. Aslında kaynak taramasında, erişilen makale sayıları tezlere göre daha fazla bulunmuş; ancak bu çalışmaların çoğu tezlerden tekrarlanan çalışmalar olduğu için analiz dışında bırakılmıştır. Tezlerde daha detaylı bilgi bulunduğundan, aynı çalışma hem tez hem makale olarak yayınlanmışsa, incelemeye tez olanı dahil edilmiştir. Bu da göstermektedir ki BOBUT konulu çalışmalar akademik alanda yoğun ilgi görmektedir. BOBUT süreci, algoritmalarının hazırlanmasında uzmanlık gerektirmesi, madde havuzu oluşturma sürecinin zorluğu, detaylı bir konu alanı hakimiyeti içerdiğinden

sıklıkla akademik çalışmalarda yer alması beklenen bir durumdur. Ancak yine de makale sayısı da %45.2 düzeyinde olup azımsanmayacak durumdadır. Bu durum tez dışında diğer bilimsel yayınlarda da BOBUT çalışmalarının hızlı bir şekilde yayıldığını göstermektedir.

Araştırmanın üçüncü alt problemine yönelik bulgulara bakıldığında, Türkiye’de yapılan BOBUT konulu çalışmaların en çok gerçek BOBUT uygulaması gerçekleştirmek ve sonuçları kâğıt kalem testleriyle karşılaştırmak amacıyla yapıldığı görülmektedir. Bu durum BOBUT’un geleneksel ölçme değerlendirme yöntemlerine göre avantajlarını ortaya koyma ve teknolojik gelişmelere ayak uydurma amacını yansıtmaktadır. Öyle ki BOBUT alanında Türkiye’de yapılan ilk çalışma olan Kaptan (1993)’ın çalışması da gerçek BOBUT uygulaması ve kâğıt kalem testi ile BOBUT’un karşılaştırmasını içermektedir. Aybek (2012), Kaya (2022), Kezer (2013), Yaşar (1999) gerçek zamanlı çalışmalarla BOBUT’un uygulanabilirliğini karşılaştıran önemli araştırmalardandır.

BOBUT’un teknik ve uygulama yöntemlerini geliştirmek adına yapılan test strateji ve parametrelerinin karşılaştırıldığı ve etkilerinin değerlendirildiği çalışmalar amaç bakımından bir diğer yoğun çalışılan konudur. Bu tür çalışmalar farklı yöntemlerin avantajlarını ve dezavantajlarını göstermek ve testlerin daha güvenilir sonuçlar vermesi sağlamak açısından önemlidir. Demir (2021), Kaplan (2015), Sarı (2016), Sulak (2013), Şahin (2017), Şahin ve Weiss (2015) bu alana katkı sağlamış önemli çalışmalardır.

Farklı BOBUT türlerinin denendiği çalışmalar test performanslarının BOBUT’un türüne ve uygulama amacına yönelik değişebileceğini gösteren bulgular içermektedir. Bu amaçla BOBUT’un çeşitli senaryolarda farklı türlerinin denenmesi öğrenci performansı ve testlerin etkisi açısından farklılıklar oluşturabilmektedir (Chen vd, 2020). Bu kapsamda çok aşamalı BOBUT, bilişsel tanıya dayalı BOBUT, çok boyutlu BOBUT, sınıflandırılmalı BOBUT türleri çalışmalarda eşit sayılarda denenmiş ve BOBUT’un amaca yönelik test sonuçlarındaki değişimleri gözlemlenmiştir. Farklı BOBUT türlerinin denenmesine yönelik çalışmaların oranının %17 gibi nispeten düşük bir seviyede olması, bu alanda daha fazla araştırma yapılması gerektiğini göstermektedir. Ancak, bu oran Türkiye’deki genel BOBUT çalışmalarının

oranına kıyaslandığında yeterli olarak değerlendirilebilir. Bu bulgu, Türkiye’de BOBUT’un henüz başlangıç aşamasında olduğunu ve araştırmaların çeşitlenmesi gerektiğini işaret etmektedir. Farklı BOBUT türlerinin daha fazla denenmesi, Türkiye’deki BOBUT konulu araştırmalarının genişlemesi ve çeşitlenmesi noktasında önemli bir adım olacağı düşünülmektedir. Bu tür çalışmalar, BOBUT’un çeşitli alanlarda ve farklı amaçlarla nasıl daha işlevsel kullanılabileceği konusunda alana büyük fayda sağlar. Örneğin, çok aşamalı BOBUT’un daha karmaşık yapıları ölçmede ve öğrenci motivasyonunu yükseltmede başarılı olduğu bilinmektedir. Aynı şekilde, bilişsel tanıya dayalı BOBUT, test katılımcılarının bilişsel becerilerini daha detaylı tespit ederek, öğrenme eksikliklerini analiz etmede önemli katkılar sunar. Bu nedenle, gelecekte yapılacak çalışmaların, BOBUT’un farklı türlerinin uygulanabilirliğini ve etkinliğini daha geniş örneklem grupları ve çeşitli eğitim seviyelerinde test etmesi gerekmektedir.

Son olarak özel alan ve konular için geliştirilen BOBUT uygulamaları, bu test türünün esnekliğini ve birçok alanda kullanılabilirliğini göstermek adına önemlidir. Bu tür çalışmalar BOBUT’un hızlı uyarlanabildiğini ve ilgili alanda sunduğu kolaylıkları vurgulayarak BOBUT’un yaygınlaşmasına katkı sağlamaktadır. Şenel (2018), görme engelliler üzerine geliştirdiği BOBUT ile herhangi bir yardım almadan özel gereksinimli bireylerin sınavlarını tamamlayabilmelerine olanak sağlarken, Elhan ve Öztuna (2008), sağlık alanında hastaların ortopedik rahatsızlıklarını ölçmek adına BOBUT’un potansiyel bir alternatif olduğunu belirtmiştir. Gür (2018) ile Çıkrıkçı ve arkadaşları (2020) sürücü adaylarının teorik sınavlarının BOBUT üzerinden yapılabilirliğini araştıran önemli araştırmalardandır.

Türkiye’de yapılan çalışmaların amaçları incelendiğinde dünya genelinde yapılan çalışmalar göz önüne alınarak mesleki yeterliklerin tespitinde ve sağlık alanında BOBUT’un daha az ilgi gördüğü ifade edilebilir. Bu durum araştırmacı türünün yoğunlukla eğitim alanında olmasından ya da ülkemizde bulunan öğrenci sayısının bu alandaki çalışmaları yönlendirmesinden kaynaklanabilir.

Araştırmanın dördüncü alt problemine yönelik bulgular incelendiğinde Türkçe olarak yayımlanan araştırmaların İngilizce olarak yayımlanan araştırmalara göre çoğunlukta olduğu görülmektedir. Çalışmaların yoğunlukla Türkçe yayımlanması

erişilebilirlik, anlaşılabilirlik ve yerel okuyucu kitlesine göre avantaj sağlamaktadır. Ancak uluslararası yayımlanan çalışmaların bilgiyi üretme, paylaşma ve global okuyucuya ulaştırma konusunda bilime olan katkısı daha fazladır (Altbach ve Knight, 2007). Bu sebeple İngilizce yayımlanan çalışmaların oran olarak azlığı uluslararası boyut göz önünde alındığında, yapılan çalışmaların tanınması, alana sunulan katkının bilinmesi, diğer araştırmacılarla bilgi paylaşımı yapılması ve diğer ülkelerle karşılaştırma yapılabilmesi noktasında önemli bir eksikliktir. Türkiye'deki çalışmalarda yayın dilinin dengeli dağılımı uluslararası literatüre daha fazla katkı sağlarken birikim ve etki alanını genişletecektir.

Araştırmanın beşinci alt problemi olan örneklem grupları çeşitliliğine yönelik bulgular incelendiğinde çalışmaların büyük oranda (%51.3) simülasyondan üretilmiş veri grubu örnekleme üzerinde yapıldığı görülmektedir. Bu durum çalışmaların gerçek uygulama yerine simülasyonda yapılmasından kaynaklanmaktadır. Diğer örneklem gruplara bakıldığında özellikle Canlı BOBUT uygulanan çalışmaların öğrenci (%34.2) grubu üzerinden yoğunlaştığı söylenebilir. Bu durum incelenen çalışmaların amaçlarıyla paralellik gösterirken, ilgili araştırmacıların sıklıkla eğitim alanından olmasından kaynaklanabilir. Örneklem grup olarak öğrencilerin tercih edilmesinin bir diğer yanı erişilebilir bir kitle olması ve test süreçlerine aşina oldukları için kolay uyum sağlayabilmeleri olarak değerlendirilebilir. Ayrıca akademik çalışmaların daha fazla olduğu göz önüne alındığında üniversite öğrenci grubunun örneklem olarak kullanılması olağan bir sonuçtur. Öğrenciler üzerinde yapılan araştırmalar BOBUT'un eğitim alanında uyarlanabildiğini ve avantajlarını gösteren önemli çalışmalardır.

Araştırmanın altıncı alt problemi doğrultusunda bulgulara bakıldığında Türkiye'deki çalışmaların sıklıkla (%70) BOBUT'un kendisi üzerinde yapıldığı görülmektedir. Wainer ve arkadaşları (2000b) genel BOBUT'un diğer türlere göre araştırmacıya uygulama kolaylığı sunduğunu ve büyük ölçekli gruplarda yetenek tespitinde daha çok tercih edildiğini belirtmiştir. Bununla birlikte çalışmalarda %30 oranında çok aşamalı BOBUT, çok boyutlu BOBUT, sınıflandırılmalı BOBUT gibi farklı BOBUT türleri de denenmiştir. Bu durum BOBUT'un kullanım amacına göre uyarlanabilir bir test aracı olduğunu, esnekliğini ve ölçme aracı olarak çok yönlü olduğunu göstermektedir.

Araştırmanın yedinci alt problemine yönelik bulgular incelendiğinde Türkiye’de yapılan çalışmaların %58.9’u simülasyonda, %41.1’i de canlı BOBUT uygulaması olarak gerçek örneklem grup üzerinde gerçekleştirildiği görülmüştür. Araştırma kapsamında incelenen 73 çalışmanın 43’ü simülasyon ortamında 30’u ise gerçek ortamda uygulanmıştır. Yapılan simülasyon çalışmalarının da 33’ü üretilmiş veriyle Monte Carlo simülasyonu ve 21’inin gerçek veriye dayalı yapılan post-hoc simülasyonu olduğu belirtilmiştir. Genel durumda simülasyon kullanım oranının yaygın olduğu görülmektedir. Lord (1980), canlı BOBUT uygulamalarının testlerin katılımcılar üzerinde daha güvenilir sonuçlar verdiğini belirtirken uygulama sürecinin daha zorlu olduğunu ifade etmiştir. Çalışmaların %32.8’i test stratejileri ve parametrelerin incelenmesi amacıyla gerçekleştirildiğinden simülasyon kullanım oranının bu denli yüksek olması simülasyon çalışmalarının hızlı sonuçlar vermesi ile araştırmalar için daha esnek imkanlar sunmasına dayandırılmaktadır. Çünkü simülasyon çalışmaları araştırmacıya yüksek örneklem gruplarında farklı algoritmaları ve parametreleri deneme konusunda büyük kolaylık sağlar. Örneğin farklı BOBUT türleri, yetenek kestirim yöntemleri, test sonlandırma kuralları test edilebilir. Simülasyon çalışmalarıyla gelecekte uygulanabilecek farklı yöntemler de test edilebilir. Canlı BOBUT gerçekleşmiş durumlara yani neden ve nasıl oldu gibi sorulara cevap verirken simülasyon çalışmaları olasılıklara dayalı gelecekte nasıl olur sorusunu yanıtlayabilir (Dooley, 2002). Gerçek örneklem grubuna testlerin uygulanmasında yaşanan zorluklar düşünüldüğünde simülasyon çalışmalarına olan eğilim anlaşılabilir. Simülasyon kullanımının dağılımı incelendiğinde araştırmalarda Monte Carlo simülasyonunun post-hoc simülasyonuna oranla daha sık kullanılmıştır. Bu durum araştırmalarda sanal olarak üretilmiş verilerden daha fazla faydalandığını göstermektedir. İncelenen çalışmalarda canlı BOBUT uygulamalarının azlığı düşünüldüğünde simülasyon çalışmalarının da yoğun olarak varsayımlara dayalı gerçeği yansıtmayan koşullarda yapılması araştırmaların güvenilirliği açısından soru işaretlerine sebep olabilir. Genel olarak simülasyon uygulamalarından edinilen veriler gerçek dünyadan bağımsız olarak elde edildiğinden güvenilirlik noktasında riskli bulunmuştur. Öyle ki simülasyon uygulamalarının testin gerçek performansını güvenilir bir şekilde yansıtmadığı Chang ve Harris (2002) tarafından da ifade

edilmiştir. Her iki uygulamanın da kendine göre avantajlı olduğu durumlar olabilmekle birlikte yapılan araştırmalarda uygulama yöntemlerinin dengeli olması gerekmektedir.

Yapılan çalışmalarda kullanılan uygulama yönteminin BOBUT türüne göre dağılımına bakıldığında farklı BOBUT türlerinin denendiği çalışmaların sıklıkla simülasyon çalışması olarak yapıldığı görülmektedir. Çok boyutlu BOBUT, çok aşamalı BOBUT, sınıflandırılmalı BOBUT gibi farklı türlerin canlı BOBUT olarak yapılmamış olması alanda ciddi eksiklik olarak değerlendirilebilir.

Araştırmanın sekizinci alt problemine yönelik bulgulara bakıldığında yazılımlar BOBUT uygulama yazılımları, simülasyon yazılımları ve madde kalibrasyon yazılımları olarak üç başlık halinde incelenmiştir. BOBUT uygulamalarında bazı araştırmacılar kendi geliştirdikleri yazılımları kullanmıştır (Kalender, 2011; Kezer 2013). Concerto (6), Firestar (5) ve SmartCAT (4) canlı BOBUT uygulamalarında öne çıkarken SimulCAT ve R paketleri simülasyon uygulamalarında sıklıkla tercih edilmiştir. Madde kalibrasyon yazılımında en fazla MULTILOG (7) yazılımından faydalanılmış olmakla birlikte yine R paketleri de yoğun olarak kullanılmıştır. BOBUT geliştirme ve uygulama sürecinde araştırmacılar tarafından çok çeşitli yazılımlar kullanılmıştır. Bu çeşitlilik farklı yazılımlardaki performansları deneme imkanı sunsa da araştırmaların tekrarlanabilirliğini ve karşılaştırılabilirliğini zayıflatmaktadır. İlave olarak araştırmalarda R temelli paketlerin sıklıkla tercih edilmesi açık kaynak kodlu olması ve kullanım kolaylığı sunmasından kaynaklanabilir.

Araştırmanın dokuzuncu alt problemine yönelik bulgular incelendiğinde örneklem grup dağılımları canlı BOBUT ve simülasyon çalışmaları olmak ayrı ayrı incelenmiştir. Canlı BOBUT uygulamalarında örneklem büyüklüğü 25-700 arasında değişmekle birlikte çalışmalar 25-100 (%53.8) aralığı olan küçük örneklem grubundan yoğunlaşmaktadır. Simülasyon çalışmaları ile canlı BOBUT uygulamalarının örneklem büyüklükleri arasında belirgin bir farklılık vardır. Simülasyon çalışmalarında kullanılan örneklem büyüklüğü 250-15000 gibi geniş bir aralıkta olup, çalışmaların çoğu 250-1000 (%40.6) aralığında olan örneklem büyüklüğünü kullanmıştır. BOBUT'un gerçek katılımcılara uygulanması hem örnekleme ulaşma açısından hem de uygulamanın zorluğu yönünden meşakkatli bir iştir. Bu durum göz

önüne alındığında gerçek uygulamalarda örneklem büyüklüğünün düşük katılımcılarla yapılması beklenen bir durumdur. Küçük örneklem gruplarıyla yapılan çalışmaların bulguların analizinde çok daha hassas davranılmalıdır. Ayrıca BOBUT'un test performansının sağlıklı incelenmesi için büyük örneklem gruplarında çalışılması önerilmektedir (Wainer, 2000a). Büyük örneklem grupları daha çeşitli veriler sağlarken araştırma sonuçlarının güvenilirliğine de olumlu etki eder (Kim ve Cohen, 1992). Küçük örneklem grupları araştırmanın genelleştirilmesini olumsuz etkilerken yanlışlık ve hata payını arttırabilir.

Araştırmanın onuncu alt problemine yönelik bulgulara bakıldığında madde havuzu büyüklüğünün çok geniş bir aralık olan 10-3000 arasında değiştiği görülmektedir. En sık kullanılan madde havuzu büyüklüğü 10-100 (%21.8) aralığı olmuştur. Bazı araştırmalar özellikle madde havuzu büyüklüğünün test performansına etkisini incelediğinden bu oranı etkilemiş olabilir. (Kaçar, 2016; Taymur, 2023; Öztürk ve Şahin, 2019). van der Linden ve Glas (2000) küçük madde havuzları testin ölçmek istediği yeteneği sınırlandırabildiğini ve testin güvenilirliğini olumsuz etkilediğini belirtmiştir. Araştırmacılar BOBUT'ta geçerlik ve güvenilirlik açısından kullanılması gereken en uygun madde havuzunun 300-500 aralığında olduğunu belirtmektedir (Thissen ve Mislevy, 2000). Bu bilgilere dayanarak Türkiye'de yapılan çalışmalarda kullanılan madde havuzu büyüklükleri yetersiz görülmüştür. Bu durum testin uyarlanabilir yapısını olumsuz etkilediği gibi madde havuzunun yetersizliği özellikle üst grup katılımcıların performansını ölçmede yetersiz kalabilir (Lord, 1980).

Araştırmanın on birinci alt problemine yönelik bulgular incelendiğinde, BOBUT çalışmalarında ölçülen yapının en çok matematik yeteneği olduğu ve bu yapının 9 kez tekrarlandığı görülmüştür. Bununla beraber İngilizce, sağlık ve engel düzeyi, mesleki ilgi ve yetenek gibi yapılar da bu havuzu çeşitlendirmektedir. Bu durum Türkiye'de yapılan BOBUT çalışmalarının çok farklı alanlara uyarlanabildiğini gösteren çalışmalar olduğunu göstermektedir. Matematik ve İngilizce gibi yapıların sıklıkla ölçülmesi gizil özellikler düşüldüğünde bu alanda soru hazırlama kolaylığından kaynaklanabilir. Mesleki ilgi ve yetenek, sağlık ve engel düzeyi gibi yapıların ölçülmesi de BOBUT'un çok fazla alana uyarlanabildiğinin bir göstergesidir. Bununla beraber çalışmaların 32'sinde ölçülen yapıların belirtilmemiş olması yine

BOBUT'un test stratejileri ve parametrelerini denemek amacıyla yapılmasından kaynaklanmaktadır. Çünkü bu tarz testlerde spesifik bir yapı belirtilmemektedir. Sonuç olarak BOBUT'un sıklıkla akademik yapıların ölçümünde kullanıldığı görüşmüştür. Çalışmaların dengelenmesi açısından farklı yapılar üzerine yapılan çalışmaların artması gerekmektedir.

Araştırmanın on ikinci alt problemine yönelik bulgulara bakıldığında çalışmalarda %42.2 oranla en çok tek boyutlu yapıların ölçüldüğü görülmüştür. Bu durum BOBUT'un MTK temelli olmasından kaynaklanabilir. Hambleton ve Swaminathan (1985)'e göre MTK varsayımlarından tek boyutluluk her bir maddenin tek bir boyutu ölçmesini gerektirir, bu da testin karmaşıklığını azaltarak sonuçlarının analiz edilmesi kolaylaştırır. İki ve daha fazla boyutun çok boyutlu olarak sınıflandırıldığı kategorinin oranı %18.7 olmuştur. Mulder ve van der Linden (2009)'e göre çok boyutlu yapılarda testin hazırlanması ve verilerin analizi fazla iş yükü getirir. Bu da çok boyutlu yapılar için BOBUT çalışmalarının daha az tercih edilmesine sebep olmuştur.

Araştırmanın on üçüncü alt problemine yönelik bulgular incelendiğinde Türkiye'de yapılan BOBUT konulu çalışmalarda MTK modellerinden %39.3 oranıyla en yaygın 3PL model kullanılmıştır. 1PL ve 2PL modeller hesaplamalar açısından daha basit olsa da ölçme hassasiyeti ve doğruluğu açısından dezavantajları vardır. 3PL model testlerde daha esnek bir yapı sunar çünkü madde güçlüğü, madde ayırt ediciliği ve madde yanıt olasılık parametrelerini içerir. Özellikle çoktan seçmeli testlerde 3PL modelin kullanımı testin güvenilirliğine katkı sağlar (Embretson ve Reise, 2000). BOBUT uygulamaların bu modelin kullanımı özellikle her birey için farklı zorlukta maddeler sunarak ve yanıt tahminlerini hesaplayarak testlerin bireyselleştirilmesini sağlar (Lord ve Novick, 1968). BOBUT katılımcılara bireye uyarlanmış bir test deneyimi sunar bu sebeple MTK modeliyle doğrudan ilişkilidir. 3PL model testlerde bireysel farklılıkları en çok yakalayabilen model olması sebebiyle testlerde daha çok tercih edilmiştir. Diğer yaygın modellerden DINA ve DINO özellikle bilişsel tanıya dayalı BOBUT çalışmalarında tercih edilmiştir. GRM ise tutum ve yetenek testlerinde tercih edilmiştir. Genel durum incelendiğinde MTK model seçiminin testin amacına uygun şekilde seçildiği görülmektedir.

Araştırmanın on dördüncü alt problemine yönelik bulgular incelendiğinde her bir yöntem ayrı ayrı değerlendirilmiştir. En yaygın kullanılan madde seçim yöntemi çok belirgin bir oranla (%30.1) 44 kez kullanılan MFI yöntemi olmuştur. MFI, her aşamada birey için en fazla bilgiyi verecek maddeyi havuzdan seçer, bu da bireyin yeteneğinin en yüksek seviyede tahmin edilmesine olanak sağlar. Daha az tercih edilen KLI, GDI gibi madde seçim yöntemleri farklı senaryolarda üretilen testler için tercih edilmiştir. Madde kestirim yöntemleri incelendiğinde çalışmaların %91.8'i bu konuda herhangi bir bilgi vermemiştir. Bu durum çalışmaların anlaşılabilirliğini ve değerlendirilmesini etkilemektedir. Yetenek kestirim yöntemlerinde ise en yaygın kullanılan yöntemin EAP (%42.5) olduğu görülmektedir. Çalışmaların örneklem grupları incelendiği 3.9. numaralı araştırma probleminin sonuçlarına göre Türkiye’de yapılan çalışmaların küçük örneklem gruplar üzerinde yoğunlaştığı görülmüştür. EAP yöntemi de daha çok küçük örneklem gruplarında kullanımının avantaj sağladığı bilindiğinden bu yöntemin sıklıkla tercih edilmesi olağandır. Çalışmalarda test başlatma kuralı olarak en yaygın nötr teta başlangıç kuralının (%67.5) kullanıldığı görülmektedir. Bu yöntem birey hakkında önceden herhangi bir bilginin olmadığı durumlarda adil bir şekilde testi başlatmak için kullanılır. Ancak diğer test başlatma yöntemlerin kullanım azlığı farklı test başlatma kurallarının denenmesi ve performanslarının detaylı şekilde değerlendirmesi konusunda eksiklik oluşturmaktadır. Test sonlandırma kurallarına bakıldığında, en sık tercih edilen sonlandırma kuralının 42 (%40) kez kullanılan standart hata sonlandırma kuralı olduğu görülmüştür. Bu yöntem standart hata değeri belli bir eşik değerinin altına düştüğünde testi sonlandırır. Ancak bu yöntem test uzunluğunu arttırabildiğinden çalışmalarda sıklıkla sabit uzunluk sonlandırma kuralıyla beraber tercih edilmiştir. Diğer az sayıda kullanılan sonlandırma kuralları testlerin amaçlarına göre değişmektedir.

Araştırmanın on beşinci alt problemine yönelik bulgular incelendiğinde analiz edilen 73 çalışmanın sadece 26’sının madde kullanım sıklığı kontrolü ve 9’unun ise içerik dengeleme yaptığı görülmüştür. Ayrıca çalışmaların % 47.4’sinde madde kullanım sıklığı kontrolü hakkında, %67.7’sinde de içerik dengeleme yapıp yapılmadığı konusunda herhangi bir bilgi vermemiştir. Madde kullanım sıklığı kontrolü bazı maddelere aşırı yüklenilmesini önleyerek havuzun verimli kullanımını sağlarken içerik dengeleme yapının ölçümüne uygun maddeleri dengeli şekilde

dağıtarak kapsam geçerliliğine etki eder. Dolayısıyla bu iki durumun kontrolü testin geçerliliğini ve güvenilirliğini arttırmak adına önemli bir yere sahiptir. Ancak Türkiye’de yapılan çalışmalar incelendiğinde bu konuya fazla önem verilmediği görülmektedir.

Araştırmanın on altıncı problemine yönelik bulgulara bakıldığında incelenen çalışmaların sonuçları ortak kategorilerde toplanmış ve genel dağılımları verilmiştir. Çalışmaların sonuçları incelendiğinde BOBUT uygulamaların test maddelerinden ve zamandan tasarruf sağladığı, ölçme sürecini daha güvenilir hale getirdiği, bireylerin yetenek ölçümünde daha doğru sonuçlar verdiği ve farklı yöntem ve stratejilerin amaca göre uygulandığında test performansını olumlu etkilediği gibi avantajlarından bahsedilmiştir. Çalışmaların genel sonuçları değerlendirildiğinde BOBUT geleneksel ölçme ve değerlendirme yöntemlerine iyi bir alternatif olabilir. BOBUT’un sıklıkla eğitim alanında tercih edilmesi diğer alanlarda çalışma sayısının nispeten az olması bu alanda bir eksiklik olarak değerlendirilebilir.

Teknolojik gelişmelerin takibi açısından makine öğrenmesi temelli BOBUT çalışmalarına yönelimin artırılması önemlidir. Literatür taramasında Türkiye’de makine öğrenmesi temelli BOBUT konulu bir çalışmaya erişilmiştir (Erbay Mermer, 2022). Ancak, bu test türü işleyiş, yöntem ve süreç bakımından farklı bir inceleme alanına odaklandığı için bu araştırma kapsamına dahil edilmemiştir. Türkiye’deki araştırmalar genellikle BOBUT’un teorik çerçevesi ve uygulamaları üzerine yoğunlaşmaktadır. Ayrıca, dünya genelinde özellikle Amerika ve Avrupa’da BOBUT ile uygulanan sınavların arttığı bilindiğinden, ülkemizdeki BOBUT çalışmalarının hala araştırma aşamasında olması üzücü bir durumdur.

6. SONUÇLAR

Bu araştırmada, Türkiye’de yapılan BOBUT konulu uygulama çalışmaları araştırma sorularına yanıt aramak için sistematik olarak incelenmiş ve sonuçlar analiz edilmiştir. Bu kapsamda Türkiye’de 1993-2024 yılları arasında araştırmanın dahil etme kriterlerine uygun 73 çalışma tespit edilerek analiz edilmiştir. Bu çalışmalar yıl, yazar, kaynak, yayın türü, çalışmanın adı, yayın dili, örneklem grup, kullanılan yazılımlar, BOBUT çeşitliliği, uygulama yöntemi, örneklem ve madde havuzu büyüklüğü, ölçülen yapıların sayısı ve türü, test başlatma kuralı, madde seçim yöntemi, madde kestirim yöntemi, yetenek kestirim yöntemi, test sonlandırma kuralı, içerik dengeleme ve madde kullanım sıklığı kontrolü, çalışmanın amacı ve genel sonucu şeklinde değişkenler kullanılarak veri dosyasına kodlanmıştır. Bu çalışma BOBUT’un ülkemizde yapılan çalışmalardaki durumu ve zamansal ilerleyişi hakkında genel ve kapsamlı bir değerlendirme sunmaktadır. Daha önce yapılan araştırmalara göre, BOBUT’un ölçme ve değerlendirme süreçlerinde önemli avantajlar sağladığı ve bireyselleştirilmiş değerlendirmelerde belirgin şekilde öne çıktığı görülmektedir (Redecker ve Punie, 2018; Wainer, 2000a). Bu bölümde Türkiye’de yapılan BOBUT konulu çalışmaların araştırmanın alt problemleri ele alınarak sonuçlarına yer verilmiştir.

Araştırmanın temel problemi olan “Türkiye’de BOBUT konulu yapılan araştırmalarda mevcut durum nedir, BOBUT geliştirme ve uygulama süreçlerinde hangi kapsamlı özellikler ve metodolojik yaklaşımlar benimsenmiştir?” sorusuna bu çalışmada alt problemler dahilinde yanıt aranmıştır. Bu kapsamda literatürdeki tüm çalışmalar detaylı olarak incelenmiş ve analiz edilmiştir. Son durumda çalışmaya dahil etme kriterlerine uygun 73 çalışmaya ulaşılmış ve erişilen bulgularla beraber sonuçlar aşağıda maddeler halinde özetlenmiştir.

- Türkiye’de BOBUT konulu çalışmalara eğilimin 2010 yılından sonra arttığı,
- Çalışmaların daha çok akademik alanda ilgi gördüğü,
- Çalışmaların sıklıkla test performanslarını incelemek ve gerçek BOBUT uygulamalarını kâğıt kalem testleri ile karşılaştırmak amacıyla yapıldığı,
- Yayın dilinde yaygın olarak Türkçe tercih edildiği,

- Çalışmalarda simülasyon verisinden sıkça yararlanılırken gerçek örneklem grup olarak öğrencilerin yoğunlukta olduğu,
- BOBUT'un farklı türleri denense de büyük oranda özellikle canlı test uygulamalarında BOBUT'un kendisine yönelik çalışmalar yapıldığı,
- Yazılım olarak çeşitlilik fazlayken Concerto, SimulCAT ve R gibi yazılımların öne çıktığı,
- Örneklem büyüklüklerinin canlı BOBUT ve simülasyon çalışmalarında değiştiği ayrıca çalışma sayısının küçük örneklem gruplarda yoğunlaştığı,
- Çalışmaların büyük kısmında madde havuzu büyüklüğü literatürün kabul gördüğü sınırın altında kaldığı,
- Ölçülen yapıların çeşitliliği fazlayken akademik yapılara daha çok yer verildiği,
- Tek boyutlu yapıların daha fazla ölçüldüğü,
- MTK modeli olarak en yaygın 3PL modelin kullanıldığı,
- Madde seçim yönteminde en yaygın MFI'nın kullanıldığı,
- Madde kestirim yöntemi hakkında çalışmaların büyük kısmının bilgi vermediği,
- Yetenek kestirim yönteminde en çok EAP'nin tercih edildiği,
- Test başlatma kuralı olarak en sık nötr teta başlangıcının seçildiği,
- Test sonlandırma kuralı olarak en fazla standart hata eşik değerinin kullanıldığı,
- Çalışmaların pek azında içerik dengeleme yapıldığı ve yine az sayıda çalışmada madde kullanım sıklığının kontrol edildiği,
- Genel sonuçlar değerlendirildiğinde BOBUT'un geleneksel ölçme değerlendirme yöntemlerine iyi bir alternatif olduğu ayrıca bazı durumlarda farklı test stratejilerinin avantaj sağladığı görülmüştür.

Bu araştırmanın bulgu ve sonuçları belirli sınırlılıklar çerçevesinde değerlendirilmelidir. Bu sınırlılıklar, araştırmanın kapsamını ve süreçlerini daha iyi anlamak adına aşağıda maddeler halinde ele alınmıştır.

- Çalışmalarda kullanılan tekniklerin isimlendirilmesinde literatürdeki standartlardan sapmalar görülmüştür. Bu durum, yöntemlerin ve tekniklerin

tutarlı bir şekilde analiz edilmesini zorlaştırmış, metodolojik tutarlılığı ve sonuçların karşılaştırılabilirliğini olumsuz yönde etkileyerek sonuçların anlamlı bir şekilde değerlendirilmesini kısıtlamıştır.

- İncelenen çalışmalarda yöntemlerin detaylı bir şekilde raporlanmaması, analiz sürecinde önemli bir engel teşkil etmiştir. Özellikle, çalışmalarda madde kestirim türüne dair bilgi eksikliği (%91.8), bu tür yöntemlerin analiz edilmesini ve karşılaştırılmasını oldukça güçleştirmiştir.
- Çalışmalarda kullanılan örneklem gruplarının büyük ölçüde öğrencilerden oluşması, elde edilen sonuçların genellenebilirliğini kısıtlamıştır. Farklı yaş grupları ve mesleklerden bireylerin dahil edilmemesi, sonuçların geniş bir popülasyona uygulanabilirliğini sınırlandırmıştır.
- Çalışma alanının ağırlıklı olarak eğitim üzerine yoğunlaşması, diğer alanlardaki uygulamaların eksik kalmasına yol açmıştır. Mesleki yeterlik ve sağlık gibi önemli alanlarda yapılan çalışmaların eksikliği, bulguların çeşitli alanlara uygulanabilirliğini sınırlandırmıştır.

7. ÖNERİLER

7.1. Gelecekte Yapılacak BOBUT Çalışmalarına Yönelik Öneriler

BOBUT'un pek çok konuda avantaj sağladığı ve geleceğin test teknolojisi olduğu araştırma bulgularıyla da ispatlanmıştır. Bu kapsamda BOBUT çalışmalarına yönelik araştırmacılara sunulan öneriler şunlardır:

1. Ülkemizde özellikle pandemi döneminde yaşanan uzaktan eğitim süreci çevrimiçi ortamda bir ölçme değerlendirme aracına olan ihtiyacı göstermiştir. BOBUT çalışmalarının uzaktan eğitim kapsamında çevrimiçi ortamlarda uygulanabilirliğine yönelik daha fazla çalışma yapılması önerilmektedir. Ayrıca bu kadar avantaja sahip bir test tekniğinin ulusal sınavlarda uygulama çalışmasının olması BOBUT'un bireyselleştirilmiş değerlendirmede büyük kitlelerde daha işlevsel bir ölçme değerlendirme aracı olduğunu ortaya koyabilir.
2. BOBUT alanında makale türünde, bağımsız çalışmaların artması adına yayın çeşitliği teşvik edilmelidir.
3. BOBUT uygulama amaçlarına bakıldığında sıklıkla geleneksel ölçme değerlendirme yöntemlerine alternatifliği araştırılmış farklı alanlara özellikle sağlık ve mesleki yeterlilik ölçümünde kullanılabilirliğine yönelik çalışmalar yetersiz kalmıştır. Yapılacak çalışmaların bu alanlara yoğunlaşması BOBUT'un esnekliğini göstermek adına önemlidir.
4. Türkiye'de yapılan çalışmalarda evrensel yayın dilinde çalışma sayısının artması uluslararası literatürde bilinirliği arttırmak ve karşılaştırılabilir çalışmalar sunmak için önemlidir.
5. Örneklem grup seçiminde sıklıkla öğrencilerin kullanılması ve eğitim alanına yoğunlaşılması BOBUT'un pek çok alana uyarlanabilirliğini göz ardı etmektedir. Bu sebeple farklı örneklem gruplar üzerinde de çalışmalar yapılması önerilmektedir.
6. BOBUT'un farklı türlerinin amaca uygun kullanıldığında ölçme değerlendirme sürecinde test performanslarını iyileştirdiği bilinmektedir. Bu kapsamda farklı BOBUT türlerinin kullanıldığı çalışmalar arttırılmalıdır. Ayrıca BOBUT

türlerinin denendiği çalışmalarda gerçek uygulama verilerine ihtiyaç duyulduğundan bu konuda çalışmalar önemsenmelidir.

7. Çalışmaların sıklıkla Monte Carlo simülasyonunda özellikle üretilmiş sanal veri kullanması, araştırma sonuçlarının genellenebilirliğini sınırlayabilmektedir. Çalışma verileri gerçek uygulamalarla desteklenmelidir.
8. BOBUT yazılımlarında bazı uygulamalar ön plana çıkmış ancak yerli uygulamaların kullanımına pek yer verilmemiştir. BOBUT uygulamaları için yazılım geliştirme çalışmalarını içeren araştırmalar desteklenmeli ve yerli yazılımlara çalışmalarda daha fazla yer verilmelidir.
9. Çalışmalarda kullanılan örneklem büyüklükleri araştırmanın genellenebilmesi adına yetersiz kalmaktadır. Daha büyük örneklem gruplar üzerinde yapılacak araştırma sayısı arttırılmalıdır.
10. Çalışmalarda madde havuzu büyüklükleri BOBUT'un performansını etkileyeceğinden testlerin geçerliğini ve güvenilirliğini arttırmak adına büyük madde havuzlarının kullanıldığı çalışmalara ihtiyaç vardır.
11. BOBUT'un karmaşık yapıları da rahatlıkla ölçebileceği göz önüne alındığında çok boyutlu yapıları ölçen çalışmalar arttırılmalıdır.
12. BOBUT'un farklı parametre ve yöntemler kullanıldığında test performansını artırabildiği bilindiğinden araştırmacılar yaygın kullanılan yöntemlere bağlı kalmadan farklı test senaryolarından da faydalanmalıdır.
13. Madde kullanım sıklığının kontrolü ve içerik dengelemenin yapılması madde havuzundan alınacak verimi arttıracığından araştırmacıların bu konuya daha fazla önem göstermesi gerekmektedir.

7.2. Benzer Konuda Çalışacak Araştırmacılara Yönelik Öneriler

1. Literatür tarama süreci çalışmanın önemli kısmını oluşturduğundan bu aşama geniş zamana yayılmalı ve çalışmalar titizlikle incelenmelidir.
2. Farklı test stratejileri ve parametrelerin incelendiği çalışmalar sayıca çok olduğundan bu çalışmaların daha detaylı analiz edilmesine yönelik çalışmalar yapılabilir.
3. Çalışmalarda simülasyon verisi kullanım oranlarının yüksekliğine yönelik çalışmalar yapılabilir.

4. BOBUT alanında kullanılan yazılımlar konulu bir çalışma alana yüksek katkı sağlayabilir. Özellikle yerli yazılımların incelenmesi bu yazılımların kullanımının yaygınlaşmasına da etki edecektir.
5. Bu çalışmada sadece tez ve makale türü incelendiğinde diğer yayın türlerini içeren çalışmalar yapılabilir.



KAYNAKLAR

- Alkan, V. (2021). *Mesleki Alan İlgi Envanteri'nin Bilgisayar Ortamında Bireye Uyarlanmış Formunun Geliştirilmesi*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Altbach, P. G., ve Knight, J. (2007). The Internationalization of Higher Education: Motivations and Realities. *Journal of Studies in International Education*, 11(3-4), 290-305.
- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38(1), 123-140.
- Aracı, F. G. İ., ve Tan, Ş. (2022). Multidimensional Computerized Adaptive Testing Simulations in R. *International Journal of Assessment Tools in Education*, 9(1), 118-137
- Arslan, Y. U. S. U. F., Çolak, M., ve Bilgin, U. (2022). Uzaktan Eğitimde Bilgisayar Uyarlamalı Test ve Kağıt-Kalem Testi ile Yetenek Düzeylerinin Belirlenmesi: Çok Merkezli Bir Çalışma. *Tıp Eğitimi Dünyası*, 21(63)
- Arslan, Y.K. (2021). *Bilgisayar uyarlamalı testlerde madde yanıt sürelerinin kullanımı*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Assessment Systems Corporation. (2004). *FastTest*. Company documentation. <https://assess.com/fasttest/> adresinden 30.05.2024 tarihinde erişildi.
- Aşiret, S. (2019). *Bilişsel tanıya dayalı bilgisayar ortamında bireye uyarlanmış testlerde madde seçim yöntemlerinin incelenmesi*. [Yayımlanmamış doktora tezi]. Mersin Üniversitesi
- Aşiret, S., ve Sünbül, S. Ö. (2022). BT-BOBUT Uygulamalarında Madde Seçim Algoritmalarının Performanslarının Ölçme Doğruluğu Açısından İncelenmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, (54), 188-214
- Ayan, C. (2018). *Bilişsel tanı modelinde geleneksel ve bilgisayarlı sınıflamalı test uygulamalarının psikometrik özelliklerinin karşılaştırılması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Aybek, E. C. (2012). *Kâğıt-kalem formu ve bilgisayar ortamında uygulanan genel yetenek testinin psikometrik özelliklerinin karşılaştırılması*. [Yayımlanmamış Yüksek Lisans Tezi]. Ankara Üniversitesi
- Aybek, E. C. (2016). *Kendini değerlendirme envanterinin Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT) olarak uygulanabilirliğinin araştırılması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- 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), 23-27.
- Aybek, E. C., ve Demirtaşlı, R. N. (2014). Bilgisayar Ortamında ve Kâğıt-Kalem Formunda Uygulanan Genel Yetenek Testinin Psikometrik Özelliklerinin Karşılaştırılması. *İlköğretim Online*, 13(4).
- Ayhan, A. S. (2015). *Comparability of scores from cat and paper and pencil implementations of student selection examination to higher education* [Yayımlanmamış yüksek lisans tezi]. Bilkent Üniversitesi
- Baker, F. B. (2001). *The basics of item response theory (2nd ed.)*. ERIC Clearinghouse on Assessment and Evaluation.
- Balta, E., ve Arzu, U. Ç. A. R. (2022). Bilgisayar Ortamında Bireye Uyarlanmış Test Uygulamalarında Ölçme Kesinliğinin ve Test Uzunluğunun Farklı Koşullar

- Altında İncelenmesi/Investigation of Measurement Precision and Test Length in Computerized Adaptive Testing Under Different Conditions. *e-Uluslararası Eğitim Araştırmaları Dergisi*, 13(1), 51-68.
- Baltacı, A. (2019). Nitel araştırma süreci: Nitel bir araştırma nasıl yapılır? *Ahi Evran Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 5(2), 368-388.
- Bejar, I. I. (1983). *Achievement testing: Recent advances*. London: Sage Publications.
- Bellibaş, M., ve Gümüş, S. (2018). Eğitim yönetiminde sistematik derleme çalışmaları. *Eğitim Yönetimi Dergisi*, 24(3), 345-367. https://doi.org/xx.xxx/yyyyEğitim_yönetiminde_arastırma_içinde (ss. 507-575). Ankara: Pegem Akademi.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F. M., ve Novick, M. R. (Eds.), *Statistical theories of mental test scores* (ss. 397-424). Reading, MA: Addison-Wesley.
- Bischof, M. (2009). catIRT: Computerized adaptive testing with R. *R package documentation*. <https://cran.r-project.org/web/packages/catIRT/>
- Bock, R. D. (1972). *Estimating item parameters and latent ability when responses are scored in two or more nominal categories*. *Psychometrika*, 37(1), 29-51.
- Bock, R. D., ve Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443-459.
- Bock, R. D., ve Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6(4), 431-444.
- Bostan, A. (2020). Using Fuzzy Logic in Computerized Adaptive Tests. *Bilişim Teknolojileri Dergisi*, 13(3), 289-300.
- Bulut, O., ve Kan, A. (2012). Application of computerized adaptive testing to entrance examination for graduate studies in Turkey. *Eurasian Journal of Educational Research*, 49, 61-80.
- Chalmers, P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*. <https://www.jstatsoft.org/article/view/v048i06>
- Chang, H. H., ve Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20(3), 213-229.
- Chang, H., ve Harris, D. J. (2002). The Impact of Data Simulation on the Development of Computerized Adaptive Testing Systems. *Educational and Psychological Measurement*, 62(5), 787-800.
- Chen, H., Chen, Y., ve Chen, M. Y. (2020). Simulation study on the efficiency of cognitive diagnostic computerized adaptive testing for grasping state standards. *Eurasia Journal of Mathematics, Science and Technology Education*, 16(10), em1881.
- Chen, S. Y., ve Ankenmann, R. D. (2004). Effects of practical constraints on item selection rules at the early stages of computerized adaptive testing. *Journal of Educational Measurement*, 41(2), 149-174.
- Choi, S. W. (2009). Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. *Applied Psychological Measurement*, 33(8), 644.
- Choi, S. W., Reise, S. P., Pilkonis, P. A., Hays, R. D., ve Cella, D. (2010). Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. *Quality of Life Research*, 19, 125-136.

- Cıkrıkçı, N., Yalçın, S., Kalender, I., Gül, E., Ayan, C., Uyumaz, G., ... ve Kamis, O. (2020). Development of a Computerized Adaptive Version of the Turkish Driving Licence Exam. *International Journal of Assessment Tools in Education*, 7(4), 570-587.
- Concerto. (2008). University of Cambridge Psychometrics Centre. *Concerto platform documentation*. <https://www.psychometrics.cam.ac.uk/newconcerto/concerto-platform>
- Cook, T. M., ve Eignor, D. R. (1991). IRT-based multistage testing in certification and licensure testing programs. *Advances in Educational and Psychological Testing*.
- Cömert, M. (2008). *Bireye uyarlanmış bilgisayar destekli ölçme ve değerlendirme yazılımı geliştirilmesi*. [Yayımlanmamış yüksek lisans tezi]. Bahçeşehir Üniversitesi
- Creswell, J. W. (2013). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Sage Publications.
- Crocker, L., ve Algina, J. (2006). Introduction to classical and modern test theory. *Introduction to classical and modern test theory*.
- Çetin, G., Ertepinar, H., ve Geban, Ö. (2004). Developing and Implementing an Instructional Technology Aided Conceptual Change Approach in Teaching Ecology Concepts at Ninth Grade. *Turkish Online Journal of Educational Technology-TOJET*, 3(1), 27-31.
- Cıkrıkçı, N., Yalçın, S., Kalender, I., Gül, E., Ayan, C., Uyumaz, G., ... ve Kamis, O. (2020). Development of a Computerized Adaptive Version of the Turkish Driving Licence Exam. *International Journal of Assessment Tools in Education*, 7(4), 570-587.
- Çoban, E. (2020). *Bilgisayar temelli bireyselleştirilmiş test yaklaşımlarının Türkiye'deki merkezi dil sınavlarında uygulanabilirliğinin araştırılması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Demir, S. (2018). *Çok kategorili bireyselleştirilmiş bilgisayarlı test uygulamalarının farklı madde seçim yöntemlerinde sonlandırma kuralları açısından incelenmesi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Demir, S., Aktan, D. Ç., ve Güler, N. (2023). Application of the professional maturity scale as a computerized adaptive testing. *International Journal of Assessment Tools in Education*, 10(3), 580-593.
- Demir, S., ve Atar, B. (2021). Investigation of classification accuracy, test length and measurement precision at computerized adaptive classification tests. *Journal of Measurement and Evaluation in Education and Psychology*, 12(1), 15-27.
- Deng, N., Wells, C., & Hambleton, R. (2008). A confirmatory factor analytic study examining the dimensionality of educational achievement tests. *Educational and Psychological Measurement*, 68(5), 695-709.
- Dirlik, E. M., Altıntaş, Ö., ve Kartal, S. (2021). Uluslararası test standartlarının ve yönergelerinin sistematik ve karşılaştırmalı incelenmesi. *OPUS International Journal of Society Researches*, 18(39), 804-830.
- Dooley, K. E. (2002). Simulation in computer-based testing: An introduction. *Journal of Applied Testing Technology*, 3(1), 1-12
- Elhan, A. H., Öztuna, D., Kutlay, Ş., Küçükdeveci, A. A., ve Tennant, A. (2008). An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain. *BMC musculoskeletal disorders*, 9, 1-15.

- Embretson, S. E. (1991). A cognitive design system approach to constructing adaptive tests. *Psychometrika*.
- Embretson, S. E., ve Reise, S. P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates.
- Erbay Mermer, Ş. (2022). *Minnesota Çok Yönlü Kişilik Envanteri İçin Makine Öğrenmesi Temelli Bireyselleştirilmiş Bilgisayarlı Test Uygulamasının Geliştirilmesi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Erdem Kara, B. (2019). *Değişen Madde Fonksiyonu Gösteren Madde Oranının Bireyselleştirilmiş Bilgisayarlı ve Çok Aşamalı Testler Üzerindeki Etkisi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Eroglu, M. G. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının ölçme kesinliği ve test uzunluğu açısından karşılaştırılması*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Ertaş Polat, F. G. (2022). *Çok Aşamalı Bireye Uyarlanmış Testlerde Farklı Koşullardan Elde Edilen Yetenek Kestirimlerinin Karşılaştırılması*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Ertuna, L. (2020). *The Effect of Item Selection and Parameter Estimation Methods to the Accuracy of Pretest Item Parameters on Online Calibration in CAT*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Fisher, R. A. (1925). Theory of statistical estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 22, 700-725.
- Frey, A., Liu, T., Fink, A., ve König, C. (2024). Meta-analysis of the effects of computerized adaptive testing on the motivation and emotion of examinees. *European Journal of Psychological Assessment*.
- Gausden, E. B., Levack, A., Nwachukwu, B. U., Sin, D., Wellman, D. S., ve Lorich, D. G. (2018). Computerized adaptive testing for patient reported outcomes in ankle fracture surgery. *Foot ve ankle international*, 39(10), 1192-1198.
- Gelibolu, M. F. (2015). *Temel bilgisayar eğitimi alanında bilgisayarlı uyarlanabilir sınavların ölçme özelliklerinin diğer sınav türleriyle karşılaştırılması* [Yayımlanmamış doktora tezi]. Gazi Üniversitesi
- Georgiadou, E., Triantafillou, E., and Economides, A. A. (2007). A Review of Item Exposure Control Strategies for Computerized Adaptive Testing Developed from 1983 to 2005. *Journal of Technology, Learning, and Assessment*, 5(8), n8.
- Gibbons, R. D., Weiss, D. J., Frank, E., ve Kupfer, D. (2016). Computerized adaptive diagnosis and testing of mental health disorders. *Annual review of clinical psychology*, 12, 83-104.
- Gökcan, M. (2023). *İngilizce kelime bilgisi testinin (VST) bireyselleştirilmiş bilgisayarlı test olarak uygulanabilirliğinin incelenmesi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Gökçe, S. (2012). *Comparison of linear and adaptive versions of the Turkish pupil monitoring system (pms) mathematics assessment*. [Yayımlanmamış doktora tezi]. Orta Doğu Teknik Üniversitesi
- Gökçe, S., ve Glas, C. A. (2018). Can TIMSS mathematics assessments be implemented as a computerized adaptive test. *Journal of Measurement and Evaluation in Education and Psychology*, 9(4), 422-436.
- Gören, S., Kara, H., Kara, B. E., ve Kelecioğlu, H. (2022). The Effect of Aberrant Responses on Ability Estimation in Computer Adaptive Tests. *Journal of Measurement and Evaluation in Education and Psychology*, 13(3), 256-268.

- Gündeğer, C., ve Doğan, N. (2018). Bireyselleştirilmiş Bilgisayarlı Sınıflama Testi Kriterlerinin Test Etkililiği ve Ölçme Kesinliği Açısından Karşılaştırılması. *Journal of Measurement and Evaluation in Education and Psychology*, 9(2), 161-177.
- Güneş, M. (2018). Teknolojinin ölçme ve değerlendirme süreçlerine etkisi. *Eğitimde Teknoloji Dergisi*, 4(1), 10-20
- Gür, R. (2017). *Bireye uyarlanmış testlerde madde kullanım sıklığı kontrol yöntemlerinin farklı koşullarda ölçme duyarlılığına ve test güvenliğine etkisi*. [Yayınlanmamış doktora tezi]. Ankara Üniversitesi.
- Gür, R., ve Karabay, E. (2018). Motorlu Taşıtlar Sürücü Kursiyerleri Sınavının Simülatif Bilgisayar Ortamında Bireye Uyarlanmış Test Uygulaması. *Journal of Theoretical Educational Science*, 11(2), 201-228.
- Hambleton R. K., Swaminathan, H. ve Rogers, H. J. (1991b). *Fundamentals of item response theory*. California: Sage Publications, Inc.
- Hambleton, R. K., ve Rogers, H. J. (1991a). Advances in criterion-referenced measurement. In *Advances in educational and psychological testing: Theory and applications* (ss. 3-43). Dordrecht: Springer Netherlands.
- Hambleton, R. K., ve Swaminathan, H. (1983). *Item Response Theory: Principles and Applications*. Springer.
- Hambleton, R. K., ve Swaminathan, H. (1985). Item Response Models. In *Item Response Theory: Principles and Applications*, 33-52. Dordrecht: Springer Netherlands.
- Hammad, W., ve Hallinger, P. (2017). A systematic review of conceptual models and methods used in research on educational leadership and management in Arab societies. *School Leadership and Management*, 37(5), 434-456.
- He, W. (2011). Firestar: Computerized adaptive testing simulation program. *Educational and Psychological Measurement*.
- Higgins, J. P. T., ve Green, S. (2011). Cochrane handbook for systematic reviews of interventions. <http://www.mrc-bsu.cam.ac.uk/cochrane/handbook/> adresinden 25.04.2024 tarihinde erişildi
- IACAT. (2022). International Association for Computerized Adaptive Testing. <https://iacat.org/> adresinden 4 Mayıs 2024 tarihinde erişildi
- Ita, M. E., Kecskemety, K. M., Ashley, K. E., ve Morin, B. (2014, June). Comparing student performance on computer-based vs. paper-based tests in a first-year engineering course. In *2014 ASEE Annual Conference ve Exposition* (ss. 24-297).
- Jeong, H. (2014). A comparative study of scores on computer-based tests and paper-based tests. *Behaviour ve Information Technology*, 33(4), 410-422.
- Jimenez, L., ve Boser, U. (2021). Future of testing in education: Artificial intelligence. *Center for American Progress*. <https://www.americanprogress.org/article/future-testing-education-artificial-intelligence>.
- Kaçar, M. (2016). *Bireye uyarlanmış bilgisayarlı test ortamında maksimum fisher madde seçme yönteminin incelenmesi* [Yayınlanmamış yüksek lisans tezi]. Necmettin Erbakan Üniversitesi
- Kalender, I., ve Berberoglu, G. (2017). Can computerized adaptive testing work in students' admission to higher education programs in Turkey?. *Educational Sciences: Theory ve Practice*, 17(2).

- Kalender, İ. (2009). Başarı ve yetenek kestirimlerinde yeni bir yaklaşım: Bilgisayar ortamında bireyselleştirilmiş testler (Computerized adaptive tests-CAT). *CITO Eğitim Kuram ve Uygulama*, 5, 39-48.
- Kalender, İ. (2011). *Effects of different computerized adaptive testing strategies on recovery of ability*. [Yayımlanmamış doktora tezi]. Ortadoğu Teknik Üniversitesi
- Kalender, İ. (2012). Yükseköğretime öğrenci seçmek için bilgisayar ortamında bireyselleştirilmiş testler. *Yükseköğretim Dergisi*, 2(1), 13-19.
- Kan, A. (2011). Test eşitleme: OKS testlerinin istatistiksel eşitliğinin sınanması. *Eğitim ve Bilim*, 36(160).
- Kaplan, M. (2016). Nitelik Sayısının Madde Seçme Algoritmalarının Performansı Üzerindeki Etkisi. *Journal of Measurement and Evaluation in Education and Psychology*, 285-295.
- Kaplan, M., de la Torre, J., ve Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied psychological measurement*, 39(3), 167-188.
- Kaplan, M., ve de la Torre, J. (2020). A Blocked-CAT Procedure for CD-CAT. *Applied Psychological Measurement*, 44(1), 49-64.
- Kaptan, F. (1993). Yetenek kestiriminde adaptive (bireyselleştirilmiş) test uygulaması ile geleneksel kâğıt-kalem testi uygulamasının karşılaştırılması. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi.
- Kaskatı, O. T. (2011). *Rasch modelleri kullanarak romatoid artirit hastaları özürüllük değerlendirimi için bilgisayar uyarlamalı test yönteminin geliştirilmesi*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Kaya, E. (2022). *A Comparability and Classification Analysis of Computerized Adaptive And Conventional Paper-Based Versions of an English Language Proficiency Reading Subtest*. [Yayımlanmamış doktora tezi]. Bilkent Üniversitesi
- Kezer, F. (2013). *Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Kezer, F. (2021). The effect of item pools of different strengths on the test results of computerized-adaptive testing. *International Journal of Assessment Tools in Education*, 8(1), 145-155.
- Kezer, F., ve Nizamettin, K. O. Ç. (2014). Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması. *Eğitim Bilimleri Araştırmaları Dergisi*, 4(1), 145-174.
- Kim, S. H., ve Cohen, A. S. (1992). Effects of linking methods on detection of DIF. *Journal of educational measurement*, 29(1), 51-66.
- Kingsbury, G. G., ve Hauser, C. (2004). Computerized Adaptive Testing and "No Child Left Behind". *Northwest Evaluation Association*.
- Kingsbury, G. G., ve Houser, R. L. (1988). A comparison of achievement level estimates from computerized adaptive testing and paper-and-pencil testing. In *annual meeting of the American Educational Research Association*, New Orleans, LA.
- Kingsbury, G. G., ve Weiss, D. J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In *New horizons in testing* (ss. 257-283). Academic Press.
- Kolen, M. G., ve Brennan, R. L. (1995). WinGen: A Windows-based program for generating item response data. *American Psychological Association*.

- Koşan-Aytuğ, M. (2013). *Tıp eğitiminde gelişim sınavı soru bankası oluşturulması ve benzetim verileri ile bilgisayar uyarlamalı test uygulaması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Kullback, S., ve Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79-86.
- Kürşad, M. Ş. (2020). *Madde parametre sapmasının bireye uyarlanmış testlerden elde edilen yetenek kestirimlerine etkisinin farklı koşullar altında incelenmesi*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Kürşad, M. Ş. (2023) The Effects of Different Item Selection Methods on Test Information and Test Efficiency in Computer Adaptive Testing. *Journal of Measurement and Evaluation in Education and Psychology*, 14(1), 33-46.
- Leighton, J. P., ve Gierl, M. J. (2007). *Cognitive Diagnostic Assessment for Education: Theory and Applications*. Cambridge University Press.
- Linacre, J. M. (1998). Winsteps: Rasch-model computer programs. MESA Press.
- Linacre, J. M. (2000). Computer-adaptive testing: A methodology whose time has come (No. 69, p. 58). *MESA memorandum*.
- Lincoln, Y. S., ve Guba, E. G. (1985). *Naturalistic inquiry*. sage.
- Ling, G., Attali, Y., Finn, B., ve Stone, E. A. (2017). Is a computerized adaptive test more motivating than a fixed-item test? *Applied psychological measurement*, 41(7), 495-511.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates.
- Lord, F. M., ve Novick, M. R. (1968). *Statistical theories of mental test scores*. IAP.
- Magis, D., Béland, S., Tuerlinckx, F., ve De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior research methods*, 42(3), 847-862.
- Magis, D., Raiche, G., ve Magis, M. D. (2018). Package 'catR'. *R package*, version, 3.
- Magis, D., ve Raîche, G. (2012). catR: An R package for computerized adaptive testing [Software]. R package version 3.0. <https://cran.r-project.org/web/packages/catR/> adresinden 01.05.2024 tarihinde erişildi.
- Magis, D., Yan, D., ve Von Davier, A. A. (2017). *Computerized adaptive and multistage testing with R: Using packages catR and mstR*. Springer.
- Mao, X., Ozdemir, B., Wang, Y., ve Xin, T. (2018). Investigating The Effect of Exposure-Control Strategies on Item Selection Methods in MCAT. *Journal of Measurement and Evaluation in Education and Psychology*, 9(3), 295-311.
- Meuleman, B., ve Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM?. In *Survey Research Methods* (Vol. 3, No. 1, ss. 45-58).
- Meyer, J. P. (2010). CATsim: An R package for simulating computerized adaptive testing. *Journal of Statistical Software*. <https://www.jstatsoft.org/article/view/v036i08>
- Microsoft Corporation. (2024). Microsoft Office 365 [Computer software]. Retrieved from <https://www.microsoft.com>
- Miles, M. B., ve Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. sage.
- Milli Eğitim Bakanlığı. (2016). *Özel eğitim hizmetleri tanıtım el kitabı*. Özel Eğitim Rehberlik ve Danışma Hizmetleri Genel Müdürlüğü. Web:

- http://mebk12.meb.gov.tr/meb_iys_dosyalar/06/14/747031/dosyalar/2013_01/11100900_zeleitimelkitab.pdf adresinden 24.04.2024 tarihinde erişildi.
- Milli Eğitim Bakanlığı. (2024). *Geçmişten günümüze sayılarla eğitim*. <http://www.meb.gov.tr> adresinden 22.03.2024 tarihinde erişilmiştir.
- Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika*, 51, 177-195.
- Mislevy, R. J., ve Bock, R. D. (1982). BILOG: Item analysis and test scoring with binary logistic models. *Research Report*.
- Mislevy, R. J., ve Stocking, M. L. (1989). A consumer's guide to LOGIST and BILOG. *Applied psychological measurement*, 13(1), 57-75.
- Moher, D., Liberati, A., Tetzlaff, J., ve Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology*, 62(10), 1006–1012.
- Moshinsky, A., ve Kazin C. (2005). Constructing a computerized adaptive test for university applicants with disabilities. *Applied Measurement In Education*, 18(4), 381-405.
- Mulder, J., ve Van der Linden, W. J. (2009). Multidimensional adaptive testing with Kullback–Leibler information item selection. In *Elements of adaptive testing* (ss. 77-101). New York, NY: Springer New York.
- Muthén, B. O., ve Asparouhov, T. (2012). OSCATS: Software for ordered categorical analysis. *Mplus Web Notes: No. 21*. <https://www.statmodel.com/oscats.shtml>
- Nalbantoglu Yilmaz, F. (2021). Comparing Paper-Pencil and Computer-Based Tests: A Meta-Analysis Study in the Sample of Turkey. *Eurasian Journal of Educational Research*, 93, 279-300.
- Norris, J. M., ve Ortega, L. (2006). *Synthesizing research on language learning and teaching*. John Benjamins Publishing.
- Nydic, S. W., ve Wiberg, M. (2012). catIRT: An R package for implementation of computerized adaptive testing with the generalized partial credit model. *Applied Psychological Measurement*, 36(1), 71-73. <https://doi.org/10.1177/0146621611430815>
- Ostini, R., ve Nering, M. L. (2006). *Polytomous item response theory models* (No. 144). Sage.
- Ozyurt, O. (2014). The Classification Of The Probability Unit Ability Levels Of The Eleventh Grade Turkish Students By Cluster Analysis. *Turkish Online Journal of Distance Education*, 15(2), 145-160.
- Öz, H., ve Özturan, T. (2018). Computer-based and paper-based testing: Does the test administration mode influence the reliability and validity of achievement tests?. *Journal of Language and Linguistic Studies*, 14(1), 67-85.
- Özbaşı, D. (2014). *Bilgisayar okuryazarlığı testinin bilgisayar ortamında bireye uyarlanmış test olarak uygulanabilirliğine ilişkin bir araştırma*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Özdemir, B. (2015). *Madde düzeyinde boyutluluk modellerinin bilgisayar ortamında bireyselleştirilmiş test yöntemleri üzerindeki etkisinin incelenmesi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Özdemir, B. (2016). Comparison of different unidimensional-CAT algorithms measuring students' language abilities: Post-hoc simulation study. *European Proceedings of Social and Behavioural Sciences*.

- Özdemir, B., ve Gelbal, S. (2022). Measuring language ability of students with compensatory multidimensional CAT: A post-hoc simulation study. *Education and Information Technologies*, 27(5), 6273-6294.
- Özgür, E. G. (2020). *Rasch Modeline uygun yanıt desenlerinde bilgisayar uyarlamalı test yöntemi ile makine öğrenmesi yöntemlerinin tahmin performanslarının benzetim çalışması ile karşılaştırılması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Öztuna, D. (2008). *Kas-iskelet sistemi sorunlarının özürlülük değerlendiriminde bilgisayar uyarlamalı test yönteminin uygulanması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Öztürk, N. B. (2014). *Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin İncelenmesi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Öztürk, N. B., ve Şahin, M. G. (2019). Effects of Item Pool Characteristics on Ability Estimate and Item Pool Utilization: A Simulation Study. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 34(2), 473-486.
- Özyurt, H., Özyurt, Ö., Baki, A., ve Güven, B. (2012). Integrating computerized adaptive testing into UZWEBMAT: Implementation of individualized assessment module in an e-learning system. *Expert Systems with Applications*, 39(10), 9837-9847.
- Park, C. (2011). IRTPARAM: An R package for parameter estimation in item response theory. *R package documentation*. <https://cran.r-project.org/web/packages/IRTPARAM/>
- Patz, R. J., & Junker, B. W. (1999). *A straightforward approach to Markov chain Monte Carlo methods for item response models*. *Journal of Educational and Behavioral Statistics*, 24(2), 146-178.
- Pearson Education. (1994). Pearson VUE: Virtual university enterprises. *Company website*. <https://www.pearsonvue.com/>
- Pearson VUE. (2002). LIVECAT: Live computerized adaptive testing platform. *Company website*. <https://www.pearsonvue.com/>
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., ve Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press.
- Prometric Inc. (1994). Prometric: Computerized adaptive testing services. *Company website*. <https://www.prometric.com/>
- Raborn, A., ve Sarı, H. (2021). Mixed Adaptive multistage testing: A new approach. *Journal of measurement and evaluation in education and psychology*, 12(4), 358-373.
- Rao, C. R. (1973). *Linear statistical inference and its applications*. John Wiley ve Sons.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Reckase, M. D. (2009). Multidimensional item response theory models. In *Multidimensional item response theory* (ss. 79-112). Springer, New York, NY.
- Redecker, C., ve Punie, Y. (2017). *European framework for the digital competence of educators: DigCompEdu*. Publications Office of the European Union.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*. <https://www.jstatsoft.org/article/view/v017i05>

- Rudner, L. M. (1998). *An on-line, interactive computer adaptive testing mini tutorial*. <http://echo.edres.org:8080/scripts/cat/catdemo.htm> adresinden 10.03.2024 tarihinde alınmıştır.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*.
- Sangmeister, J. (2017). Commercial competence: Comparing test results of paper-and-pencil versus computer-based assessments. *Empirical Research in Vocational Education and Training*, 9, 1-19.
- Sarı, H. İ. (2016). *Examining content control in adaptive tests: Computerized adaptive testing vs. computerized multistage testing*. [Yayımlanmamış doktora tezi]. Florida Üniversitesi
- Sarı, H.İ. (2019). Investigating consequences of using item pre-knowledge in computerized multistage testing. *Gazi Üniversitesi Eğitim Fakültesi Dergisi*, 39(2), 1113-1134.
- Segall, D. O., (2004). Computerized Adaptive Testing. In Kempf-Leanard (Eds.). *The Encyclopedia of Social Measurement*, 429-438. San Diego, CA: Academic Press.
- Spenassato, D., Bornia, A. C., ve Tezza, R. (2015). Computerized adaptive testing: A review of research and technical characteristics. *IEEE Latin America Transactions*, 13(12), 3890-3898.
- Stone, E., ve Davey, T. (2011). Computer-adaptive testing for students with disabilities: A review of the literature. *ETS Research Report Series*, 2011(2), i-24.
- Sulak, S. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında kullanılan madde seçme yöntemlerinin karşılaştırılması*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Suna, E., Şensoy, S., ve Özer, M. (2021). Milli Eğitim Bakanlığı Tarafından Ölçme ve Değerlendirme Alanında Atılan Güncel Adımlar. *Eğitim ve İnsani Bilimler Dergisi: Teori ve Uygulama*, 12(23), 51-76
- Şahin, A., ve Özbaşı, D. (2017). Effects of content balancing and item selection method on ability estimation in computerized adaptive tests. *Eurasian Journal of Educational Research*, (69).
- Şahin, A., ve Weiss, D. J. (2015). Effects of calibration sample size and item bank size on ability estimation in computerized adaptive testing. *Educational Sciences: Theory ve Practice*, 15(6).
- Şahin, M. D. (2017). *Gerçek ve Üretilmiş Veri Setlerinde Çok Boyutlu Bireyselleştirilmiş Bilgisayarlı Test Uygulama Sonuçlarının İncelenmesi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Şahin, M. D., ve Gelbal, S. (2020). Development of a multidimensional computerized adaptive test based on the bifactor model. *International Journal of Assessment Tools in Education*, 7(3), 323-342.
- Şahin, M. G. (2020). Analyzing different module characteristics in computer adaptive multistage testing. *International Journal of Assessment Tools in Education*, 7(2), 191-206.
- Şahin, M. G., ve Öztürk, N. B. (2019). Analyzing the maximum likelihood score estimation method with fences in ca-MST. *International Journal of Assessment Tools in Education*, 6(4), 555-567.

- Şen, Ö. F. (2022). *Bilgisayarda Bireyselleştirilmiş Test Uygulamalarında Maddeyi Yeniden Cevaplayabilmenin Ölçme Hatasına ve Yanlılığına Etkisinin İncelenmesi*. [Yayımlanmamış Doktora Tezi]. Hacettepe Üniversitesi
- Şen, S., Cohen, A. S., ve Kim, S. H. (2013). A comparison of algorithms for dimensionality analysis. In *New developments in quantitative psychology: Presentations from the 77th annual Psychometric Society Meeting* (ss. 215-233). Springer New York.
- Şen, S., Cohen, A., ve Kim, S. H. (2020). A short note on obtaining item parameter estimates of IRT models with Bayesian estimation in Mplus. *Journal of Measurement and Evaluation in Education and Psychology*, 11(3), 266-282.
- Şenel, S. (2017). *Bilgisayar ortamında bireye uyarlanmış testlerin görme engelli öğrencilere uygunluğunun incelenmesi*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Şenel, S. (2021). *Bilgisayar Ortamında Bireye Uyarlanmış Testler*. Pegem Akademi Yayıncılık
- Şenel, S., ve Kutlu, Ö. (2018). Görme engelli öğrencilere yönelik bilgisayar ortamında bireye uyarlanmış test tasarımı. *Eğitim ve Bilim*, 43(194).
- Şimşek, A. S. (2017). *Becerilere güven mesleki ilgi envanterinin uyarlanması ve bilgisayarlı bireyselleştirilmiş test uygulamasının geliştirilmesi*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Tapan, H. R., ve Kılıç, T. (2020). Comparison of the graded response model and the generalized partial credit model in computerized adaptive testing. *Assessment in Education: Principles, Policy ve Practice*, 27(3), 273-290.
- Tat, O. (2020). *Açıklayıcı Madde Tepki Modellerinin Bilgisayar Ortamında Bireye Uyarlanmış Testlerde Kullanımı*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Taymur, M. Ö. (2023). *BOBUT uygulamalarında farklı koşullardan elde edilen ölçme kesinliği kestirim değerlerinin karşılaştırılması*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Tercan, S. S. (2010). *Çevrimiçi uyarlamalı bir test aracı geliştirilmesi, uygulanması ve öğrenci görüşlerinin belirlenmesi* [Yayımlanmamış yüksek lisans tezi]. Sakarya Üniversitesi
- Thissen, D. (1986). MULTILOG: Multiple, categorical item analysis and test scoring using item response theory. *Scientific Software, Inc.*
- Thissen, D., ve Mislevy, R. J. (2000). Testing algorithms. In *Computerized adaptive testing* (ss. 101-133). Routledge.
- Thompson, N. A. (2010). *Adaptive testing: Is it right for me?* University of Cincinnati, Assessment Systems Corporation. [http://www.assess.com/docs/Thompson_\(2010\)_-_Adaptive_Testing_Right.pdf](http://www.assess.com/docs/Thompson_(2010)_-_Adaptive_Testing_Right.pdf) adresinden 11.02.2024 tarihinde alınmıştır.
- Thompson, N. A., ve Weiss, D. J. (2009). *Computerized and adaptive testing in educational assessment*. The Transition to Computer-Based Assessment, 127.
- Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., ve Hagen, E. P. (1991). *Measurement and evaluation in psychology and education*. Macmillan Publishing Co, Inc.
- Thurlow, M., Lazarus, S. S., Albus, D., ve Hodgson, J. (2010). *Computer-based testing: Practices and considerations*. University of Minnesota, National Center on Educational Outcomes.

- Tseng, W. T. (2016). Measuring English vocabulary size via computerized adaptive testing. *Computers ve Education*, 97, 69–85. <https://doi.org/10.1016/j.compedu.2016.02.018>
- Tull, E., Dabner, N., ve Ayebi-Arthur, K. (2020). Social media and e-learning in response to the Covid-19 pandemic. *Journal of Open, Flexible and Distance Learning*, 24(1), 1-14.
- UCLA Center for the Study of Evaluation (CSE) (1980). SimulCAT Software. University of California, Los Angeles.
- Uysal, I., ve Doğan, N. (2021). Automated essay scoring effect on test equating errors in mixed-format test. *International Journal of Assessment Tools in Education*, 8(2), 222-238.
- van der Linden, W. J. (2006). Equating scores from adaptive to linear tests. *Applied Psychological Measurement*, 30(6), 493-508.
- van der Linden, W. J., ve Glas, C. A. W. (Eds.). (2010). *Elements of Adaptive Testing*. Springer.
- van der Linden, W. J., ve Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. Springer.
- van der Linden, W. J., ve Pashley, P. J. (2000). Item selection and ability estimation in adaptive testing. In W. J. van der Linden ve C. A. W. Glas (Eds.), *Computerized Adaptive Testing: Theory and Practice* (ss. 1-25). Springer.
- van Der Linden, W. J., ve Reese, L. M. (1998). A model for optimal constrained adaptive testing. *Applied Psychological Measurement*, 22(3), 259-270.
- Verschoor, A., Berger, S., Moser, U., ve Kleintjes, F. (2019). On-the-fly calibration in computerized adaptive testing. In *Theoretical and practical advances in computer-based educational measurement*, 307-323.
- Wainer, H. (2000a). *Computerized Adaptive Testing: A Primer*. Lawrence Erlbaum Associates.
- Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., ve Mislevy, R. J. (2000b). Computerized adaptive testing: A primer. *Routledge*.
- Wainer, H., ve Kiely, G. L. (1987). Multistage testing: A review. *Educational Measurement: Issues and Practice*.
- Walker, J., Böhnke, J. R., Cerny, T., ve Strasser, F. (2010). Development of symptom assessments utilising item response theory and computer-adaptive testing—A practical method based on a systematic review. *Critical reviews in oncology/hematology*, 73(1), 47-67.
- Wang, P., Dai, H., ve Ding, S. (2010). Computerized adaptive testing in Chinese mainland: A review. *International Journal of Digital Content Technology and its Applications*, 4(6), 94-99.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427-450.
- Weiss, D. J. (1983). Latent trait theory and adaptive testing. In David J. Weiss (ed.), *New horizons in testing* (ss. 5-7). New York: Academic Press.
- Weiss, D. J. (Ed.). (2014). *New horizon testing: Latent trait test theory and computerized adaptive testing*. Elsevier.
- Weiss, D. J., ve Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of educational measurement*, 21(4), 361-375.

- Wise, S. L., ve Kingsbury, G. G. (2016). Modeling student test-taking motivation in the context of an adaptive achievement test. *Journal of Educational Measurement*, 53(1), 86-105.
- Wright, B. D., ve Masters, G. N. (1982). *Rating Scale Analysis: Rasch Measurement*. Chicago: MESA Press.
- Wu, M. L., Adams, R. J., ve Wilson, M. R. (1997). ConQuest: Multi-aspect test software [computer program]. Camberwell, Vic.: Australian Council for Educational Research.
- Yaşar, M. (1999). *Bireyselleştirilmiş testler üzerine bir çalışma*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Yılmaz, K. (2015). Eğitimde ölçme ve değerlendirmenin önemi. *Eğitim Bilimleri Üç Aylık Dergisi*, 12(2), 25-35.
- Yigit, H. D., Sorrel, M. A., ve De La Torre, J. (2018). Computerized adaptive testing for cognitively based multiple-choice data. *Applied Psychological Measurement*, 43(5), 388-401.
- Yiğiter, M. S. (2023a). *Sabit ve anında bireyselleştirilmiş çok aşamalı testlerin karşılaştırılması*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Yiğiter, M. S., ve Doğan, N. (2023b). Comparison of Different Computerized Adaptive Testing Approaches with Shadow Test Under Different Test Length and Ability Estimation Method Conditions. *Journal of Measurement and Evaluation in Education and Psychology*, 14(4), 396-412.
- Yurtcu, M., ve Güzeller, C. O. (2021). Bibliometric analysis of articles on computerized adaptive testing. *Participatory Educational Research*, 8(4), 426-438.
- Zhang, Y. (2010). RSCAT: An R package for real-time computerized adaptive testing. *Journal of Statistical Software*.

EKLER

EK-1: Araştırmaya Dahil Edilen Çalışmalar

- Alkan, V. (2021). *Mesleki Alan İlgi Envanteri'nin Bilgisayar Ortamında Bireye Uyarlanmış Formunun Geliştirilmesi*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Aracı, F. G. İ., ve Tan, Ş. (2022). Multidimensional Computerized Adaptive Testing Simulations in R. *International Journal of Assessment Tools in Education*, 9(1), 118-137
- Arslan, Y. U. S. U. F., Çolak, M., ve Bilgin, U. (2022). Uzaktan Eğitimde Bilgisayar Uyarlamalı Test ve Kağıt-Kalem Testi ile Yetenek Düzeylerinin Belirlenmesi: Çok Merkezli Bir Çalışma. *Tıp Eğitimi Dünyası*, 21(63)
- Arslan, Y.K. (2021). *Bilgisayar uyarlamalı testlerde madde yanıt sürelerinin kullanımı*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Aşiret, S. (2019). *Bilişsel tanıya dayalı bilgisayar ortamında bireye uyarlanmış testlerde madde seçim yöntemlerinin incelenmesi*. [Yayımlanmamış doktora tezi]. Mersin Üniversitesi
- Aşiret, S., ve Sünbül, S. Ö. (2022). BT-BOBUT Uygulamalarında Madde Seçim Algoritmalarının Performanslarının Ölçme Doğruluğu Açısından İncelenmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, (54), 188-214
- Ayan, C. (2018). *Bilişsel tanı modelinde geleneksel ve bilgisayarlı sınıflamalı test uygulamalarının psikometrik özelliklerinin karşılaştırılması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Aybek, E. C. (2012). *Kâğıt-kalem formu ve bilgisayar ortamında uygulanan genel yetenek testinin psikometrik özelliklerinin karşılaştırılması*. [Yayımlanmamış Yüksek Lisans Tezi]. Ankara Üniversitesi
- Aybek, E. C. (2016). *Kendini değerlendirme envanterinin Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT) olarak uygulanabilirliğinin araştırılması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Ayhan, A. S. (2015). *Comparability of scores from cat and paper and pencil implementations of student selection examination to higher education* [Yayımlanmamış yüksek lisans tezi]. Bilkent Üniversitesi
- Balta, E., ve Arzu, U. Ç. A. R. (2022). Bilgisayar Ortamında Bireye Uyarlanmış Test Uygulamalarında Ölçme Kesinliğinin ve Test Uzunluğunun Farklı Koşullar Altında İncelenmesi/Investigation of Measurement Precision and Test Length in Computerized Adaptive Testing Under Different Conditions. *e-Uluslararası Eğitim Araştırmaları Dergisi*, 13(1), 51-68.
- Bostan, A. (2020). Using Fuzzy Logic in Computerized Adaptive Tests. *Bilişim Teknolojileri Dergisi*, 13(3), 289-300.
- Bulut, O., ve Kan, A. (2012). Application of computerized adaptive testing to entrance examination for graduate studies in Turkey. *Eurasian Journal of Educational Research*, 49, 61-80.
- Cıkrıkcı, N., Yalçın, S., Kalender, I., Gül, E., Ayan, C., Uyumaz, G., ... ve Kamış, O. (2020). Development of a Computerized Adaptive Version of the Turkish

- Driving Licence Exam. *International Journal of Assessment Tools in Education*, 7(4), 570-587.
- Çoban, E. (2020). *Bilgisayar temelli bireyselleştirilmiş test yaklaşımlarının Türkiye'deki merkezi dil sınavlarında uygulanabilirliğinin araştırılması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Demir, S. (2018). *Çok kategorili bireyselleştirilmiş bilgisayarlı test uygulamalarının farklı madde seçim yöntemlerinde sonlandırma kuralları açısından incelenmesi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Demir, S., ve Atar, B. (2021). Investigation of classification accuracy, test length and measurement precision at computerized adaptive classification tests. *Journal of Measurement and Evaluation in Education and Psychology*, 12(1), 15-27.
- Elhan, A. H., Öztuna, D., Kutlay, Ş., Küçükdeveci, A. A., ve Tennant, A. (2008). An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain. *BMC musculoskeletal disorders*, 9, 1-15.
- Erdem Kara, B. (2019). *Değişen Madde Fonksiyonu Gösteren Madde Oranının Bireyselleştirilmiş Bilgisayarlı ve Çok Aşamalı Testler Üzerindeki Etkisi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Eroglu, M. G. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının ölçme kesinliği ve test uzunluğu açısından karşılaştırılması*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Ertaş Polat, F. G. (2022). *Çok Aşamalı Bireye Uyarlanmış Testlerde Farklı Koşullardan Elde Edilen Yetenek Kestirimlerinin Karşılaştırılması*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi.
- Ertuna, L. (2020). *The Effect of Item Selection and Parameter Estimation Methods to the Accuracy of Pretest Item Parameters on Online Calibration in CAT*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Gelibolu, M. F. (2015). *Temel bilgisayar eğitimi alanında bilgisayarlı uyarlanabilir sınavların ölçme özelliklerinin diğer sınav türleriyle karşılaştırılması* [Yayımlanmamış doktora tezi]. Gazi Üniversitesi
- Gökcan, M. (2023). *İngilizce kelime bilgisi testinin (VST) bireyselleştirilmiş bilgisayarlı test olarak uygulanabilirliğinin incelenmesi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Gökçe, S. (2012). *Comparison of linear and adaptive versions of the Turkish pupil monitoring system (pms) mathematics assessment*. [Yayımlanmamış doktora tezi]. Orta Doğu Teknik Üniversitesi
- Gökçe, S., ve Glas, C. A. (2018). Can TIMSS mathematics assessments be implemented as a computerized adaptive test. *Journal of Measurement and Evaluation in Education and Psychology*, 9(4), 422-436.
- Gören, S., Kara, H., Kara, B. E., ve Kelecioğlu, H. (2022). The Effect of Aberrant Responses on Ability Estimation in Computer Adaptive Tests. *Journal of Measurement and Evaluation in Education and Psychology*, 13(3), 256-268.
- Gündeğer, C., ve Doğan, N. (2018). Bireyselleştirilmiş Bilgisayarlı Sınıflama Testi Kriterlerinin Test Etkililiği ve Ölçme Kesinliği Açısından

- Karşılaştırılması. *Journal of Measurement and Evaluation in Education and Psychology*, 9(2), 161-177.
- Gür, R., ve Gülleroğlu, H. D. (2020). Bireye uyarlanmış testlerde madde kullanım sıklığı kontrol yöntemlerinin farklı koşullarda ölçme duyarlılığına ve test güvenliğine etkisi. *Eğitim ve Bilim*, 45(202).
- Gür, R., ve Karabay, E. (2018). Motorlu Taşıtlar Sürücü Kursiyerleri Sınavının Simülatif Bilgisayar Ortamında Bireye Uyarlanmış Test Uygulaması. *Journal of Theoretical Educational Science*, 11(2), 201-228.
- Kaçar, M. (2016). *Bireye uyarlanmış bilgisayarlı test ortamında maksimum fisher madde seçme yönteminin incelenmesi* [Yayımlanmamış yüksek lisans tezi]. Necmettin Erbakan Üniversitesi
- Kalender, İ. (2011). *Effects of different computerized adaptive testing strategies on recovery of ability*. [Yayımlanmamış doktora tezi]. Ortadoğu Teknik Üniversitesi
- Kalender, İ. (2012). Yükseköğretime öğrenci seçmek için bilgisayar ortamında bireyselleştirilmiş testler. *Yükseköğretim Dergisi*, 2(1), 13-19.
- Kalender, I., ve Berberoglu, G. (2017). Can computerized adaptive testing work in students' admission to higher education programs in Turkey?. *Educational Sciences: Theory ve Practice*, 17(2).
- Kaplan, M. (2016). Nitelik Sayısının Madde Seçme Algoritmalarının Performansı Üzerindeki Etkisi. *Journal of Measurement and Evaluation in Education and Psychology*, 285-295.
- Kaplan, M., ve de la Torre, J. (2020). A Blocked-CAT Procedure for CD-CAT. *Applied Psychological Measurement*, 44(1), 49-64.
- Kaplan, M., de la Torre, J., ve Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied psychological measurement*, 39(3), 167-188.
- Kaptan, F. (1993). Yetenek kestiriminde adaptive (bireyselleştirilmiş) test uygulaması ile geleneksel kâğıt-kalem testi uygulamasının karşılaştırılması. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi.
- Kaskatı, O. T. (2011). *Rasch modelleri kullanarak romatoid artirit hastaları özürüllük değerlendirimi için bilgisayar uyarlamalı test yönteminin geliştirilmesi*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Kaya, E. (2022). *A Comparability and Classification Analysis of Computerized Adaptive And Conventional Paper-Based Versions of an English Language Proficiency Reading Subtest*. [Yayımlanmamış doktora tezi]. Bilkent Üniversitesi
- Kezer, F. (2013). *Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Kezer, F. (2021). The effect of item pools of different strengths on the test results of computerized-adaptive testing. *International Journal of Assessment Tools in Education*, 8(1), 145-155.
- Koşan-Aytuğ, M. (2013). *Tıp eğitiminde gelişim sınavı soru bankası oluşturulması ve benzetim verileri ile bilgisayar uyarlamalı test uygulaması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi

- Kürşad, M. Ş. (2020). *Madde parametre sapmasının bireye uyarlanmış testlerden elde edilen yetenek kestirimlerine etkisinin farklı koşullar altında incelenmesi*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Kürşad, M. Ş. (2023) The Effects of Different Item Selection Methods on Test Information and Test Efficiency in Computer Adaptive Testing. *Journal of Measurement and Evaluation in Education and Psychology*, 14(1), 33-46.
- Mao, X., Ozdemir, B., Wang, Y., ve Xin, T. (2018). Investigating The Effect of Exposure-Control Strategies on Item Selection Methods in MCAT. *Journal of Measurement and Evaluation in Education and Psychology*, 9(3), 295-311.
- Ozyurt, O. (2014). The Classification Of The Probability Unit Ability Levels Of The Eleventh Grade Turkish Students By Cluster Analysis. *Turkish Online Journal of Distance Education*, 15(2), 145-160.
- Özbaşı, D. (2014). *Bilgisayar okuryazarlığı testinin bilgisayar ortamında bireye uyarlanmış test olarak uygulanabilirliğine ilişkin bir araştırma*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Özdemir, B. (2015). *Madde düzeyinde boyutluluk modellerinin bilgisayar ortamında bireyselleştirilmiş test yöntemleri üzerindeki etkisinin incelenmesi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Özdemir, B. (2016). Comparison of different unidimensional-CAT algorithms measuring students' language abilities: Post-hoc simulation study. *European Proceedings of Social and Behavioural Sciences*.
- Özgür, E. G. (2020). *Rasch Modeline uygun yanıt desenlerinde bilgisayar uyarlamalı test yöntemi ile makine öğrenmesi yöntemlerinin tahmin performanslarının benzetim çalışması ile karşılaştırılması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Öztuna, D. (2008). *Kas-iskelet sistemi sorunlarının özürlülük değerlendirmesinde bilgisayar uyarlamalı test yönteminin uygulanması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Öztürk, N. B. (2014). *Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin İncelenmesi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Öztürk, N. B., ve Şahin, M. G. (2019). Effects of Item Pool Characteristics on Ability Estimate and Item Pool Utilization: A Simulation Study. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 34(2), 473-486.
- Özyurt, H., Özyurt, Ö., Baki, A., ve Güven, B. (2012). Integrating computerized adaptive testing into UZWEBMAT: Implementation of individualized assessment module in an e-learning system. *Expert Systems with Applications*, 39(10), 9837-9847.
- Raborn, A., ve Sarı, H. (2021). Mixed Adaptive multistage testing: A new approach. *Journal of measurement and evaluation in education and psychology*, 12(4), 358-373.
- Sarı, H. İ. (2016). *Examining content control in adaptive tests: Computerized adaptive testing vs. computerized multistage testing*. [Yayımlanmamış doktora tezi]. Florida Üniversitesi
- Sarı, H.İ. (2019). Investigating consequences of using item pre-knowledge in

- computerized multistage testing. *Gazi Üniversitesi Eğitim Fakültesi Dergisi*, 39(2), 1113-1134.
- Sulak, S. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında kullanılan madde seçme yöntemlerinin karşılaştırılması*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Şahin, A., ve Özbaşı, D. (2017). Effects of content balancing and item selection method on ability estimation in computerized adaptive tests. *Eurasian Journal of Educational Research*, (69).
- Şahin, A., ve Weiss, D. J. (2015). Effects of calibration sample size and item bank size on ability estimation in computerized adaptive testing. *Educational Sciences: Theory ve Practice*, 15(6).
- Şahin, M. D. (2017). *Gerçek ve Üretilmiş Veri Setlerinde Çok Boyutlu Bireyselleştirilmiş Bilgisayarlı Test Uygulama Sonuçlarının İncelenmesi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Şahin, M. G. (2020). Analyzing different module characteristics in computer adaptive multistage testing. *International Journal of Assessment Tools in Education*, 7(2), 191-206.
- Şen, Ö. F. (2022). *Bilgisayarda Bireyselleştirilmiş Test Uygulamalarında Maddeyi Yeniden Cevaplayabilmenin Ölçme Hatasına ve Yanlılığına Etkisinin İncelenmesi*. [Yayımlanmamış Doktora Tezi]. Hacettepe Üniversitesi
- Şenel, S. (2017). *Bilgisayar ortamında bireye uyarlanmış testlerin görme engelli öğrencilere uygunluğunun incelenmesi*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Şenel, S., ve Kutlu, Ö. (2018). Görme engelli öğrencilere yönelik bilgisayar ortamında bireye uyarlanmış test tasarımı. *Eğitim ve Bilim*, 43(194).
- Şimşek, A. S. (2017). *Becerilere güven mesleki ilgi envanterinin uyarlanması ve bilgisayarlı bireyselleştirilmiş test uygulamasının geliştirilmesi*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi
- Tat, O. (2020). *Açıklayıcı Madde Tepki Modellerinin Bilgisayar Ortamında Bireye Uyarlanmış Testlerde Kullanımı*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Taymur, M. Ö. (2023). *BOBUT uygulamalarında farklı koşullardan elde edilen ölçme kesinliği kestirim değerlerinin karşılaştırılması*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Yaşar, M. (1999). *Bireyselleştirilmiş testler üzerine bir çalışma*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Yigit, H. D., Sorrel, M. A., ve De La Torre, J. (2018). Computerized adaptive testing for cognitively based multiple-choice data. *Applied Psychological Measurement*, 43(5), 388-401.
- Yiğiter, M. S. (2023a). *Sabit ve anında bireyselleştirilmiş çok aşamalı testlerin karşılaştırılması*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi
- Yiğiter, M. S., ve Doğan, N. (2023b). Comparison of Different Computerized Adaptive Testing Approaches with Shadow Test Under Different Test Length and Ability Estimation Method Conditions. *Journal of Measurement and Evaluation in Education and Psychology*, 14(4), 396-412.