



HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

BOBUT UYGULAMALARINDA FARKLI KOŞULLARDAN ELDE EDİLEN ÖLÇME KESİNLİĞİ KESTİRİM DEĞERLERİNİN KARŞILAŞTIRILMASI

Melike ÖZER TAYMUR

Doktora Tezi

Ankara, 2023

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

Daha ileriye ... En İyiyeye ...



Eğitim Bilimler Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

BOBUT UYGULAMALARINDA FARKLI KOŞULLARDAN ELDE EDİLEN ÖLÇME KESİNLİĞİ
KESTİRİM DEĞERLERİNİN KARŞILAŞTIRILMASI
COMPARISON OF MEASUREMENT PRECISION VALUES OBTAINED FROM DIFFERENT
CONDITIONS IN COMPUTERIZED ADAPTIVE TESTING

Melike ÖZER TAYMUR

Doktora Tezi

Ankara, 2023

Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Melike ¼ZER TAYMUR'un hazırladıđı “BOBUT Uygulamalarında Farklı Koşullardan Elde Edilen ¼lçme Kesinliđi Kestirim Deđerlerinin Karşılaştırılması” bařlıklı bu çalıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitim ¼lçme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiřtir.

J¼ri Bařkanı	Prof. Dr. Zekeriya NARTG¼N	İmza
J¼ri Üyesi (Danıřman)	Prof. Dr. Duygu ANIL	İmza
J¼ri Üyesi	Prof. Dr. Burcu ATAR	İmza
J¼ri Üyesi	Prof. Dr. Nil¼fer KAHRAMAN	İmza
J¼ri Üyesi	Doç. Dr. Erg¼l DEMİR	İmza

Bu tez Hacettepe ¼niversitesi Lisans¼st¼ Eđitim, ¼đretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından 23 / 06 / 2023 tarihinde uygun g¼r¼lm¼ř ve Enstit¼ Y¼netim Kurulunca / / tarihi itibarıyla kabul edilmiřtir.

Prof. Dr. İsmail Hakkı MİRİCİ
Eđitim Bilimleri Enstitüsü M¼d¼r¼

Öz

Bu arařtırmada, bir bilgisayar ortamında bireye uyarlanmış test uygulaması (BOBUT)'nda iki farklı büyüklüğe sahip madde havuzundan elde edilen ölçme kesinlięi kestirim deęerlerinin; farklı madde seçme, farklı yetenek kestirim yöntemlerine göre içerik aęırlıklandırma ve madde kullanım sıklığı kontrolü gibi bazı madde havuzu sınırlandırmaları altında nasıl deęiřtięinin incelenmesi amaçlanmıřtır. Bu amaçla, simülatif olarak 2000 kiřilik bir gruba iliřkin yetenek parametreleri ile 320 ve 1000 maddelik iki farklı büyüklükte madde havuzu oluřturulmuřtur. Daha sonra içerik aęırlıklandırmanın yapılmadıęı ve yapıldıęı her iki durumda çalıřmanın kořulları olan; madde seçme yöntemi, yetenek kestirim yöntemi ve madde kullanım sıklığı kontrol yöntemi ile bir BOBUT uygulaması RStudio üzerinde “catR” paketi yardımıyla oluřturulmuř; BOBUT uygulamasına ait kořullar, bu iki ayrı madde havuzu ve aynı bireyler üzerinden simülasyonlar ile tek tek gerçekteřtirilmiřtir. Çalıřmada oluřturulan toplam 48 kořul, 156 replikasyon yapılarak karřılařtırılmıřtır. Simülasyonlar sonucunda, ölçme kesinlięi kestirim deęerleri (RMSE, yanlılık ve korelasyon) hesaplanmış ve bu deęerlerin 48 kořul altında nasıl deęiřtięi deęerlendirilmiřtir. Çalıřmadan elde edilen sonuçlar, madde havuzu büyüklüęü 1000 madde olan havuzun, madde havuzu büyüklüęü 320 madde olan havuza göre RMSE, yanlılık ve korelasyon deęerleri bakımından daha iyi sonuçlar verdięini göstermektedir. Ayrıca RMSE, yanlılık ve korelasyon deęerleri bakımından, çalıřmaya dahil edilen madde seçme yöntemleri arasında MFI ve yetenek kestirim yöntemleri arasında ise EAP yöntemlerinin daha etkili olduęu görülmüřtür. Bununla birlikte; içerik aęırlıklandırmanın yapıldıęı durumlarda, içerik aęırlıklandırmanın yapılmadıęı durumlara göre RMSE deęerlerinde bir miktar artış gözlenmiřtir. Madde kullanım sıklığı kontrolünde ise 320 maddelik madde havuzunda, MUİ yönteminin MFI ve MIS yöntemi ile birlikte kullanılmasının daha iyi performans gösterdięi belirlenmiřtir.

Anahtar sözcükler: bilgisayar ortamında bireye uyarlanmış testler, madde seçme yöntemleri, yetenek kestirim yöntemleri, içerik aęırlıklandırma

Abstract

In this study it is aimed to investigate the change in measurement precision estimation values that has been obtained from two different sizes of item pools in a CAT application under several item pool restrictions such as content balancing and item exposure control according to several item selection and ability estimation methods. For this aim, ability parameters related to a group of 2000 people and two different item pools one of which is 320 and the other is 1000 have been developed simulatively. Later, the conditions of the study in both balanced and unbalanced contents are constructed by developing a CAT application using catR package in RStudio with item selection, ability estimation and item exposure control methods and then these conditions of the CAT application are carried out by using simulations of two different item pools and the same participants. The 48 different conditions developed in the study are compared by making 156 replications. As a result of simulations, measurement precision estimation values (RMSE, correlations and bias) are computed and how these values are changed under 48 conditions are investigated. The findings obtained from the study have showed that the item pool with 1000 items has better results compared to the item pool with 320 items in terms of RMSE, bias and correlation values. Moreover, according to RMSE, bias and correlation values; MFI is seen to be the most effective among item selection methods and among ability estimation methods, EAP is found out to be the most efficient method. In addition; an increase in RMSE value has been observed in balanced content when it is compared with the value in unbalanced content. It is also found out that IEI method is more effective and has higher performance when it is used with MFI and MSI methods together.

Keywords: computerized adaptive testing, item selection methods, ability estimation methods, content balancing

Teşekkür

Doktora eğitim süresince ve tez çalışmamın her adımında neşesini ve gülen yüzünü eksiltmeden, bir anne şefkatiyle bana yol gösteren, bu süreçte her konuda destek olan, değerli görüş ve önerileriyle önemli katkılar sağlayan, emeğini ve sabrını esirgemeyen çok değerli hocam ve danışmanım Prof. Dr. Duygu ANIL'a;

Doktora eğitimim boyunca bilgi birikimleriyle beni yetiştiren ve bu yolda ilerlememi sağlayan kıymetli hocalarım Prof. Dr. Selahattin GELBAL ve Prof. Dr. Hülya KELECİOĞLU'na;

Tez izleme komitemde bulunarak değerli görüş ve önerileriyle tez çalışmamın ilerlemesine katkı sağlayan Prof. Dr. Burcu ATAR ve Doç. Dr. Ergül DEMİR'e;

Tez yazma sürecinde kod yazma ile ilgili karşılaştığım her sorunda ve güçlükte, verdiği fikir ve desteklerle yardımını esirgemeyen Doç. Dr. Cafer Mert YEŞİLKANAT'a;

Doktora yolculuğuna birlikte adım attığımız, tezimi yazma aşamasında yardım ve görüşlerini esirgemeyip ve beni son ana kadar destekleyen, bıkmadan usanmadan sorularımı cevapsız bırakmayan, varlıklarını her zaman yanımda hissettiğim ve her konuda görüşlerine başvurduğum canım arkadaşlarım Fatma Güneş ERTAŞ POLAT ve Begüm ÖZTEMUR'a;

Yıllardır iyi veya kötü her anımda yanımda olup, bu uzun yolculukta dertlerime ortak olan, tez çalışmam boyunca bana hep inanan ve tez sürecimdeki stresimi paylaşan, kan bağı olmaksızın kardeşim gibi sevdiğim Sevim Burçin GÜLTEKİN'e;

Tüm eğitim hayatım boyunca beni yalnız bırakmayan ve bugünlere gelmemde sayısız emekleri olan canım annem Selvi ÖZER ve vefatıyla beni derinden yaralayan canım babam Kemal ÖZER'e, her zaman desteğini hissettiğim abim Emrah Erkan ÖZER'e;

Tüm süreçleri benimle birlikte yaşayan, her zaman desteğini hissettiğim, yorulduğum anda bana güç veren sevgili eşim Erkan TAYMUR'a ve her güldüğünde tüm

olumsuzlukları unutmamı sağlayıp, bana bir ömür bahşeden canım oğlum Aras'a sonsuz teşekkürlerimi sunarım.

İçindekiler

Kabul ve Onay.....	ii
Öz.....	iii
Abstract.....	iv
Teşekkür.....	v
Tablolar Dizini.....	ix
Şekiller Dizini.....	x
Simgeler ve Kısaltmalar Dizini.....	xi
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	7
Araştırma Problemi.....	8
Sayıltılar.....	9
Sınırlılıklar.....	9
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	10
Madde Tepki Kuramı.....	10
Bilgisayar Ortamında Bireye Uyarlanmış Testler.....	15
Madde Havuzunda Bazı Sınırlandırmalar.....	32
İlgili Araştırmalar.....	41
Bölüm 3 Yöntem.....	49
Araştırmanın Türü.....	49
Araştırma Deseni.....	49
Verilerin Üretilmesi.....	51
Verilerin Analizi.....	57
Simülasyon Geçerliği.....	60
Bölüm 4 Bulgular, Yorumlar ve Tartışma.....	63
Birinci Alt Probleme İlişkin Bulgular ve Yorum.....	63

İkinci Alt Probleme İlişkin Bulgular ve Yorum	72
Tartışma.....	80
Bölüm 5 Sonuç ve Öneriler.....	84
Sonuçlar	84
Öneriler	87
Kaynaklar	91
EK-A: 320 ve 1000 Maddelik Havuzlara Ait Test Bilgi Fonksiyonları	cvi
EK-B: 320 ve 1000 Maddelik Havuzdan Rastgele Seçilen 10 Maddeye Ait a , b ve c Değerleri.....	cvii
EK-C: Rastgele Seçilen Bir Bireyin Cevaplaması Gereken 40 Maddenin İçerik Alanlarına Göre Dağılımı	cviii
EK-Ç: 320 Maddelik Havuza Ait a , b ve c Parametrelerinin Güven Aralığı Değerleri	cix
EK-D: 1000 Maddelik Havuza Ait a , b ve c Parametrelerinin Güven Aralığı Değerleri	cx
EK-E: Simülasyon Geçerliği İçin Kestirimi Yapılan Madde Parametrelerinin RMSE ve Yanlılık Değerleri (320 Madde Havuz Büyüklüğü)	cxı
EK-F: Simülasyon Geçerliği İçin Kestirimi Yapılan Madde Parametrelerinin RMSE ve Yanlılık Değerleri (1000 Madde Havuz Büyüklüğü)	cxii
EK-G: Simülasyon Geçerliği İçin Kestirimi Yapılan Yetenek Parametrelerinin RMSE ve Yanlılık Değerleri (320 Madde Havuzu Büyüklüğü)	cxiii
EK-Ğ: Simülasyon Geçerliği İçin Kestirimi Yapılan Yetenek Parametrelerinin RMSE ve Yanlılık Değerleri (1000 Madde Havuzu Büyüklüğü)	cxiv
EK-H: Araştırma Etik Komisyon İzin Muafiyeti Formu/ Araştırma Etik Komisyonu Onay Bildirimi	cxv
EK-I: Etik Beyanı	cxvi
EK-İ: Doktora Tez Çalışması Orijinallik Raporu	cxvii
EK-J: Dissertation Originality Report	cxviii
EK-K: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	cxix

Tablolar Dizini

Tablo 1 <i>BOBUT Simülasyon Koşulları</i>	50
Tablo 2 <i>320 Maddelik Havuza Ait Betimsel İstatistikler</i>	54
Tablo 3 <i>1000 Maddelik Havuza Ait Betimsel İstatistikler</i>	54
Tablo 4 <i>320 Maddelik Havuzda İçerik Alanları Boyunca Parametre Dağılımları</i>	57
Tablo 5 <i>1000 Maddelik Havuzda İçerik Alanları Boyunca Parametre Dağılımları</i>	57
Tablo 6 <i>Madde Havuzu Büyüklüğü 320 Madde Olan Havuzdan Elde Edilen Ölçme Kesinliği Kestirim Değerlerine Ait ANOVA Sonuçları</i>	64
Tablo 7 <i>Madde Havuzu Büyüklüğü 320 Madde Olan Havuzdan Elde Edilen Ölçme Kesinliği Kestirim Değerleri</i>	68
Tablo 8 <i>Madde Havuzu Büyüklüğü 1000 Madde Olan Havuzdan Elde Edilen Ölçme Kesinliği Kestirim Değerlerine Ait ANOVA Sonuçları</i>	74
Tablo 9 <i>Madde Havuzu Büyüklüğü 1000 Madde Olan Havuzdan Elde Edilen Ölçme Kesinliği Kestirim Değerleri</i>	76

Şekiller Dizini

Şekil 1 İki Kategorili Bir Maddeye Ait Madde Karakteristik Eğrisi	11
Şekil 2 1, 2 ve 3PLM ile Uyumlu Madde Karakteristik Eğrileri	13
Şekil 3 BOBUT Uygulama Adımları	17
Şekil 4 Yetenek Parametrelerinin Dağılımı	52
Şekil 5 320 Maddelik Havuza Ait Madde Parametre Dağılımları	55
Şekil 6 1000 Maddelik Havuza Ait Madde Parametre Dağılımları	56
Şekil 7 Madde Havuzu Büyüklüğü 320 Madde Olan Havuza Ait RMSE Değerleri	69
Şekil 8 Madde Havuzu Büyüklüğü 320 Madde Olan Havuza Ait Yanlılık Değerleri	71
Şekil 9 Madde Havuzu Büyüklüğü 320 Madde Olan Havuza Ait Korelasyon Değerleri	72
Şekil 10 Madde Havuzu Büyüklüğü 1000 Madde Olan Havuza Ait RMSE Değerleri	77
Şekil 11 Madde Havuzu Büyüklüğü 1000 Madde Olan Havuza Ait Yanlılık Değerleri	79
Şekil 12 Madde Havuzu Büyüklüğü 1000 Madde Olan Havuza Ait Korelasyon Değerleri	80

Simgeler ve Kısaltmalar Dizini

BOBUT: Bilgisayar Ortamında Bireye Uyarlanmış Testler

EAP: Beklenen Sonsal Dağılım

GEK: Güçlük Eşleştirme Kriteri

KTK: Klasik Test Kuramı

KY: Kısıtlayıcı Yöntem

MFI: Maksimum Fisher Bilgisi

MIS: Maksimum Bilgi Tabakalama

MLE: En Çok Olabilirlik Yöntemi

MUI: Madde Uygunluk İndeksi

MTK: Madde Tepki Kuramı

RMSE: Root Mean Squared Error (Hata kareleri ortalamasının karekökü)

1PLM: Bir Parametrelili Lojistik Model

2PLM: İki Parametrelili Lojistik Model

3PLM: Üç Parametrelili Lojistik Model

Bölüm 1

Giriş

Bu bölümde araştırmamızın problem durumu, amacı ve önemi, araştırma problemi, alt problemler, sayıtlılar ve sınırlılıklara yer verilmiştir.

Problem Durumu

Günümüz koşullarında teknolojinin gelişmesi sadece günlük yaşantımızı değiştirmekle kalmayıp, aynı zamanda yaşamımızda önemli bir yere sahip olan eğitimi ve dolayısıyla da eğitim sisteminde öğrencileri seçme ve yerleştirmede kullanılan geniş ölçekli sınavları etkilemektedir. Son yıllarda bilgisayarların ileri düzeylerde kullanımı, ölçme ve değerlendirmede kullanılan bu geniş ölçekli sınavların birçok açıdan değişmesine sebebiyet vermiştir.

Geleneksel anlamda kağıt-kalem testleri teknolojinin gelişmesiyle beraber yavaş yavaş etkinliğini yitirmektedir. Bunun nedeni kağıt-kalem testlerinin kolay hazırlanıp, kolay uygulanması gibi avantajlarının olmasının yanında bazı dezavantajlarının da olmasından kaynaklanmaktadır. Örneğin kağıt-kalem testlerinde, yetenek düzeyi üst sınırlarda yer alan bir öğrenci ile yetenek düzeyi alt sınırlara yakın bir değerde olan öğrenci aynı maddeleri yanıtlamaktadır. Bu durum, bireylerin yetenek düzeylerine uygun olmayan maddeleri çözmeye çalışarak testin uzunluğunun artmasına bu bakımdan da zaman kaybına yol açmasına neden olmaktadır. Diğer bir yandan yetenek düzeyi üst sınırlarda yer alan bir öğrenci, madde güçlüğü orta ve alt düzeyde olan maddeleri çok çabuk çözdüğü için testten sıkılmakta, yetenek düzeyi daha aşağılarda olan bir öğrenci ise madde güçlüğü yüksek olan maddeleri çözmek için çok vakit harcamakta ve genel olarak da sorunun çözümüne ulaşamayıp moral bozukluğu yaşamaktadır. Geleneksel kağıt-kalem testleriyle ilgili ifade edilen bu sınırlılıklardan dolayı bireye biçimlendirilmiş (tailored tests) veya bireye uyarlanmış (adaptive testing) testlere ihtiyaç duyulmuştur (Lord, 1970; Weiss, 1983; Hambleton and Swaminathan, 1985).

Bilgisayar ortamında bireye uyarlanmış test (BOBUT) uygulamaları, geleneksel kâğıt-kalem testlerinden daha etkilidir ve geleneksel kâğıt-kalem testlerine alternatif bir test uygulama biçimi olarak kabul edilmektedir. Tüm bireylerin aynı madde dizisini aldığı kâğıt-kalem testlerinin aksine BOBUT; uygulamaya katılan her birey için farklı bir test düzenler. Ayrıca her bireyin yetenek düzeyi ile eşleşen maddelerle karşılaşmasını sağlamaktadır. Bu durum bireylerin kafa karışıklığı, hayal kırıklığı veya can sıkıntısından kaynaklanan ölçme hatalarını azaltırken, ne çok kolay ne de çok zor olmayan, uygun güçlük düzeyi aralığında bir test almasını sağlayacaktır. Çünkü BOBUT uygulamaları sürecinde maddelerin güçlük derecesi bireylerin yetenek düzeyine uyarlanabilmektedir (Chang, 2004; Cheng vd., 2009; Segall, 2004; Georgiadou vd., 2007). Diğer bir yandan BOBUT uygulamaları, geleneksel kâğıt-kalem testlerine göre daha kısa testler, yetenek kestiriminde daha kesin ve etkili sonuçlar, anında puan dönütü verme, test planlamasında ve yönetiminde daha fazla esneklik ve yenilikçi madde formatlarının kullanımı gibi kolaylıklar sağlamaktadır (He, 2010). Özetlenecek olursa, BOBUT uygulamalarının en büyük avantajı, bireyin yeteneğinin ölçme kesinliğini artırırken test uzunluğunu kısaltabilmesidir (Evans, 2010; Jodoin, 2003; Weiss, 1982; Wainer, 2000).

BOBUT, seçilen bir sonraki soru veya maddenin, uygulamaya katılan bireyin önceki madde veya maddelere verdiği cevaba dayalı olarak belirlendiği bilgisayar tabanlı bir test biçimidir. Başka bir deyişle, birey kendine yöneltilen soruya doğru bir cevap vermişse daha zor bir soruyla, yanlış bir cevap vermişse daha kolay bir soruyla karşılaşacaktır. Sonuç olarak, uygulamaya katılan her birey, göstermiş oldukları performans modeline dayalı benzersiz bir test alacaktır. Bu bağlamda BOBUT uygulamaları; eğitsel değerlendirme, psikolojik testler, personel seçimi ve tıbbi teşhis dahil olmak üzere birçok alanda yaygın olarak kullanılmaktadır. BOBUT'un dayandığı temel mantık, her bir bireyin gizil yetenek düzeyine (uzamsal zeka, İngilizce yeterliliği, depresyon vb.) uygun hale getirilmiş bir madde havuzundan sırayla maddelerin seçilmesi sürecidir. Böylece, uygulamaya katılan bireyler kendileri için çok kolay veya çok zor olan birçok maddeyi cevaplamak zorunda

kalmayacaklardır. Sonuç olarak, BOBUT uygulamaları geleneksel kağıt-kalem testlerinden daha az maddeyle potansiyel olarak daha iyi yetenek düzeyi kestirimleri sağlayabilir (Cheng & Chang, 2009).

Bir BOBUT uygulaması, orta güçlük düzeyinde bireye sunulan ilk maddeye, bireyin verdiği cevaba göre ilk yetenek kestirimine dayalı olarak başlar. Her maddeye verilen cevaptan sonra, yeni bir yetenek kestirimi yapılır ve güncellenen yetenek düzeyine göre bir sonraki en uygun madde seçilir. Bu süreç, belirli durdurma kuralları karşılanana kadar, örneğin yeterliliğin kesinliği sağlanana veya sabit sayıda madde uygulanana kadar tekrarlanır. BOBUT uygulamaları, varsayımlarının karşılanması koşuluyla madde parametrelerini ve bireyin puanını aynı ölçüğe yerleştiren Madde Tepki Kuramı (MTK)'na dayanır. Bundan dolayı, çok farklı test örüntülerini cevaplayan bireyler karşılaştırılabilir bir puana sahip olabilir (Stone & Davey, 2011).

Her ne kadar hem teorik hem de deneysel çalışmalar, BOBUT uygulamalarının geleneksel kağıt-kalem testleri üzerindeki üstünlüğünü ve avantajlarını kanıtlamış olsa bile bazı sınırlılıkları da bulunmaktadır (Zheng vd., 2013). Rudner (1998) bu sınırlılıkları şu şekilde sıralamıştır:

1. BOBUT uygulamaları tüm konu ve beceriler için geçerli değildir. Bir BOBUT, genel olarak MTK modeline dayalıdır ancak MTK modelleri tüm becerilere ve madde türlerine uygulanamaz.
2. Donanım sınırlamaları, bilgisayar tarafından yönetilebilen madde türlerini kısıtlayabilir. Örneğin, ayrıntılı sanat eseri ve grafikler ya da kapsamlı okuma pasajları içeren maddeleri uygulamaya dahil etmek zor olabilir.
3. BOBUT uygulamaları, itinalı madde kalibrasyonu gerektirir. Bir kağıt ve kalem testinde kullanılan madde parametreleri, bir BOBUT bir uygulamasında geçerli olmayabilir.

4. BOBUT'lar, yalnızca bir kuruluştta çok sayıda sınava girecek kadar bilgisayar varsa ve sınava girenler en azından kısmen bilgisayar okuryazarı ise yönetilebilir. Bu büyük bir sınırlılık olabilir.
5. Test uygulama prosedürleri farklıdır. Bu, uygulamaya katılan bazı bireyler için sorunlara neden olabilir.
6. Uygulamaya katılan her bireyin farklı bir madde örüntüsü ile karşılaşması, algılanan eşitsizliklere neden olabilir.
7. Uygulamaya katılan bireylerin genellikle geri dönüp cevaplarını değiştirmelerine izin verilmez.
8. Cevapların değiştirilmesine izin verilirse, bazı bireyler kasıtlı olarak ilk soruları kaçırabilir. Algoritma daha sonra bu bireyi düşük yetenek düzeyinde varsayar ve bir dizi kolay soru seçer. Bu birey geri dönüp cevapları değiştirerek hepsini doğru yapabilir ve bu da uygulamaya katılan bireyin kestirilen yeteneğinin en yüksek yetenek düzeyi olmasına neden olur.

BOBUT uygulamalarının; güvenlik riski oluşturan maddelerin yalnızca küçük bir bölümünün kullanılma eğiliminde olması; içerik üzerindeki sınırlandırmaların yönetiminin zor olması gibi yetenek kestirimi üzerinde olumsuz bir etkiye sahip olan yaygın dezavantajları da vardır. Madde havuzunda pratik sınırlandırmalar olarak belirtilen bu duruma göre, bir testi oluşturan maddelerin içerik ve kullanım sıklığı bakımından madde havuzunda dengeli bir şekilde dağılım göstermesi önemlidir. Madde havuzunda sınırlandırmalar yapılan bir BOBUT uygulaması, içerik ağırlıklandırma ve madde kullanım sıklığı kontrolü sınırlandırmaları olan bir BOBUT uygulamasını ifade etmektedir (He, 2010). İçerik ağırlıklandırma, BOBUT uygulamasına katılan her bireyin benzer içerik bileşimine sahip bir test almasını sağlamaya yardımcı olur. Madde kullanım sıklığı kontrolü ise havuzda yer alan maddelerin aşırı kullanılmasını önlerken, nadiren kullanılan maddelerin de kullanım oranını artırarak test güvenliğini sağlamaya katkıda bulunur.

BOBUT uygulamalarında içerik ağırlıklandırma yapılması kağıt-kalem testlerine göre daha zor olabilir. Çünkü kağıt-kalem testlerinde testi geliştiren kişi, uygulamadan önce testin düzenlenmesini ve içerik özelliklerine göre test formlarının oluşturulmasını önceden belirleyebilir. Fakat kağıt-kalem testlerinin aksine, BOBUT uygulamalarında süreç, bir testi alan kişinin her bir maddeye verdiği cevaba göre şekillenmektedir. Dolayısıyla, testi geliştiren kişiler madde örüntüsü ve içerik sıralaması gibi konular üzerinde fazla kontrol sahibi olamamaktadır. Test geliştiricileri, içerik ağırlıklandırma gibi istatistiksel olmayan madde özelliklerinin dağılımını kontrol edebilmeleri için, bütün maddelerin özelliklerini ve algoritmaları tanımlamak hususunda, madde seçimi algoritmalarını da içeren çok sayıda içerik sınırlandırmalarına ihtiyaç duyabilir (Yan vd., 2014).

Uyarlanabilir bir testin doğal akışında, aynı test uygulamasına katılan bireyler farklı maddelerle karşılaşmaktadır. Fakat bu bireylerin, her içerik alanına göre aynı madde dağılımını alması gerekmektedir. Örneğin 28 maddelik bir matematik testi için, bir öğrenciye 28 maddenin sayı bilgisi alanından gelmesi, başka bir öğrenciye 28 maddenin geometriden gelmesi geçerli bir uygulama olmayacaktır. Çünkü ölçülen içerik alanları arasında bir ağırlıklandırma olmalıdır. Bu durum, uygulamadan elde edilen puanların geçerliğinin sorgulanmasına işaret etmektedir (Segall, 2004). BOBUT uygulamalarında, uygulamaya katılan her bireyin ağırlıklandırılmış bir madde dağılımı ile karşılaşmaması testin geçerliğini tehdit ederken, diğer taraftan ölçme kesinliğini, elde edilen puanların yorumlanmasını ve böylece de testin adilliğini etkilemektedir. Bu bakımdan içerik ağırlıklandırma, BOBUT uygulamaları için önem arz etmektedir. Bu nedenle araştırmacılar tarafından geliştirilen farklı içerik ağırlıklandırma yöntemleri, seçilen maddenin içerik alanını kontrol eden komutları madde seçme algoritmasına yerleştiren stratejilerdir (Segall, 2004; Zheng vd., 2013). İçerik ağırlıklandırma yöntemlerinin konu alanları arasındaki dengeyi sağlaması sayesinde, BOBUT uygulamaları süreci için geriye kalan tek şey, testin genel faktöre göre uyarlanmasıdır (Zheng vd., 2013). İçerik ağırlıklandırma yapılarak, içerik planının yerine getirilmesi, BOBUT uygulamalarının içerik geçerliliği için kanıt sağlar. Bir içerik planı,

içeriğin geçerliliğini sağlamaya yardımcı olacak kısıtlamaların belirlenmesine katkıda bulunur. Bir BOBUT uygulaması için içerik planları, genellikle her bir içerik alanında uygulanması amaçlanan alt ve üst sınırları (en az ve en çok madde sayısı) içerir. İçerik planının yerine getirilmesi hususunda ise içerik ağırlıklandırma kontrolünün dikkate alınması gerekir. İçerik ağırlıklandırma, bir test birden çok içerik alanından oluştuğunda ve her içerik alanının istendik oranları karşılması gerektiğinde kullanılır (Liu, 2019).

Madde havuzunda yapılması gereken pratik sınırlandırmalardan biri de madde kullanım sıklığı kontrolüdür. Örneğin araştırmalarda yaygın olarak kullanılan Maksimum Fisher Bilgisi (MFI) yöntemi, geçici yetenek düzeyi kestirimi için en çok bilgi veren maddeyi seçmektedir. Bununla birlikte bu yöntem, iyi maddelerin madde havuzunda olmasından gereken fazla kullanılması ve geriye kalan diğer maddelerin çok az kullanılması sebebiyle dengesiz ve orantısız bir madde havuzu kullanımına sebebiyet vermektedir (Cheng vd., 2009). Eğitimde kullanılan testler bağlamında maddenin deşifre olması ya da maddenin çok fazla kullanımı (yani teste katılan bireylerin maddeleri hatırlama ve diğer katılımcılarla paylaşma olasılığı) test güvenliğine yönelik bir tehdit olarak görülmekte ve literatürde bunun önemine dikkat çekilmektedir (Barrada vd., 2009a; Chang, 2004; Georgiadou vd., 2007; Hau & Chang, 2001). Test güvenliğini tehdit eden bu sınırlılığı ortadan kaldırmak için sunulan çözüm madde kullanım sıklığı kontrol yöntemleridir.

BOBUT uygulaması geliştiren kişiler, madde havuzunda pratik sınırlandırmaları göz önünde bulundurmalıdır. Çünkü öğrencinin geleceği ile ilgili belirleyici sonuçları olan sınavlarda içerik geçerliliği ve test güvenliği çok önemlidir ve bir test için yetenek düzeyi dağılımındaki sınırlandırma beklenen bir durumdur. Bununla birlikte, içerik geçerliliği ve test güvenliği, öğrencinin geleceği ile ilgili belirleyici sonuçları olan sınavlarda kritik kaygılardır (Lin, 2011). Bu bağlamda, madde havuzunda yapılan içerik ağırlıklandırma ve madde kullanım sıklığı kontrolü sınırlandırmaları, içerik geçerliliği ve test güvenliği hedeflerini sağlamaya yardımcı olan önemli faktörlerdir (Liu, 2019). Diğer bir yandan, içerik

ağırlıklandırma ve madde kullanım sıklığı kontrolü gibi istatistiksel olmayan sınırlandırmalar, savunulabilirliği ve geçerliliği test etmek için hayati hale gelmiştir (Hambleton, 2005).

Araştırmanın Amacı ve Önemi

Bu araştırma; BOBUT uygulamalarında ölçme kesinliği kestirim değerlerinin; farklı madde seçme, farklı yetenek kestirim yöntemlerine göre içerik ağırlıklandırma ve madde kullanım sıklığı kontrolü gibi bazı madde havuzu sınırlandırmaları altında nasıl değiştiğini incelemeyi amaçlamaktadır. Elde edilen sonuçlara göre en düşük hata ve yanlılık değeri ile en yüksek korelasyon değerine sahip koşulun hangi koşul olduğu konusunda fikir edinilebileceği düşünülmektedir.

Yurtdışı alanyazınında yer alan BOBUT uygulamaları ile ilgili araştırmalar, ülkemizde nadir olarak çalışılmakta olup halen yetersiz kalmaktadır. Yurtiçi alanyazın incelendiğinde ise, BOBUT uygulamasında madde havuzuna içerik ağırlıklandırma sınırlandırmasının dâhil edildiği sayıca çok az çalışmaya (Şahin, 2017) rastlanmaktadır. Alanyazın incelendiğinde, içerik ağırlıklandırmanın yapılmadığı ve yapıldığı her iki durumda; farklı madde havuzu büyüklükleri, farklı madde seçme ve farklı yetenek kestirim yöntemlerinin yanı sıra farklı madde kullanım sıklığı kontrol yöntemlerinin çaprazlanmasıyla oluşturulan simülasyon koşullarının ölçme kesinliği kestirim değerleri bakımından incelendiği bir çalışmaya rastlanmamıştır. Çalışmanın bu bakımdan yurtiçi ve yurtdışı alanyazına sağlayacağı katkı nedeniyle önemli olduğu düşünülmektedir. Araştırmada BOBUT uygulamalarında ölçme kesinliği kestirim değerlerinin; farklı madde seçme, farklı yetenek kestirim yöntemlerine göre içerik ağırlıklandırma ve madde kullanım sıklığı kontrolü gibi bazı madde havuzu sınırlandırmaları altında nasıl değiştiği incelenecektir. Ayrıca içerik ağırlıklandırmanın yapılmadığı ve yapıldığı durumların, BOBUT uygulamaları sürecinde diğer bileşenler ile birlikte uygulamaya dâhil edilmesinin, araştırma sonuçlarına zenginlik katacağı düşünülmektedir. Diğer bir yandan araştırmada BOBUT uygulamalarında yer alan diğer bileşenlere dâhil edilen içerik ağırlıklandırma yapılmayan ve yapılan durumların,

ülkemizde bu konuda yer alan az sayıda çalışmalara eklenecek olması araştırmanın önemli bir boyutunu oluşturmaktadır.

Zaman ilerledikçe, teknolojinin bizlere sağladığı kolaylıklar ve imkânlar bakımından hayatımızda büyük bir yere sahip olduğunu göz önüne alırsak; bu bağlamda BOBUT uygulamalarının avantajları düşünüldüğünde; ülkemizde de başarıyı ölçmek ve bireyleri yerleştirmek için kullanılan geleneksel kâğıt-kalem uygulamalarının yerine BOBUT uygulamalarının kullanılacağı günlerin pek yakında görülmesinin muhtemel olduğu söylenebilir. Bu durum, ülkemizde BOBUT uygulamaları ile ilgili farklı kategorilerde çalışmaların yapılmasının gerekliliğini ortaya koymaktadır.

Araştırma Problemi

BOBUT uygulamalarında ölçme kesinliği kestirim değerleri; farklı madde seçme, farklı yetenek kestirim yöntemlerine göre içerik ağırlıklandırma ve madde kullanım sıklığı kontrolü gibi bazı madde havuzu sınırlandırmaları altında nasıl değişmektedir?

Alt Problemler

1. BOBUT uygulamasında, madde havuzu büyüklüğü 320 madde olan havuzdan elde edilen ölçme kesinliği kestirim değerleri; içerik ağırlıklandırma yapılmadığı ve yapıldığı durumda, madde seçme yöntemine (MFI, MIS, GEK), yetenek kestirim yöntemine (EAP, MLE) ve madde kullanım sıklığı kontrol yöntemine (KY, MUI) göre nasıl değişmektedir?
2. BOBUT uygulamasında, madde havuzu büyüklüğü 1000 madde olan havuzdan elde edilen ölçme kesinliği kestirim değerleri; içerik ağırlıklandırma yapılmadığı ve yapıldığı durumda, madde seçme yöntemine (MFI, MIS, GEK), yetenek kestirim yöntemine (EAP, MLE) ve madde kullanım sıklığı kontrol yöntemine (KY, MUI) göre nasıl değişmektedir?

Sayıtlılar

Bu çalışmada kullanılan veriler simülatif olarak üretilmiş olup, iki kategorili 320 ve 1000 maddeden oluşan iki ayrı madde havuzu oluşturulmuş ve MTK modeli olarak 3PLM kullanılmıştır. Madde parametreleri; yetenek kestiriminde daha dengeli bir kestirim yeteneğine sahip bir madde havuzu elde etmek için a , b ve c parametreleri tek biçimli dağılımdan (Şahin, 2017) sırasıyla $U[0.8, 2.0]$, $U[-3, +3]$ ve $U[0.05, 0.2]$ olacak şekilde üretilmiştir. Bunun yanı sıra bireylere ait yetenek parametreleri $(-3, +3)$ aralığında ve normal dağılımdan ($N(0,1)$) toplam 2000 birey için rastgele üretilerek gerçek yetenek düzeyleri elde edilmiştir. Bu çalışma kapsamında belirlenen simülasyon koşullarının gerçek durum koşullarını sağladığı varsayılmıştır.

Sınırlılıklar

1. Araştırma kapsamında birden fazla içerik ağırlıklandırma yöntemi kullanılarak bu yöntemlerin performanslarının karşılaştırılması amaçlanmış ancak kullanılan catR paketinde, içerik ağırlıklandırma yöntemlerinden sadece bir tanesi olan sınırlandırılmış bilgisayar ortamında bireye uyarlanmış test uygulamaları yönteminin bulunması sebebiyle içerik ağırlıklandırma yöntemlerinin performansları karşılaştırılamamıştır. Çalışma bu yönüyle sınırlıdır.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde, çalışmanın kuramsal temeli olan madde tepki kuramı (MTK), bilgisayar ortamında bireye uyarlanmış testler (BOBUT) ve madde tepki kuramının bilgisayar ortamında bireye uyarlanmış testlerde kullanımına ilişkin genel bilgi verilmiştir. Son olarak alanyazındaki BOBUT uygulamalarının geliştirilmesi ve uygulanmasına ilişkin araştırmalara yer verilmiştir.

Madde Tepki Kuramı

20. yüzyılın büyük bir bölümünde, Klasik Test Kuramı (KTK) test geliştirme ve puanlamada dayanak noktası olmuştur. Bununla birlikte, Lord ve Novick (1968) tarafından ortaya atılan model tabanlı ölçme kuramı ile bu durum değişmiş ve test teorisinde sessiz bir devrim meydana gelmiştir. Madde tepki kuramı adı verilen bu kuram, psikolojik ölçmeler için bir temel olarak hızla ana akım olarak anılmaya başlanmıştır (Embretson & Reise, 2000).

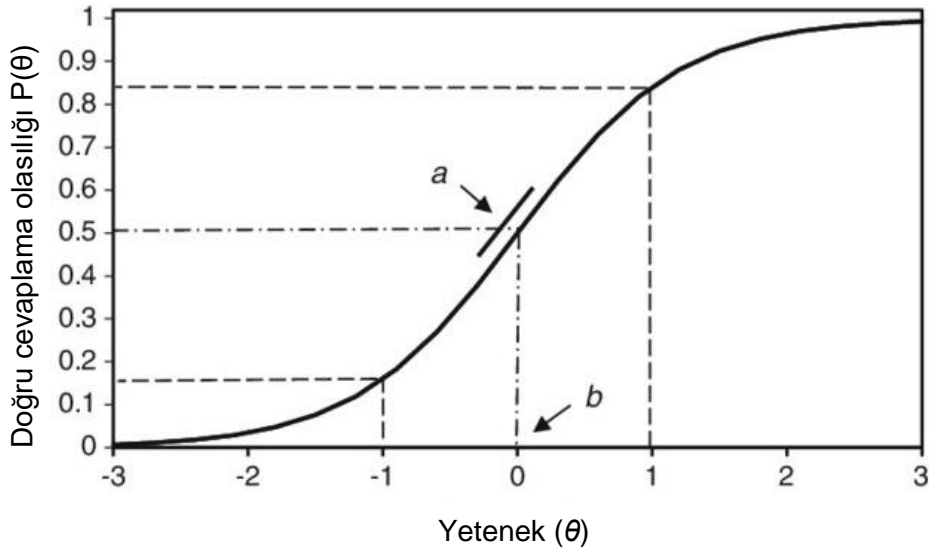
MTK'nın KTK'ya göre bir takım üstün yönleri olduğu düşünülmektedir. Örneğin, KTK'da ölçmenin standart hatası belirlenen katılımcı grubundaki tüm test puanları için aynıyken MTK'da test puanları (ya da cevap dizileri) boyunca farklılık gösterir ve katılımcılar boyunca genelleştirilebilir. Ayrıca, KTK'da kullanılan ölçek metriği temelindeki özellikle ilgili bir aralık değildir oysa MTK'daki ölçek metriği temelde olan özellikle ilgili bir aralıktır. MTK'nın diğer bir avantajı ise uzun testlerden ziyade daha kısa testler kullanıldığında bile güvenilir ölçme sonuçlarının elde edilebilmesidir. Diğer bir yandan, MTK modeli hem ikili hem de çok kategorili maddeleri içeren daha karmaşık karma modelleri ele alabilir (Embretson & Reise, 2000).

MTK'da ölçme konusu olan zeka, yetenek, beceri gibi özellikler gizil veya örtük (latent) değişken olarak adlandırılır ve θ (teta) olarak simgelenir (Şenel, 2021). MTK, bir anlamda, gizil değişken olarak nitelendirilen bireyin yeteneği ile bu bireyin bir maddeye verdiği cevap arasındaki uyumu kurmanın yolunu tanımlayan bir modeller sistemidir (De

Ayala, 2009; DeMars, 2010; Nguyen vd., 2014). Kurmaya çalıştığı bu uyumu, olasılığa dayalı olarak yapmaktadır. Diğer bir ifadeyle maddeye doğru yanıt verme olasılığını, bireyin yeteneği (θ) ve madde parametreleri arasında kurulan matematiksel fonksiyonlarla açıklayan bir kuramdır. Bir bireyin bir maddeyi doğru yanıtlama olasılığı, yetenek ve madde parametreleri arasındaki ilişkiyi tanımlayan fonksiyona “madde karakteristik fonksiyonu”; fonksiyonun oluşturduğu eğriye de “madde karakteristik eğrisi” denir. Madde karakteristik fonksiyonunda yetenek düzeyi (θ) yükseldikçe, $P(\theta)$ ile sembolize edilen maddeye doğru yanıt verme olasılığı da artmaktadır (Hambleton vd., 1991; Lord, 1980). İki kategorili bir madde için bir madde karakteristik eğrisi Şekil 1’de gösterilmektedir (Nguyen vd., 2014).

Şekil 1

İki Kategorili Bir Maddeye Ait Madde Karakteristik Eğrisi



MTK modeli, bir testin verileri ile uyum içerisinde olduğu zaman bazı avantajlı durumlar elde edilebilir. Bunlar; sınava giren bireylerin yetenek kestirimlerinin teste ve madde indekslerinin gruba bağlı olmadığı durumlardır. Yapılan ölçmelerde farklı madde kümelerinden elde edilen yetenek kestirimleri (ölçme hataları hariç) ile farklı gruplardan elde edilen madde parametre kestirimleri aynı olacaktır (ölçme hataları hariç). Bu MTK’da madde ve yetenek parametrelerinin değişmezliği olarak tanımlanır. Madde ve yetenek parametrelerinin değişmezlik özelliği, maddelerle ilgili bilgilerin yetenek kestirim sürecine

dahil edilmesiyle ve sınava giren bireylerin yetenek düzeyleri ile ilgili bilgilerin madde parametre kestirim sürecine dahil edilmesiyle elde edilir (Hambleton vd., 1991).

MTK bazı genel varsayımları veya koşulları gerektirmektedir. Bunlar; tek boyutluluk ve yerel bağımsızlık varsayımlarıdır (DeMars, 2010; Hambleton & Swaminathan, 1985; Hambleton vd., 1991; Mahmud, 2017). Tek boyutluluk, bir testte yer alan maddelerle yalnızca bir yeteneğin ölçülmesi anlamına gelir (Crocker & Algina, 1986; Hambleton vd., 1991; Meijer & Tendeiro, 2018). Bu varsayım her zaman tam olarak karşılanamaz çünkü test performansını etkileyen motivasyon düzeyi, sınav kaygısı, hızlı okuma becerisi, şans başarısı ve bilişsel beceriler gibi bazı faktörler vardır. Bir dizi test maddesi tarafından ölçülen baskın bir bileşen veya faktörün varlığı tek boyutluluk varsayımının karşılanması için yeterlidir. Bu baskın bileşen veya faktör, test tarafından ölçülen yetenek olarak adlandırılır (Hambleton vd., 1991). Yerel bağımsızlık varsayımı ise, herhangi bir davranışın ölçülmesinde sınava giren bireyin bir testteki maddelere verdiği cevapların istatistiksel olarak bağımsız olması demektir (De Ayala, 2009; DeMars, 2010; Hambleton vd., 1991; Meijer & Tendeiro, 2018). Tek boyutluluk varsayımı sağlandığında, yerel bağımsızlık varsayımı da sağlanmış olur (Lord, 1980). Tek boyutluluk varsayımının karşılanmadığı durumlar için çok boyutlu MTK modellerinin kullanımı önerilmiştir. Bu çalışmada tek boyutlu MTK modelleri kullanıldığından sadece bu modellere değinilecektir.

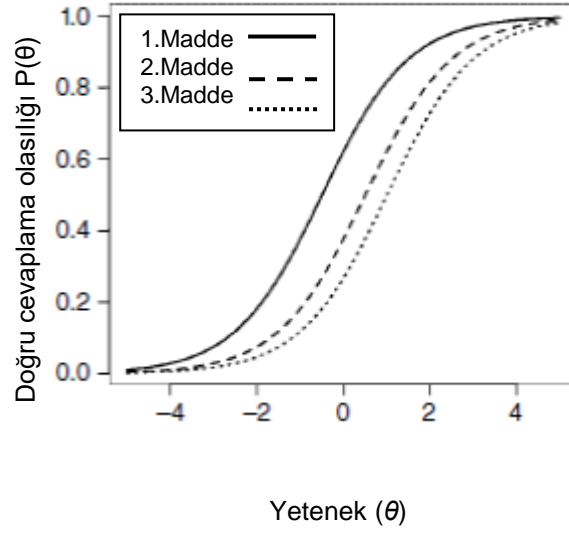
MTK'da bir test maddelerine verilen cevap kategorilerine göre modeller iki kategorili ve çok kategorili olarak ikiye ayrılmıştır. Bu çalışmada iki kategorili maddeler için kullanılan modeller ele alındığı için yalnızca bu modellere ait detaylı bilgi verilecektir. İki kategorili modeller kullanılan parametre sayısına göre farklılık gösterir (Meijer & Tendeiro, 2018; Zanon vd., 2016). Bir parametrelili lojistik model (1PLM) veya Rasch (1960) modelinde yalnızca bir madde güçlük parametresi vardır ve b_i olarak gösterilir. İki parametrelili Birnbaum modelinde (2PLM) güçlük parametresine madde ayırt edicilik parametresi eklenir ve a_i ile gösterilir. Son olarak üç parametrelili lojistik modelde (3PLM) diğer iki parametreye şans

başarısı parametresi eklenir ve c_i ile gösterilir. Şekil 2'de sırasıyla 1, 2 ve 3PLM ile uyumlu madde karakteristik eğrileri verilmiştir (Meijer & Tendeiro, 2018).

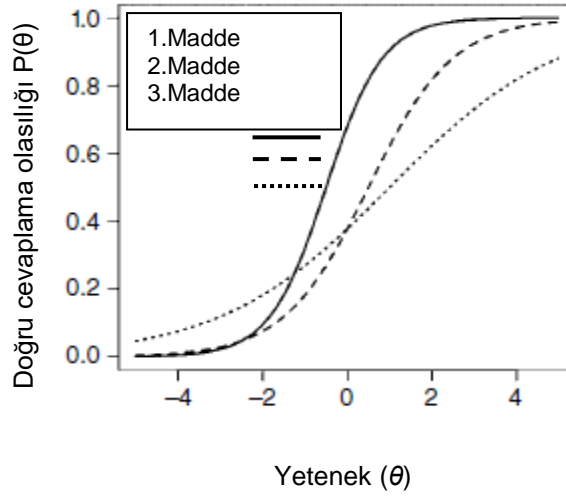
Şekil 2

1, 2 ve 3PLM ile Uyumlu Madde Karakteristik Eğrileri

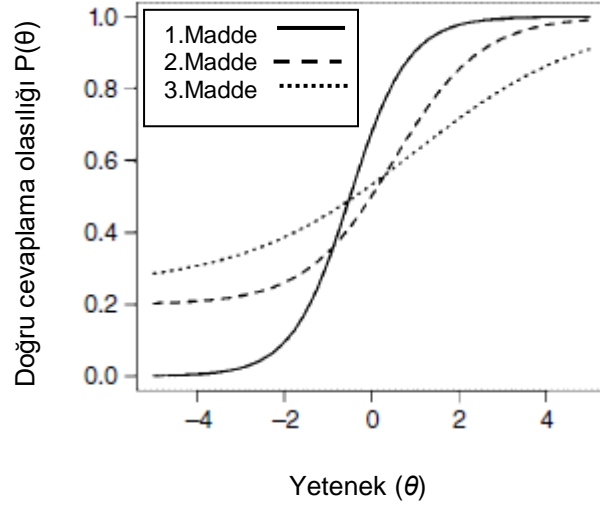
Şekil 2a



Şekil 2b



Şekil 2c



Üç parametrelili lojistik (3PL) model (Birnbbaum, 1968) için eşitlik aşağıdaki gibidir:

$$P(u_{ij} = 1 | \theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}}$$

Burada $P(u_{ij} = 1 | \theta_j, a_i, b_i, c_i)$ j kişinin i maddesine doğru cevap verme olasılığıdır; u_{ij} , j kişinin i maddesine verdiği cevaptır (1 doğru ve 0 yanlış); θ_j , j kişinin yeteneğidir. b_i , i maddesinin madde güçlük parametresini; a_i , i maddesi için ayırt edicilik parametresini ve c_i , i maddesi için şans başarısı parametresini göstermektedir.

Şans başarısı parametresi tüm maddeler için 0 olarak ayarlanırsa, 3PLM aşağıda eşitliği verilen 2PLM haline gelir (Birnbbaum, 1968):

$$P(u_{ij} = 1 | \theta_j, a_i, b_i) = \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}}$$

Son olarak madde ayırt edicilik parametresi tüm maddelerde 1 olacak şekilde daha da kısıtlanırsa, 2PLM şu şekilde tanımlanan bir Rasch modeliyle sonuçlanır (Lord & Novick, 1968):

$$P(u_{ij} = 1 | \theta_j, b_i) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}}$$

Madde konum parametresi olarak da adlandırılan b_i , θ ölçeğinde bir maddeye doğru cevap verme olasılığının $(1 + c_i) / 2$ 'ye eşit olduğu nokta olarak tanımlanır. $c_i = 0$ olduğunda (1PLM ve 2PLM'de) maddenin konumu, bu maddeyi cevaplama olasılığının 0,5'e eşit olduğu θ ölçeğindeki nokta olarak tanımlanır. Bu nedenle, madde karakteristik eğrisini ölçeğin sağ tarafına taşıdığımızda, herhangi bir maksimum performans testinin uygulanması durumunda daha zor bir madde olacaktır; madde karakteristik eğrisini ölçeğin sol tarafına aldığımızda daha kolay bir madde olacaktır. Bu nedenle b_i parametresi daha önce de bahsedildiği gibi güçlük parametresi olarak da bilinir. Madde konum parametreleri genellikle -2,5 ile +2,5 arasındadır (Meijer & Tendeiro, 2018).

Madde karakteristik eğrisinin dikliği, ayırt edicilik parametresi a_i 'de ifade edilir. Bu parametre, $\theta = b_i$ noktasında eğriye teğet bir fonksiyondur. Ayırt edicilik parametre değerinin sıfıra yakın olması durumunda, θ ölçeğinde birbirine yakın bireyler arasında bilenle bilmeyeni ayıramayan maddelerin varlığından bahsedilebilir (yani ilgili madde karakteristik eğrileri düzdür). Son olarak 3PLM'in şans başarısı parametresi c_i , madde karakteristik eğrisinin alt asimptotunu belirtir. Bahsedilen MTK modellerinden Rasch modeli, yalnızca bir parametreye sahip olduğu için en kısıtlayıcı modeldir oysa 3PLM daha esnek bir modeldir (Meijer & Tendeiro, 2018). Bu sebeple bu araştırmada 3PLM kullanılmıştır.

Bilgisayar Ortamında Bireye Uyarlanmış Testler

Bir aracı bireye uyarlama kavramı, bir asırdan fazla bir süre boyunca geçmişten günümüze uzanan bir kavramdır. Bu süre zarfında adaptif testler, bilgisayar ortamında bireye uyarlanmış testler isminin yanı sıra, uyarlanmış testler, yanıt-uyumlu testler, sıralı testler ve programlanmış testler gibi birçok farklı isme sahip olmuştur (De Ayala, 2009).

Bireye uyarlanmış ilk testin, 1900'lü yılların başında geliştirilen Binet-Simon zeka testi olduğu düşünülmektedir (Weiss, 1982). Bu testte uygulanan belirli alt testler, testin uygulanması esnasında belirlenen bireyin mevcut yetenek düzeyine göre seçilmiştir. Yani bir birey alt testlerin tümünü veya herhangi birini, bir yetenek düzeyi kapsamında geçerse,

daha sonra daha yüksek yetenek düzeyine sahip bir alt test uygulanır. Tam aksine, eğer bir birey belirli bir yetenek düzeyinde tüm alt testlerden başarısız olursa, test sona erdirilir. Bu nedenle, Binet testi yetenek düzeyine göre uyarlanabilmektedir (De Ayala, 2009).

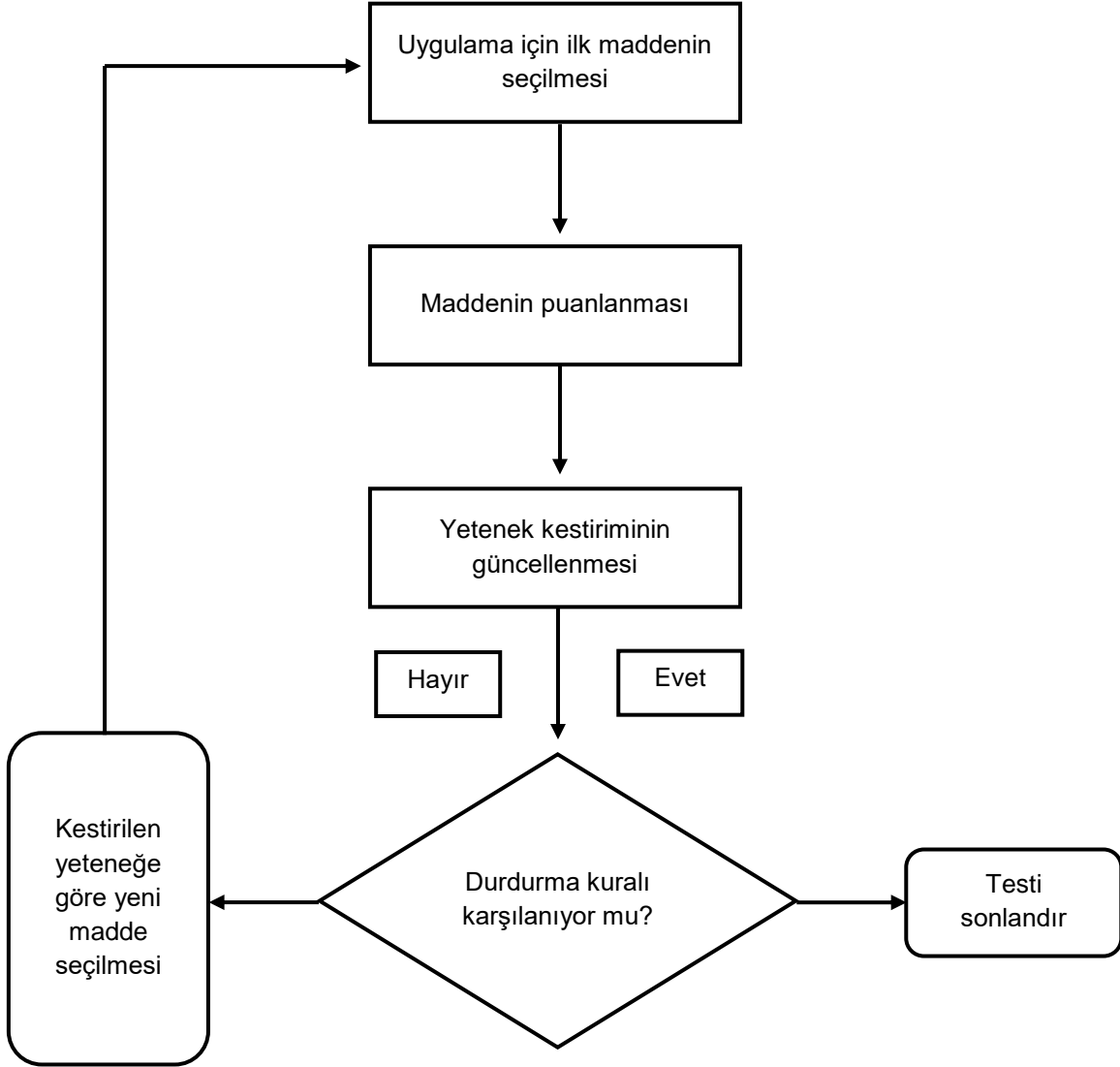
Bilgisayar ortamında bireye uyarlanmış test (BOBUT) uygulamaları eğitime ait test programlarında çok yaygın bir şekilde kullanılan, sistemle etkileşim sırasında oluşturulan ve güncellenen bir öğrenci profiline dayalı olarak uygulanacak bir sonraki maddenin seçiminin ve testi durdurma kararının dinamik olarak gerçekleştirildiği, bilgisayar tarafından uygulanan bir testtir (Conejo vd, 2004). BOBUT'un uygulanmasındaki temel mantık, maddelerin her bireyin yetenek düzeyine göre sıralı olarak uygulanması yöntemidir. Bu yöntemde, her madde bireyin testte vermiş olduğu cevaplara dayanan güncel yetenek düzeyine göre belirlenen madde seçme kuralı tarafından seçilmektedir. Daha sonra her maddenin güçlüğü, bireyin yetenek düzeyi ile içerik ağırlıklandırma ve madde kullanım sıklığı oranı gibi diğer pratik gereksinimlerle uyumlu hale getirilir. Bu madde seçme süreci durdurma kuralı karşılanana dek devam eder (Wang, 2017). He (2010)'a göre, BOBUT uygulamalarına ait sürecin şematik gösterimi Şekil 3'de sunulmuştur:

BOBUT uygulamalarında testin yapılandırılması ve uygulanması bilgisayarlaştırılmakta ve kişiselleştirilmektedir (Eggen, 2004). Başka bir ifadeyle BOBUT uygulamaları, teste katılan bir katılımcının daha önceki maddelere verdiği cevaplara dayanarak bir sonraki maddenin seçildiği uygulama sürecidir. Bu süreçte, katılımcıya bir madde yöneltilir, bu madde cevaplanır, cevap değerlendirilir ve katılımcının yetenek düzeyi kestirilir. Bu yetenek düzeyine göre de yeni bir madde seçilir (Orcutt, 2002; Thissen & Mislevy, 2000; Wainer, 2000; Weiss, 1983; Weiss & Kingsbury, 1984; Yao, 2019). Bu süreç neticesinde maksimum performansın ölçüldüğü test uygulamalarında yetenek düzeyi yüksek bireyler daha güç maddelere maruz kalırken, yetenek düzeyi düşük bireyler daha kolay maddelere maruz kalmakta ve her birey için özelleştirilmiş bir test formu ortaya çıkmaktadır (Bartram & Hambleton, 2006; De Ayala, 2009; DeMars, 2010; Eggen, 2004). BOBUT uygulamalarının ardında yatan temel mantık, testi alan bireye çok zor ya da çok

kolay gelebilecek maddelerden kaçınma ve bireyin özelliğine en uygun maddelerin seçilmesidir (Weiss & Kingsbury, 1984; Davey, 2010).

Şekil 3

BOBUT Uygulama Adımları (He, 2010)



Diğer bir ifadeyle BOBUT, uygulamaya katılan bir bireyin ilk yetenek düzeyinin kestirimi ile başlayan ve aşağıdaki adımları içeren yinelemeli bir algoritmadır (Conejo vd, 2004):

1. Madde havuzundaki (henüz uygulanmamış) tüm maddeler, uygulamaya katılan bireyin yetenek düzeyine ilişkin mevcut kestirimine göre bir sonraki sorulacak en iyi maddenin hangisi olacağını belirlemek için incelenir.

2. Madde bireye yöneltilir ve uygulamaya katılan birey bu maddeyi cevaplar.
3. Verilen cevaba göre yeni bir yetenek kestirimi yapılır.
4. 1'den 3'e kadar olan adımlar, belirlenen durdurma kuralı sağlanana kadar tekrarlanır.

BOBUT uygulamaları için ana sebep etkili ölçmedir. Bu bakımdan daha az madde ile daha duyarlı ve kesin ölçmeler yapılmaktadır (Eggen,2004). BOBUT uygulamalarında ölçme süreci ise aşağıdaki bileşenlerden oluşmaktadır (Weiss & Kingsbury, 1984; Kingsbury & Zara, 1989):

- a) Madde havuzu,
- b) Madde tepki modeli,
- c) Teste başlama kuralı,
- d) Madde seçme yöntemi,
- e) Yetenek kestirim yöntemi ve
- f) Sonlandırma kuralı.

Bu bileşenlerin her biri, belirli bir ölçme ihtiyacını karşılamak için bir test geliştirmeye çalışan kişi tarafından ayrı ayrı incelenmelidir (Kingsbury & Zara, 1989).

Madde Havuzu

Madde havuzu, bireylerin yeteneklerini iyi tanımlayacak şekilde oluşturulmuş maddelerin bir bütünüdür (Eggen, 2004). BOBUT uygulamalarında madde seçimi, amaca en uygun şekilde ulaşmak için oluşturulan madde havuzundan yapılır. Oluşturulan bu madde havuzu, maddelerin yanı sıra bu maddelere ait çeşitli özellikleri de içermektedir. Bu özellikler, içerik kategorilendirilmesi ya da maddeler üzerindeki uygulama verileri ile ilgili olmaktadır. Fakat bu özelliklerden en önemlisi maddelerin psikometrik özellikleridir (Eggen, 2004).

BOBUT uygulamalarının ve dolayısıyla da yaptığı kestirimlerin niteliği, diğer tüm test türlerinde olduğu gibi büyük ölçüde kullanılan madde havuzunun kalitesine bağlıdır (Flaughner, 2000; Jensema, 1977; Paap vd., 2017; van der Linden vd., 2006; Wise vd., 2015). Madde havuzu BOBUT uygulamalarının büyük öneme sahip olan bir bileşenidir (Gu & Reckase, 2007). İdeal bir madde havuzu ise, bir dizi katılımcının yetenek düzeyleri boyunca çoklu test formları oluşturacak kadar maddeye sahip olmasıyla ilişkilidir (Davey & Nering, 2002; Wise vd., 2015). Wise (1997)'a göre madde havuzu kalitesinin iki temel kritere göre kavramsallaştırılabileceği önerilmektedir. Bunlardan ilki madde havuzunun bir test oturumu boyunca bilgi verecek yeterli sayıda maddeye sahip olması iken, ikinci kriter ise madde havuzundaki maddelerin araştırmacıya katılımcıların yetenek düzeyleri konusunda yeterli bilgi sağlayacak özellikte olmasıdır.

Madde havuzunun nitelikli olması, uyarlanabilir algoritmanın da işlevini iyi yapması demektir. Eğer bir madde havuzu sınırlı sayıda ya da düşük kaliteli maddelerden oluşuyorsa, en iyi ve en gelişmiş uyarlanabilir program bile işlevini iyi bir şekilde yerine getiremez (Flaughner, 2000). Reckase (2010) mümkün olan en iyi madde havuzunu, madde seçme algoritması tarafından seçilen maddenin uygulanabildiği havuz olarak tanımlamıştır.

BOBUT uygulamalarının birçok avantajını gerçekleştirebilmek için, maddelerin seçildiği madde havuzu, birçok farklı yetenek düzeyine uygun nitelikli maddeler içermelidir. Bunun aksine, geleneksel kağıt-kalem testleri ise orta düzeydeki katılımcılar açısından en iyi ölçme yapılabilirsin diye maddeler seçilerek oluşturulur. Ayrıca, madde havuzu nitelikli ve farklı yetenek düzeylerine uygun maddeler gerektirmesinin yanı sıra madde kalibrasyonu, uygulama ve puanlamanın temelini oluşturan psikometrik modelin varsayımlarını da karşılamalıdır. Bu nedenle, madde havuzu oluşturma sürecinde madde yazımı geleneksel kağıt-kalem testlerine göre daha zor ve zaman alıcıdır (Flaughner, 2000).

Flaughner (2000)'e göre madde yazım sürecinde izlenmesi gereken adımlar şu şekildedir:

1. Daha önce belirlenen test özelliklerine göre her içerik kategorisinde yeterli sayıda madde oluşturulur.
2. Maddelerin niteliğine bakılır; ölçme değerlendirme ve alan uzmanı incelemesi yapılır.
3. Yeni yazılan maddelerin pilot uygulaması yapılır. Bilgisayar formatına dönüştürüldüğünde ortaya çıkabilecek olası sorunlara rağmen, ilk madde havuzlarının kağıt-kalem testi formatında oluşturulması gerekebilir.
4. Pilot uygulamadan elde edilen madde parametrelerine dayanarak, orijinal maddelerden hem geleneksel madde analizi istatistiklerine hem de madde tepki kuramına göre bir alt test seçilir.
5. Nihai madde havuzunun farklı içerik alanlarına göre bölümlendirilmesi önceki test formlarınıninkiyile karşılaştırılır ve farklı yetenek düzeylerine sahip bireylerin davranışlarının bir simülasyonu yapılarak sistemin işleyişi değerlendirilir.
6. Geriye kalan maddeler bilgisayar formatına dönüştürülür.

BOBUT uygulamaları için kabul edilebilir bir madde havuzu, yüksek ayrımcılığa sahip (a'nın 1'den büyük), dikdörtgen bir zorluk dağılımı ve düşük tahminli (c'nin, 2'den küçük) parametrelerine sahip maddelerle oluşturulan bir havuzdur (Flaughner, 2000).

Madde havuzu oluşturulurken sıkça sorulan sorulardan biri de “Havuzda kaç tane madde olmalı?” sorusudur. Bu sorunun cevabı olacak en iyi cümle “Ne kadar çok madde o kadar iyi madde havuzu” olacaktır. Bu sebeple madde havuzunun yeterli sayıda nitelikli maddeye sahip olması gerekmektedir (Gu & Reckase, 2007; He & Reckase, 2013; Millman & Arter, 1984). Çünkü madde havuzunun büyük olması BOBUT uygulamalarında daha çok seçenek olmasını sağlayarak maddenin ifşa edilebilirlik oranını düşürmektedir (Gu & Reckase, 2007). Bu da katılımcıların maddeleri hatırlama probleminin önüne geçmektedir. Ayrıca madde havuzunun büyük olması havuzda bulunan maddelerin içerik, madde formatı

ve istatistiksel gereksinimlerle büyük ölçüde eşleşmesi anlamına gelmektedir (Millman & Arter, 1984).

Madde Tepki Modeli

BOBUT uygulamaları, MTK'ya dayalı olarak yapılandırılmıştır (Folk & Smith, 2002; Thompson & Weiss, 2011). MTK, test geliştirme, madde analizi ve bireylerin yeteneğini kestirmede birçok üstün yönü olan güçlü bir psikometrik paradigmadır. BOBUT bakımından düşünüldüğü zaman ise en önemli avantajının maddelerin ve uygulamaya katılan bireylerin aynı ölçeğe yerleştirilerek, bireylerin en uygun maddelerle doğrudan eşleşmesinin sağlanması olduğu söylenebilir. BOBUT uygulamaları hala KTK ile tasarlanabilirken diğer bir yandan çalışmaların büyük çoğunluğunun MTK'ya dayalı olması bu avantajın önemini vurgulamaktadır. Bu nedenle, BOBUT'un işleyişini anlamak için MTK'ya aşina olmak gerekmektedir (Thompson & Weiss, 2011).

BOBUT ve MTK'nın bir arada kullanılması psikometristlere bazı avantajlar sağlamaktadır. Bunlar (Bjorner vd., 2005):

1. Her birey için en uygun maddelerin seçilmesiyle, verilen bir test uzunluğu için değerlendirme kesinliği uygun hale getirilir ve gereksiz madde kullanımından kaçınılabilir.
2. Değerlendirme kesinliği, uygulamanın ihtiyaçlarına göre ayarlanabilir. Örneğin tanılayıcı amaçlı yapılacak değerlendirmelerde, değerlendirme kesinliği kesme noktalarına yakın puanlar için yüksek olmalıdır veya test kesinliği bireylerin takibi amacıyla tüm puan aralığında yüksek olarak ayarlanabilir.
3. Farklı maddeler kullanılmış veya farklı kesinlik seviyeleri belirtilmiş olsa bile değerlendirmeler karşılaştırılabilir.

4. Madde havuzları, geriye dönük karşılaştırılabilirlikten ödün vermeden yeni maddelerin eklenmesi ve değerlendirilmesiyle kademeli olarak genişletilebilir.
5. Madde havuzuna geleneksel soru formlarından maddeler eklenerek, yaygın olarak kullanılan formları çapraz kalibre etmek mümkündür.
6. Değerlendirme kalitesinin ve anormal yanıt modellerinin belirlenmesini sağlamak için cevaplama süreci eş zamanlı olarak izlenebilir.
7. Değerlendirmenin sonunda, uygulamaya katılan bireye puanın nasıl yorumlanacağına ilişkin yönergelerle birlikte hemen bir puan verilebilir.

Bu avantajlardan bazıları diğer metodolojilerle elde edilebilse de, BOBUT uygulamalarının kullanımı ve MTK modellemesi için gerekli olan maddelerin dikkatli bir şekilde incelenip geliştirilmesi, değerlendirme kalitesini önemli ölçüde artırmaktadır (Bjorner vd., 2005).

3PLM, büyük ölçekli test ve BOBUT uygulamalarında en yaygın olarak uygulanan MTK modelidir (Green, vd., 1984; Wainer & Mislevy, 2000). Çünkü diğer iki MTK modelinde (1PLM, 2PLM), şans başarısı parametresinin olmaması çoktan seçmeli madde kullanımının önüne geçmektedir (Green, vd., 1984). Aslında BOBUT'larda şans başarısının etkisi ve bir şans parametresine duyulan ihtiyaç çok da göz önünde bulundurulmuş bir durum değildir. Çünkü bir BOBUT uygulaması düzgün çalışıyorsa ve yeterince işlevselse uygulamaya katılan bireyin uygun olmayan zor bir maddeyle karşılaşması çok düşük bir olasılıktır. Bununla birlikte, şans başarısı parametresi bazı önemli avantajlar da sunar. Örneğin BOBUT'larda, bireylerin genellikle bir maddeye cevap vermeden geçmelerine izin verilmez ve bu nedenle cevabı bilmediklerinde bir bakıma cevabı tahmin etmeye mecbur bırakılırlar. Bu sebeple şans başarısı parametresinin bulunduğu 3PLM'nin BOBUT için önemi büyüktür (Wainer & Mislevy, 2000).

Başlama Kuralı

BOBUT uygulamalarında, teste başlamak için birden fazla yol kullanılabilir. BOBUT uygulamalarında uygulanacak olan ilk madde, farklı bireyler için farklı zorluk seviyelerindeki maddeler olabilir. Örneğin bir katılımcının başarı düzeyinin yüksek olduğu düşünülüyorsa, uygulama nispeten zor bir madde ile başlayabilir. Seçilen maddelerin zorluk seviyesi, test ilerledikçe katılımcının yetenek düzeyine göre belirleneceğinden, hatalı bir ilk madde seçimi sonuçları ciddi şekilde etkilemeyecek ancak ilk madde seçiminin doğru düzeyde yapılması yetenek kestirimi için gereken madde sayısını azaltacaktır (Weiss & Kingsbury, 1984).

Teste başlamak için ilk madde seçiminde bireylere ait yetenek düzeyi bilinmediğinden uygulamaya hangi güçlük düzeyinde bir madde ile başlanması gerektiğini belirlemek zor bir süreçtir (De Ayala, 2009). Bunun için teste başlama kuralı olarak kullanılan birkaç farklı yöntem vardır. Bunlardan ilki, bireyin yeteneği hakkında ön bilginin olmadığı durumda başlama kuralı olarak uygulamanın orta güçlük düzeyinde bir madde ($\theta=0$) ile başlatılmasıdır (Magis vd., 2017; Mills & Stocking, 1996; Thompson & Weiss, 2011). Fakat bu durum aynı yetenek ranjında bulunan bireylerin aynı maddelerle karşılaşması olasılığını arttıracaktır. Bu da test güvenliğini tehdit edecek ve bazı maddelerin aşırı kullanımına sebebiyet verecektir. İlk madde seçiminde güçlük düzeyinin -0.5 ile +0.5 aralığında rastgele olarak belirlenmesi oluşan bu durumu önlemeye yardımcı olan başka bir yöntemdir (Thompson & Weiss, 2011).

Bahsi geçen başlama kurallarının her ikisinde de bireyler hakkında hiçbir ön bilginin olmadığı varsayılmaktadır. Bununla birlikte, bireylerin yetenek düzeyine ait ön bilgilerin mevcut olduğu birçok durum da vardır (Thompson & Weiss, 2011). Eğer bireylerin yetenek düzeyine ait bir ön bilgi varsa, başlama kuralı olarak ilk madde seçiminde bu ön bilgi kullanılabilir (Embretson & Reise, 2000; Thissen & Mislevy, 2000; Thompson & Weiss, 2011). Böyle bir durumda, bireylerin yetenek düzeyine ait bu ön bilgi sonraki uygulamalar

için ideal bir başlangıç noktasıdır. Çünkü bireylere ait yetenek düzeyinin benzer bir aralıkta olması muhtemeldir.

İlerleme Kuralı

BOBUT uygulamalarında, bireyin ilk maddeye verdiği cevaba göre bilgisayar bir sonraki maddenin seçimine karar vermektedir. Ancak burada düşünülmesi ve üzerinde durulması gereken iki soru vardır:

1. Bireylerin maddelere verdiği cevaplar nasıl puanlanacaktır?
2. Bireylerin verdiği cevaba göre bir sonraki madde nasıl seçilecektir?

Bu iki sorunun cevaplanması için düşünülmesi gereken puanlama (yetenek kestirim) ve madde seçme yöntemi bileşenleridir. Bunlardan ilk olarak bireylerin vermiş olduğu cevapların puanlanmasında kullanılan yetenek kestirim yöntemlerine değinilecektir.

Yetenek Kestirim Yöntemleri. Bir BOBUT uygulamasında, uygulanan her maddeden sonra bireyin önceki tüm maddelere verdiği cevaplara dayalı olarak yetenek kestirimi güncellenir. Yetenek düzeyinin belirlenmesinde, uygun bir yetenek kestirim yönteminin seçilmesi çok önemli bir noktadır. Yetenek kestirimi, yalnızca testin nihai sonucunu değil, aynı zamanda hangi maddelerin uygulanacağını da etkilemektedir (Wang & Vispoel, 1998). Bu bağlamda, yaygın olarak kullanılan iki yetenek kestirim yöntemi en çok olabilirlik ve Bayes yöntemleridir (Bejar & Weiss, 1979; Chen, 1999; Wang, 2017; Wang & Vispoel, 1998). Bu yöntemlere ait detaylı bilgi aşağıda verilmiştir.

En Çok Olabilirlik Kestirim Yöntemi (MLE). Bu yaklaşımın amacı, gözlemlenen madde cevapları örüntüsü için maksimum olabilirlikle sonuçlanan bir kestirim yapmaktır. Bir cevap örüntüsü ve bilinen parametrelere sahip bir dizi madde verildiğinde, olabilirlik işlevi şu şekildedir (Song, 2010):

$$L(U_j|\theta_j) = \prod_{i=1}^n P(u_{ij}|\theta_j)$$

Burada $L(U_j|\theta_j)$, θ_j 'de bulunan bir j kişisi için U_j cevap dizisinin olasılığıdır;

u_{ij} , j kişinin i maddesine verdiği madde cevabıdır/ (doğru cevap için 1 ve yanlış cevap için 0);

$P(u_{ij}|\theta_j)$, θ_j konumunda bulunan bir j kişinin i maddesi için u cevabını alma olasılığıdır.

Eşitliğe göre bir bireyin gerçek θ yeteneğinin maksimum olabilirlik tahminini veren $\hat{\theta}$, bu olabilirlik fonksiyonunu (veya log olabilirlik fonksiyonunu) maksimize eden değerdir. Bu değer, olabilirlik fonksiyonunun türevinin sıfıra eşitlenmesi ve elde edilen denklemin çözülmesi ile elde edilebilir. Newton-Raphson yöntemi gibi yinelemeli sayısal yöntemler tipik olarak bu denklemi çözmek için kullanılır (Wang & Vispoel, 1998). Fakat uygulamaya katılan bireyler tüm maddeleri doğru veya yanlış cevapladığında olabilirlik denkleminin çözümünde sorunlar ortaya çıkar ve en çok olabilirlik yöntemi bu tür cevap örüntülerinde bireylerin yetenek kestirim değerlerini sağlayamaz (Chen, 1999; Mao, 2014; Song, 2010; Thompson & Weiss, 2011; Wang, 2017; Wang & Vispoel, 1998). En çok olabilirlik yönteminin bu sınırlılığına çözüm olarak, $\hat{\theta}$ 'nin makul bir aralıkta (örneğin, -4 ila 4) sınırlanması veya Bayesian yöntemleri gibi alternatif yöntemlerin kullanılması gösterilebilir.

Bayesian Yetenek Kestirim Yöntemleri. Yetenek kestirime yönelik Bayesian tabanlı yaklaşımlar, yetenek dağılımı varsayımlarının yetenek kestirime dahil edilmesi bakımından en çok olabilirlik yönteminden farklıdır. Kavramsal olarak, uygulamaya katılan bireylerin yetenek dağılımının biçimi, ortalaması ve varyansı hakkındaki varsayımlarla maddelerin doğru ve yanlış cevaplanma olabilirlik bilgilerini birleştiren yetenek kestirimlerini ortaya koyar. Başlangıçta, bireylerin yetenek dağılımlarının ortalaması sıfır ve standart sapması bir olan normal bir dağılımdan geldiği varsayılır. Bu başlangıçta elde edilen yetenek dağılımına önsel dağılım denir. Uygulamaya katılan birey ilk maddeyi cevapladıktan sonra, verdiği cevaba ilişkin olabilirlik, önsel yetenek dağılımı hakkındaki bilgilerle birleştirilerek sonsal dağılım oluşturulur. Daha sonra bu sonsal dağılım, uygulamaya katılan bireyin ikinci maddeye verdiği cevaba ilişkin olabilirlikle birleştirilecek olan önsel dağılım haline gelir. Bu süreç, önsel dağılım olarak kullanılan bir önceki adımdan

gelen sonsal dağılım, uygulamaya katılan bireyin olabilirliği ile birleştirilmek üzere sıralı bir şekilde madde madde devam eder (Wang & Vispoel, 1998). Bu durumu gösteren eşitlik aşağıda verilmiştir (Song, 2010):

$$h(\theta|U_j) = \frac{L(U_j|\theta)f(\theta)}{\int_{\theta} L(U_j|\theta)f(\theta)d\theta}$$

Burada $f(\theta)$, θ için önsel olasılık yoğunluk fonksiyonudur,

U_j , j kişisi için madde cevap dizisidir,

$L(U_j|\theta)$, θ düzeyinde verilen madde cevap dizisinin olasılığıdır (olabilirlik fonksiyonu),

ve $h(\theta|U_j)$, madde yanıt dizisi verilen θ 'nin sonsal olasılık yoğunluğudur.

Bayesian yöntemleri, önsel ve sonsal dağılımların biçimi hakkında yapılan varsayımlar ve bireyin yetenek kestiriminde kullanılan sonsal dağılımların özellikleri açısından farklılık göstermektedir. Bu bağlamda yaygın olarak kullanılan üç farklı Bayesian yöntemi vardır. Bunlar; Beklenen Sonsal Dağılım (EAP), Maksimum Sonsal Dağılım (MAP) ve Owen'in yöntemleridir. Bayesian yöntemi olarak sadece EAP yetenek kestirim yöntemi kullanıldığından diğer iki yöntemde değinilmemiştir.

EAP yöntemi, yetenek kestirimi olarak sonsal dağılımın ortalamasını, $h(\theta|U_j)$ 'yi kullanır (Song, 2010):

$$\hat{\theta} = E(\theta|U_j) = \int_{-\infty}^{\infty} \theta h(\theta|U_j) d\theta$$

Burada integralin doğrudan hesaplanması yerine, Gauss-Hermite kareleme noktaları kullanılarak yaklaşık bir hesaplama yapılır,

$$\hat{\theta} = E(\theta|U_j) = \frac{\sum_{k=1}^q X_k L_i(U_j|X_k) W(X_k)}{\sum_{k=1}^q L_i(U_j|X_k) W(X_k)}$$

X_k , q kareleme noktalarından biridir, $W(X_k)$ bu noktayla ilişkili bir ağırlıktır ve $L_i(U_j|X_k)$ i maddeden sonra X_k 'da değerlendirilen olasılık fonksiyonudur.

Bayesian yetenek kestirim yöntemleri arasında, EAP yöntemi en istikrarlı ve daha küçük ortalama karesel hataya sahip kestirimleri sağlarken bu kestirimler ortalamadan uzak yetenek düzeylerinde yanlıdır (Bock & Mislevy, 1982). MLE yöntemi ile karşılaştırıldığında ise EAP yöntemi bazı avantajlara sahiptir. EAP, bireye ait yeteneği kestirememeye riski olmadan daha basit bir prosedür kullanır. Ayrıca EAP yöntemi yinelemeli hesaplama gerektirmez ve birey tüm maddeleri doğru ya da yanlış cevapladığında bile yetenek kestirimi yapabilir. EAP yöntemi MLE yönteminde ortaya çıkan problemi çözebilse de bir BOBUT uygulamasının ilk aşamalarında önsel dağılımdan etkilenmesi gibi bir dezavantajı vardır. Önsel dağılım uygun değilse, bireyin yetenek düzeyine ilişkin EAP yönteminin yaptığı kestirim yanlı olacaktır (Wang & Vispoel, 1998). Bu araştırmada, bireylerin tüm soruları doğru cevapladığı ya da hiç doğru cevap veremediği tüm cevap örüntülerinde kullanılabilirliği ve tüm yetenek düzeylerinde hata ortalamalarının karekök değerinin MLE'ninkinden daha küçük olması sebebiyle (Bock & Mislevy, 1982) EAP yöntemi ve MLE yöntemine göre sağladığı üstünlükleri karşılaştırabilmek adına MLE yetenek kestirim yöntemi kullanılmıştır.

Madde Seçme Yöntemleri. BOBUT uygulamalarında test süreci ilerlerken bireyin cevaplayacağı bir sonraki madde, belirlenen madde seçme yöntemine göre algoritma tarafından seçilmekte ve bireye uygulanmaktadır. BOBUT uygulamalarının temel bir bileşeni olan madde seçme kriteri, uygulamaya katılan bireylere en uygun test sorularının veya maddelerinin uygulanmasını sağlayarak test ölçüm kalitesini ve verimliliğini artırabilir (Choi & Swartz, 2009).

Bir BOBUT uygulaması gerçekleştirilirken bireyin cevaplayacağı bir sonraki maddeyi seçme aşamasında en yaygın olarak kullanılan iki yöntem; bireyin verdiği yanıtlara dayalı olarak kestirilen geçici θ değeri için “maksimum bilgi” ve “maksimum sonsal kesinlik” sağlayan yöntemleridir (Thissen & Mislevy, 2000; van der Linden, 1998; Wang, 2017).

Maksimum bilgi madde seçme yönteminde, mevcut θ kestiriminde maksimum bilgiye sahip olan madde seçilir (Eggen, 2007; Lin, 2012; Weiss & Kingsbury, 1984). Bu, θ belirlendikten sonra, madde havuzunda henüz uygulanmayan her madde için bilgi değerinin bu kestirim değerinde hesaplandığı ve bir sonraki adımda maksimum bilgi değerine sahip maddenin uygulandığı anlamına gelmektedir (Eggen, 2007). Maksimum sonsal kesinlik diğer adıyla Owen'in Bayes yaklaşımı yönteminde ise, yetenek kestiriminin beklenen sonsal kesinliğini maksimize eden bir madde seçilir (Wang, 2017). Bu yöntemde, her birey uygulamaya başlangıç yetenek düzeyine ilişkin bir önsel bilgi ile başlar. Bireyin cevapladığı her bir maddeden sonra, verdiği cevap ve önsel dağılımlar kullanılarak yetenek kestirimi yeniden yapılır ve bu yetenek kestirimine ait sonsal dağılım oluşturulur. Bu aşamadan sonra, beklenen sonsal dağılımın varyansını azaltan maddeler seçilir ve madde seçimi beklenen varyans düzeyine ulaşana kadar süreç devam eder (Weiss & Kingsbury, 1984). Bu çalışmada MFI, MIS ve GEK yöntemleri kullanıldığından, bu yöntemlere ait ayrıntılı bilgi aşağıda verilmiştir.

Maksimum Fisher Bilgisi (MFI). BOBUT uygulamalarında en yaygın olarak kullanılan ve muhtemelen en eski madde seçme yöntemlerinden biri olan MFI yöntemi, uygulamaya katılan bireye önceden uygulanan test maddelerine dayanan geçici yetenek kestiriminde maksimum Fisher bilgisine sahip bir maddenin seçilmesini içermektedir. Temel olarak bu, geçici yetenek kestirimi $\hat{\theta}$ 'da uygulanan $m - 1$ madde için $I[\hat{\theta}_{m-1}]$ 'i maksimum yapan x maddesini bulmayı içermektedir. Madde parametrelerinin üç parametrelilik lojistik model tarafından tanımlandığı çoktan seçmeli bir madde havuzunun tipik bir örneğini ele alarak MFI yöntemine dayalı madde seçimi aşağıdaki eşitlikte verilmiştir:

$$I_i[\theta_{m-1}] = \frac{(Da_i)^2(1 - c_i)}{\left[c_i + e^{Da_i(\theta_{m-1} - b_i)} \right] \left[1 + e^{-Da_i(\theta_{m-1} - b_i)} \right]^2}$$

Verilen eşitlikte sırasıyla a_i, b_i ve c_i ; ayırdedicilik, güçlük ve şans başarısı parametrelerini; D ise değeri 1,702 olan bir ölçeklendirme sabitini göstermektedir (Han, 2012).

MFI yöntemi arařtırmalarda çoęunlukla tercih edilen bir yaklařımdır. Çünkü her birey için maksimize edilmiş test bilgileriyle sonuçlanan BOBUT uygulamasının etkili bir yoludur. Ancak MFI yöntemi BOBUT'un başlangıç aşamasında, örneğin bireylere beş veya daha az madde uygulandığında, geçici yetenek kestirimlerini nadiren doğru olarak verir. Bir BOBUT uygulaması başladığında, MFI yöntemine göre seçilen maddeler, geçici yetenek kestirimlerinde yapmak üzere tasarlandıkları kadar fazla bilgi sağlama eğilimindedir (Han, 2012). MFI yöntemiyle ilgili diğer bir sorun, daha yüksek a parametresi değerlerine sahip maddeleri daha düşük a parametresi değerlerine sahip maddelere göre daha sık seçme eğiliminde olmasıdır (Chang, 2004; Han, 2012). MFI yöntemiyle madde havuzunda oluşan bu dengesiz madde kullanımı, madde havuzunun sürdürülebilirliğinde ciddi sorunlar yaratabilir.

Maksimum Bilgi Tabakalama (MIS). BOBUT uygulamalarında madde seçimine yönelik standart bir yaklařım, bir sonraki madde seçimi yapılırken kestirilen yetenek düzeyine ilişkin en yüksek bilgiyi veren maksimum Fisher bilgisine sahip maddenin seçilmesi olmuřtur (Lord, 1977). Bunu yaparken, bazı maddeler diğerlerinden daha sık kullanılırken, bazıları hiçbir zaman kullanılmamaktadır. Bu durum madde kullanım oranlarının oldukça dengesiz hale gelmesiyle, madde havuzunda kullanılmayan bu maddeleri geliřtirmek için harcanan emek ve para göz önüne alındığında birincisi ekonomik, ikincisi ise sık kullanılan maddelerin de ifřa olma riski nedeniyle test güvenliğini etkilemesi gibi sorunların ortaya çıkmasına neden olmuřtur (Barrada vd., 2006).

Madde havuzundaki maddelerin aşırı kullanımını önlemeye yönelik bazı madde seçme yöntemleri geliřtirilmiştir. Bunlardan bazıları a tabakalama ve b-blocking a tabakalama yöntemleridir. Fakat bu yaklařımlar sadece a ve b parametrelerini dikkate almaktadır. Bununla birlikte bahsi geçen yaklařımlarda maddelerin sınıflandırılması veya seçimi için göz ardı edilen ve üç parametrelili lojistik modelde (3PLM) řans başarısı olarak adlandırılan bir parametre daha bulunmaktadır. c parametresi, a tabakalama ve b-blocking a tabakalama yaklařımlarını geliřtiren arařtırmacılar tarafından yöntemde dahil edilmemiřtir.

Barrada vd. (2006), c parametresini de dahil ederek yöntemlerde iki basit deęişiklik yapmış ve maksimum bilgi tabakalama yöntemini geliştirmişlerdir. Bu bağlamda, madde havuzunu tabakalandırmak için a parametresini kullanmak yerine, onu Fisher bilgi fonksiyonu $I(\theta)_{max}$ 'taki bir madde tarafından elde edilen maksimum deęerle deęiştirmişlerdir. Bu durum aşığıdaki eşitlikte verilmiştir:

$$I(\theta)_{max} = \frac{1.7^2 a^2}{8(1 - c^2)} [1 - 20c - 8c^2 + (1 + 8c)^{3/2}]$$

Daha sonrasında, a tabakalama ve b-blocking a tabakalama yöntemlerindeki madde seçme kuralında ve b-blocking a tabakalama yönteminde tabakalama işlemindeki b deęerini θ_{max} ile deęiştirmişlerdir. θ_{max} 'ın elde edildięi θ deęeri aşığıdaki eşitlikte verilmiştir:

$$\theta_{max} = b + \frac{\ln[1 + (1 + 8c)^{1/2}] - \ln(2)}{1.7a}$$

Güçlük Eşleştirme Kriteri (GEK, b-matching). Urry'nin kriteri olarak da adlandırılan bu madde seçme yönteminde, güçlük düzeyi geçici yetenek kestirimine mümkün olduğunca en yakın olan madde seçilir (Han, 2018; Magis vd, 2018). Bu yaklaşım, madde güçlüklerinin θ deęerine en yakın olduęu noktada en fazla bilgiyi verdięi için bir parametrelili lojistik model veya Rasch modeli altında yaygın olarak kullanılmaktadır. Ayrıca bu yöntem, bir parametrelili lojistik veya Rasch modeli kullanıldığında MFI yöntemiyle temelde aynı madde seçim örüntüsü ile sonuçlanır (Han, 2018; Magis & Raiche, 2012). Güçlük eşleştirme kriteri, MFI yönteminden farklı olarak, yalnızca daha yüksek a parametre deęerlerini seçme eğilimi göstermediğinden, 1PL modelin yanı sıra genellikle 2PL veya 3PL modelleriyle de kalibre edilen maddelerle birlikte kullanılmaktadır (Han, 2018).

Sonlandırma Kuralı

Wainer'e (2000) göre, önceden belirlenmiş sayıda madde uygulandıktan, önceden belirlenmiş bir ölçme kesinlięi düzeyine ulaşıldıktan veya önceden belirlenmiş bir süre geçtikten sonra bir BOBUT uygulaması tamamlanmış sayılabilir. Bir BOBUT uygulamasının ne zaman tamamlandığını belirlemek için en sık kullanılan iki yöntem, sabit uzunluk ve

değişken uzunluk durdurma kurallarıdır (Choi vd., 2010; Sireci, 2003; Stafford vd., 2019; Wainer, 2000).

Sabit uzunluk durdurma kuralı altında, önceden belirlenmiş sayıda madde uygulandığında BOBUT uygulaması sonlandırılır (Babcock & Weiss, 2009; Choi vd., 2010; Stafford vd., 2019). Buna göre, testin sona ermesinden sonra elde edilen ölçme kesinliği düzeyine bakılmaksızın, uygulamaya katılan tüm bireylere aynı sayıda madde uygulanır. Sabit uzunluk durdurma kuralının en önemli avantajı basit olmasıdır. Bununla birlikte, sabit uzunluk durdurma kuralının uygulanmasının bir sonucu da bireylerin uç yetenek seviyelerinde meydana gelen daha büyük ölçme hatasıyla, farklı kesinlik dereceleriyle ölçülecek olmasıdır. Ayrıca, bireylerin yetenek düzeyi hakkında çok az bilgi sağlayan maddelerin gereksiz yere kullanılması BOBUT uygulamasının etkinliğini sınırlayabilir (Choi vd., 2010).

Değişken uzunluk durdurma kuralları, önceden belirlenmiş bir ölçme standardı karşılandıktan sonra madde uygulamasını sonlandırarak BOBUT uygulamasına katılan bireyler arasında eşdeğer ölçme duyarlılığı sağlamak için tasarlanmıştır. Bu testler değişken uzunluktadır çünkü uygulamaya katılan bireyler, testi sonlandırmak için kriter karşılanmadan önce farklı sayıda madde alabilirler. Araştırmacılar, uygulamaya katılan tüm bireylerin becerisine ilişkin psikometrik olarak sağlam bir tahmin yürütebilmek için gerektiği kadar az maddenin uygulanmasını amaçlayan, ancak uygulamaya katılan bireylerin yeteneğinin yeterince ölçüldüğünü belirtmek için kullanılan kriterlerde farklılık gösteren birkaç değişken uzunluklu durdurma kuralı geliştirmiştir (Stafford vd., 2019). Bu kurallardan biri, mevcut yetenek tahmininin önceden belirlenmiş bir standart hatasına ulaşıldığında madde uygulamasını sonlandıran standart hata (SE) durdurma kuralıdır (Dodd vd., 1989). Örneğin, standart hata 0,25 veya daha az olduğunda test durdurulabilir. Bu, her iki tarafta ± 2 standart hata ile %95'lik bir güven aralığının yaklaşık bir θ birim genişliğinde olacağı anlamına gelir. Bu durdurma kuralı, madde havuzunun uygun şekilde geliştirildiğini

varsayarak, uygulamaya katılan tüm bireyler için eş kesin puanlar üretme avantajına sahiptir (Thompson & Weiss, 2011).

Değişken uzunluk durdurma kurallarından başka bir yaklaşım olan minimum bilgi kriterinde ise; sonlandırma kriteri bir katılımcıya ait parametrenin yerine madde havuzuna dayandırılmaktadır (Thompson & Weiss, 2011). Bu kuralda, madde havuzunda madde seçim algoritmasına dayalı en azından bir miktar minimum bilgi sağlayabilecek hiçbir madde kalmamışsa, uygulamaya değer başka madde kalmadığı için test durdurulabilir (Choi vd., 2010; Gialluca & Weiss, 1979; Maurelli & Weiss, 1981; Thompson & Weiss, 2011).

Bu çalışmada, uygulanması daha kolay ve madde havuzu kullanım oranlarını daha iyi kestirme avantajına sahip olduğu için sabit uzunluklu durdurma kuralı kullanılmıştır (Thissen & Mislevy, 2000). Diğer bir yandan, sabit uzunluklu durdurma kuralının değişken uzunluk durdurma kurallarından daha iyi bir içerik planı gerçekleştirme sağladığı bulunmuştur (Cohen & Albright, 2014; Babcock & Weiss, 2014) . Ayrıca, sabit uzunluklu durdurma kuralı kullanıldığı zaman bu durumun uygulamaya katılan bireylere daha kolay açıklanmasının yanısıra bireylerde adalet algısının oluşmasına katkı sağlayacaktır. Çünkü uygulamaya katılan farklı bireyler farklı sayıda madde cevaplırsa, nispeten kısa testlerde düşük performans gösteren bireyler, yeterliliklerini kanıtlamak için diğerleriyle aynı fırsatı elde etmediklerini iddia edebilirler (Bergstrom & Lunz, 1999).

Madde Havuzunda Bazı Sınırlandırmalar

BOBUT uygulamalarının en önemli bileşeni madde seçim algoritmasıdır. Yaygın olarak kullanılan yöntem ise mevcut yetenek düzeyinde değerlendirilen en büyük Fisher bilgisine sahip bir sonraki maddeyi seçen maksimum bilgi yöntemidir (Thissen & Mislevy, 2000). Bu yaklaşım teorik olarak en etkili yetenek kestirimini sağlayabilir. Fakat bir BOBUT uygulamasına ait madde havuzunda istatistiksel iyileştirmeye ek olarak çeşitli istatistiksel olmayan sınırlandırmaların da dikkate alınması gerekmektedir. İstatistiksel olmayan sınırlandırmalar ise şu şekilde sıralanabilir (Cheng & Chang, 2009):

- Her içerik alanından belirli bir oranda madde seçilmelidir (içerik ağırlıklandırma olarak bilinir),
- Teste ait doğru seçenekler A, B, C ve D vb. seçeneklere yaklaşık olarak eşit dağıtılmalıdır (cevap anahtarı ağırlıklandırma olarak bilinir),
- Testte kökü olumsuz olan maddeler gibi “özel” maddeler sınırlı sayıda kullanılmalıdır.

Madde havuzunda bazı sınırlandırmaların dahil edildiği BOBUT uygulaması, tipik olarak içerik ağırlıklandırma ve madde kullanım sıklığı kontrolü sınırlandırmaları olan bir BOBUT uygulamasını ifade etmektedir (He, 2010). Bu bağlamda istatistiksel olmayan diğer bir sınırlandırma ise madde kullanım sıklığı kontrolüdür, yani her maddenin kullanım oranının belirli bir eşiğin altında tutulmasıdır. Madde kullanım sıklığının kontrol edilmesinin nedeni, yüksek kullanım sıklığı oranlarına sahip maddeler ile uygulamaya katılan farklı bireylerin karşılaşma ihtimalinin yüksek olması ve bundan dolayı da katılımcıların daha önce uygulamayı alan bireylerden bilgi toplayarak bu maddeler hakkında ön bilgi edinebilmelerinin önlenmesine yöneliktir (Cheng & Chang, 2009; Gu & Reckase, 2007). Bu nedenle madde havuzunda, maddelerin yüksek kullanım sıklığı oranına sahip olmasını önlemek son derece önemlidir (Chang & Ying, 1999; Davey & Parshall, 1995; Lunz & Stahl, 1998; Stocking & Lewis, 1995, 1998; Sympson & Hetter, 1985).

İçerik Ağırlıklandırma

İçerik alanı ve madde formatı gibi sınırlandırmaların karşılanması durumu içerik ağırlıklandırma kavramı ile ifade edilmektedir (Wang, 2017). Farklı içerik alanlarında madde güçlük dağılımları farklı olabileceğinden, bireylere uygulanan her BOBUT’da benzer içerik yapılandırmasını sağlamak için bu tür sınırlandırmaların dikkate alınması gerekir. Bir BOBUT uygulamasında, bireylerin göreceli karşılaştırılabilir test puanlarını elde etmek için içerik alanına göre aynı madde dağılımını almaları gerekmektedir (Stocking & Swanson, 1993). Bu durumu karşılamak için, içerik ağırlıklandırma sınırlandırması her zaman madde seçim algoritması aracılığıyla uygulanır (Kingsbury & Zara, 1991).

İçerik ağırlıklandırma, BOBUT uygulamalarında madde havuzunda pratik sınırlandırmalarla ilgili bir özelliktir. Test planı veya içerik özellikleri üzerine inşa edilen geleneksel kâğıt-kalem testlerinden farklı olarak, BOBUT uygulamaları madde seçimi sırasında içerik özelliklerini takip etmez. Bu nedenle, uygulamaya katılan bireylere içerik alanına göre farklı madde dağılımı uygulanabilir. Örneğin, bir matematik testinde, bireylerden biri tamamen aritmetik sorularından oluşan bir test alırken diğeri tamamen geometri sorularından oluşan bir test alabilir. Bu farklı içerik dağılımlarının uygulanması durumu, puanların geçerliliği için bir tehdit oluşturabilir ve bireyler için kabul edilebilir olmayabilir (Kinsbury & Zara, 1991).

Bireylerin aynı içerik dağılımı ile karşılaşmasını sağlayan içerik ağırlıklandırma yöntemleri, uygulamaya katılan her bireyin her içerik alanından yaklaşık olarak aynı oranda madde ile karşılaşmasını sağlar. Farklı araştırmacılar tarafından farklı içerik ağırlıklandırma yöntemleri geliştirilmiştir. Sınırlandırılmış bilgisayar ortamında bireye uyarlanmış test uygulamaları yöntemi (SBOBUT; Kingsbury & Zara, 1991), ağırlıklı sapma modeli yaklaşımı (ASM; Swanson & Stocking, 1993), gölge testi yaklaşımı (GTY; van der Linden & Reese, 1998), modifiye edilmiş sınırlandırılmış bilgisayar ortamında bireye uyarlanmış test uygulamaları yöntemi (MESBOBUT; Leung, Chang & Hau, 2003) ve maksimum öncelik indeksi yaklaşımı (MÖİ; Cheng & Chang, 2009) gibi çeşitli yaklaşımlar içerik ağırlıklandırma yöntemleri arasındadır (He, 2010).

Cheng vd. (2007) yaptıkları araştırmada, içerik ağırlıklandırma yöntemlerini kapsamlı bir şekilde ele alan bir çalışma yapmışlardır. Çalışmalarına bağlı olarak, içerik ağırlıklandırma her bir içerik alanındaki madde sayısı sabit sınırlandırma yapılarak sabitlenebilir veya her içerik alanındaki madde sayısının üst ve alt sınırlar tarafından tanımlanan bir aralık içinde sınırlandırılmasıyla esnek olabilmektedir. Birçok çalışma, sabit içerik ağırlıklandırmada içerik ağırlıklandırma yöntemlerini kullanmıştır. Örneğin Kingsbury ve Zara (1989), istedik içerik oranının en altında olan içerik alanından en uygun maddeyi seçerek sınırlandırılmış bilgisayar ortamında bireye uyarlanmış test uygulamaları

(SBOBUT) yöntemini önermektedir. Leung vd. (2003b), istedik içerik oranlarını karşılamayan tüm içerik alanlarından maksimum bilgiye dayalı olarak maddeleri seçerek SBOBUT yöntemini değiştirmişlerdir. Chen ve Ankenmann (2004), modifiye edilmiş multinominal modeli (MMM) önermektedir (Lin, 2011).

Alanyazına göre, ölçme kesinliği açısından SBOBUT yönteminin diğer içerik ağırlıklandırma yöntemleriyle benzer sonuçlar verdiği belirtilmiştir (Leung vd., 2003b). Bu nedenle bu çalışmada içerik ağırlıklandırma yöntemi olarak SBOBUT yöntemi kullanılmıştır.

Sınırlandırılmış Bilgisayar Ortamında Bireye Uyarlanmış Test Uygulamaları Yöntemi (SBOBUT). Kingsbury ve Zara (1989)'ya göre bir BOBUT uygulamasında karşılaşılması muhtemel sorunlardan biri de uygulama süresince bireylere uygulanan maddelerin, ölçmeyi hedeflediğimiz yapıya ait içerik alanına dengeli bir şekilde dağılıp dağılmayacağı konusudur. Bunun için madde havuzunda yer alan maddelere ait çeşitli içerik alanlarının her birine ait bir yüzde oranı önceden belirlenmelidir.

Geleneksel BOBUT uygulamaları için madde seçim yöntemleri, bir bireyin karşılaşılabileceği maddeleri seçme aşamasında sadece madde parametre kestirim değerlerine bağlıdır. Bu uygulamaya katılan bir birey için çok bilgilendirici bir test puanıyla sonuçlanacak olsa da, okul yöneticileri veya uygulamaya katılan diğer bireyler için kabul edilebilir olmayabilir (özellikle lisans sınavı uygulamasında). Örneğin, uygulamaya katılan bir bireyin matematik testi yalnızca uzun bölme problemlerinden oluşuyorsa, öğrencinin genel matematik becerisinin yeterli bir değerlendirmesinin yapıldığı söylenemez. Veya bir lisans sınavı, doğrulanmış test planıyla sıkı bir uyum içerisinde olan içeriğe sahip maddelerden oluşmuyorsa, uygulamaya katılan bireylerin testin geçerli olmadığına dair düşüncelere kapılmasına neden olabilir. Testin müfredat veya test planıyla eşleşmesini sağlamak için bir testteki belirli herhangi bir alt hedef (veya içerik) alanını fazla test etmekten kaçınan bir yol izlenmelidir (Kingsbury & Zara, 1989).

Kingsbury ve Zara'nın (1989) geliştirmiş olduğu SBOBUT yöntemi, uygulanan içerik oranı ile önceden belirlenmiş istedik içerik oranı arasında oluşan farklılığın gözlemlendiği alt

içerik alanından en uygun maddenin seçilip bireye uygulanmasıyla gerçekleştirilir. Bir testin istendik içerik oranı, testteki alt hedef alanlarının her birinden gelmesi gereken test maddelerinin yüzdesi olarak tanımlanabilir. Örneğin, temel bir aritmetik testi %30 toplama, %30 çıkarma, %20 çarpma ve %20 bölme alt içerik alanlarından oluşabilir. Test için bu plan oluşturulup istendik içerik oranları belirlendikten sonra, bu oranlara göre dağılımı yapılan maddelerin yer aldığı testi oluşturmak için içerik ağırlıklandırmadan yararlanılır (Kingsbury & Zara, 1989).

Bir BOBUT uygulamasında, SBOBUT yöntemi ile içerik ağırlıklandırma adımları şu şekilde gerçekleştirilmektedir (Kingsbury & Zara, 1989):

1. Bireye ait geçici yetenek düzeyi, bir maddenin uygulanmasının ardından kestirilir.
2. Mevcut testte her bir alt içerik alanında uygulanmış olan maddelerin yüzdesi hesaplanır.
3. Bir önceki adımda hesaplanan yüzdeler, önceden belirlenmiş istendik içerik oranları ile karşılaştırılır ve en büyük tutarsızlığa sahip alt içerik alanı belirlenir.
4. En büyük farkın olduğu alt içerik alanı içinde, bireyin geçici yetenek düzeyi kestiriminde en çok bilgi veren madde seçilir ve bireye uygulanır.

Bu adımlar bir BOBUT uygulaması sonlanana kadar her bir madde için sürekli olarak tekrar eder ve uygulama sonlandığında istendik alt içerik oranlarına göre madde seçimi ve kullanımı yapılmış olur.

Madde Kullanım Sıklığı

BOBUT uygulamalarının geliştirilmesi sürecinde madde havuzundaki diğer bir sınırlandırma da test güvenliği ve madde havuzu kullanımı ile ilgilidir. Madde havuzunda madde seçimi yapılırken, bazı maddeler bilgisayar tarafından sıklıkla seçilme eğilimindedir ve geriye kalan birçok madde hiç seçilmez, bu nedenle de madde havuzunda yer alan maddelerin kullanım sıklığı oranları oldukça dengesiz hale gelir. Ayrıca BOBUT'lar

genellikle küçük katılımcı gruplarına sık zaman aralıklarında uygulandığından, uygulamaya bir önceki zaman diliminde giren bireyler daha sonra girecek olanlarla bilgi paylaşabilir ve bu da pek çok maddenin ifşa olma riskini artırır (Chang, 2004; Cheng vd., 2009; Davey & Nering, 2002; Yi vd., 2006).

Dengesiz kullanılan bir madde havuzu sürekli aynı maddeleri seçme eğiliminde olduğu için, gerçekte olduğundan çok daha küçük bir madde havuzu gibi çalışır. Bu nedenle madde havuzunu büyütmek maddelerin kullanım sıklığı sorunundan kurtulmak için tek başına yeterli bir çözüm değildir. Yalnızca işlevsel bir madde kullanım sıklığı kontrol yöntemi ile birleştiğinde, gerçekten büyük bir madde havuzu değeri elde edilmiş olur. Tüm madde kullanım sıklığı kontrol yöntemlerinin amacı, seçilme olasılığı yüksek maddelerin uygulanabilme sıklığını sınırlayarak madde kullanım sıklığı oranlarını daha dengeli bir seviyeye getirmektir (Davey & Nering, 2002).

Madde kullanım sıklığı; madde havuzunda bir kısıtlama olmadan BOBUT uygulamasındaki madde seçme sürecinde istatistiksel olarak iyi olan maddelerin daha çok seçilmesi ve özellikle BOBUT'un başlangıcında bazı maddelerin diğerlerine göre çok daha fazla kullanılması anlamına gelmektedir (Rudner, 1998). Bu problemi gidermek adına, geliştirilen madde kullanım sıklığı kontrol yöntemlerini Stocking (1993) iki grupta sınıflandırmaktadır. Bunlar; a) maksimum bilgi madde seçme yöntemine rastgele bir bileşen ekleyen yöntemler ve b) maddenin aşırı kullanımını kontrol etmek için her bir maddeye bir parametre atamaya dayalı yöntemlerdir. Benzer şekilde Way (1998), madde kullanım sıklığı kontrol yöntemlerini rastgelelik ve koşullu seçim olarak iki grupta sınıflandırmıştır. Fakat bu alanda yapılan çalışmalar ilerledikçe, bu sınıflandırmanın ötesinde yeni yöntemler ortaya atılmıştır. Chang ve Ying (1999) maddeleri, artan bir madde ayırt ediciliği düzeninde birkaç tabakaya ayıran çok aşamalı bir tabakalama tasarımı önermiştir. Bu yöntemin temelinde önce düşük ayırt ediciliğe sahip maddelerin bireylere sunulması ve daha sonrasında bireylerin yetenek kestirimi yapılırken yüksek ayırt ediciliğe sahip maddelerin uygulanması yatmaktadır. Madde kullanım sıklığı kontrol yöntemlerinin geliştirilmesi süreci ilerledikçe,

farklı arařtırmacılar farklı stratejileri bir araya getirmiřtir. Örneęin, rastgelelik ve kořullu seçim; kořullu seçim ile a-tabakalama; a-tabakalama ile gölge testi yaklařımları gibi yöntemleri kombine etmiřlerdir. Ayrıca, Luecht ve Nungester (1998) tarafından geliřtirilen Bilgisayar Ortamında Bireye Uyarlanmıř Ardıřık Test (CAST) gibi dięer yöntemler test formlarını önceden oluřturarak maddelerin ařırı kullanımını önceden kontrol etmeyi amaçlamaktadır. Bahsi geçen tüm bu yöntemler beř grup altında sınıflandırılabilir. Bunlar; a) rastgelelik stratejileri; b) kořullu seçim stratejileri; c) tabakalama stratejileri; d) birleřik stratejiler; ve e) çok ařamalı bilgisayar ortamında bireye uyarlanmıř test tasarımlarıdır (Georgiadou vd., 2007).

Rastgelelik stratejisi, maksimum bilgi veren maddelerden oluřan bir gruptan rastgele bir madde seçerek madde kullanım sıklıęını kontrol etmeye çalıřmaktadır. Kořullu seçim stratejisi ise, madde kullanım sıklıęı kontrol parametresi ile maddelerin kullanım sıklıęı oranını kontrol eder ve Sympson-Hetter yöntemi (SH; Sympson & Hetter, 1985) bu stratejinin ilk örnek versiyonudur (Lin, 2011). SH, istenen maksimum madde kullanım sıklıęı oranını belirleyen (r^{max}) ve madde havuzundaki her bir maddenin kullanım sıklıęı oranını bu deęerden daha az veya yaklařık olarak bu deęere eřit olacak řekilde sınırlamayı amaçlayan yerleřik bir madde kullanım sıklıęı kontrolü yöntemidir (Huebner, 2012). Bu yöntemde, bir madde seçildikten sonra, maddenin uygulanıp uygulanmadıęını belirlemek için bir olasılık deneyi gerçekteřtirilir. Bu deneydeki (kořullu) olasılıklar manipüle edilerek, maddelerin kullanım sıklıęı oranları sınırların altında tutulur (Eggen, 2007; Veldkamp & van der Linden, 2000). Daha sonrasında bu yöntemin çeřitli modifikasyonları da geliřtirilmiřtir (Davey & Nering, 1998; Stocking & Lewis, 1998). Bununla birlikte bu yöntemde simülasyonların çok uzun sürmesi ve madde kullanım sıklıęı kontrol parametrelerinin yetenek daęılımı ve madde havuzu gibi BOBUT bileřenlerine baęımlı olması, yöntemin dezavantajları arasında görülebilir (Lin, 2011). Dięer bir dezavantajı ise, SH yönteminin doęası gereęi olasılıksal olduęu ve özellikle ardıřık madde seçiminde maddelerin büyük bir bölümünün r^{max} oranını ařması olası bir durumdur. Ayrıca SH yöntemi, madde havuzunda

aşırı kullanılan maddelere karşı duyarlı iken, hiç kullanılmayan ya da çok az kullanılan maddelere karşı duyarlı değildir (Eggen, 2007). Bu araştırmada; rastgelelik stratejilerinin, karmaşık ama gerçekçi madde havuzu yapıları ve test özelliklerinin olduğu durumlarda bazı maddelerin aşırı kullanımını engellemede yetersiz kaldığından (Stocking & Lewis, 1995) ve SH yönteminin dezavantajları sebebiyle, koşullu seçim stratejilerinden olan kısıtlayıcı yöntem (KY; Revuelta & Ponsada, 1998) ve madde uygunluk indeksi yöntemi (MUİ; van der Linden & Veldkamp, 2004) kullanılmıştır.

Maksimum madde kullanım sıklığı oranlarını kontrol etmeye yönelik SH, KY ve MUİ yöntemlerinin tümü, madde havuzundaki her bir maddeye bir madde kullanım sıklığı kontrol parametresi atar. Bu üç yöntemin arasındaki fark, madde kullanım sıklığı kontrol parametrelerinin nasıl ve ne zaman hesaplandığıdır. $m = 1, \dots, M$ havuzda yer alan maddeler için bir indeks olmak üzere ve m 'inci madde için madde kullanım sıklığı kontrol parametresinin değeri K_m ile gösterilir. Bu yöntemler, madde seçimini madde uygulama sürecinden ayırır; seçilen bir maddenin uygulanması zorunlu değildir. $P(S_m)$ ve $P(A_m)$ sırasıyla m maddesinin seçimi ve uygulanması için olasılıklardır ve böylece $P(A_m) \leq P(S_m)$ olmaktadır. K_m , her yöntemde $P(S_m)$ veya $P(A_m)$ değerleri kullanılarak hesaplanır ve K_m her zaman $[0,1]$ aralığında değerler alır. Sınava giren i için bir madde seçildiğinde, 0 ile 1 arasında rasgele bir tek biçimli U sayısı üretilir ve madde yalnızca $U < K_m$ ise uygulanır. Bir madde seçildikten sonra, uygulanmış olsun ya da olmasın, uygulamaya katılan birey artık o maddeyle karşılaşmayacaktır (Huebner, 2012). KY'de ve MUİ'de K_m 'nin nasıl elde edildiğine yönelik bilgiye aşağıda değinilmiştir.

Kısıtlayıcı Yöntem (KY). Revuelta ve Ponsada (1998), KY'nin maksimum madde kullanım sıklığı kontrolünü uygulama ile eşzamanlı yapan bir yöntem olduğunu belirtmişlerdir. SH yönteminin aksine, gerçek test başlamadan önce madde kullanım sıklığı kontrol parametreleri belirlenmez. Bunun yerine, uygulamaya katılan birey sayısı arttıkça gözlemlenen madde kullanım sıklığı oranlarına göre madde kullanım sıklığı kontrol parametrelerini sürekli olarak düzenler. Genel olarak, eşzamanlı yöntemler, $K_m^{(i)}$ 'nin

belirlenmesinde ön simülasyonlar gerektirmez ve $K_m^{(i)}$, havuzdaki diğer maddelere bağlı değildir. KY'in matematiksel olarak ifadesi şu şekildedir: i . birey uygulamaya katıldıktan sonra m maddesi için madde kullanım sıklığı kontrol parametresi $K_m^{(i)}$ olarak gösterilir ve bir maddenin uygulanma olasılığı 1'den i 'ye kadar olan bireylerden elde edilen $p^{1...i}(A_m)$ ile gösterilir. Ardından, uygulamaya katılan $(i + 1)$ birey için kontrol parametreleri şu şekilde verilir (Huebner, 2012):

$$K_m^{(i+1)} = \begin{cases} 1 & \text{eğer } P^{(1...i)}(A_m) \leq r^{max} \\ 0 & \text{eğer } P^{(1...i)}(A_m) > r^{max} \end{cases}$$

Eşitlikte madde kullanım sıklığı kontrol parametrelerinin yalnızca 0 ve 1 değerlerini alabileceği ve herhangi bir maddenin kullanım sıklığı oranının r^{max} 'ı aşmasının imkansız olduğu görülmektedir. KY, kavramsal açıdan basit bir yöntem olarak ifade edilebilir (Huebner, 2012). Chen vd. (2008) ile Barrada vd. (2009a) KY'nin bazı maddeler için öngörülebilir kullanım sıklığı örüntüleri üretmesinin yöntemin bir dezavantajı olduğuna dikkat çekmişlerdir. Örneğin bir ya da bir dizi madde n 'in belirli bir madde havuzuna ve r^{max} 'a bağlı olduğu durumda her n . kişiye gösterilebilir. Bu nedenle uygulamaya katılan bireyler arasındaki madde örtüşme oranı, genel örtüşme oranından çok daha büyük olacaktır (Huebner, 2012).

Madde Uygunluk İndeksi Yöntemi (MUİ). van der Linden ve Veldkamp (2004), madde kullanım sıklığı kontrol sorununu sınırlandırılmış bir test uygulamasının problemi olarak formüle etmeyi önermişlerdir. Bu yöntemde SH yönteminde olduğu gibi bir olasılık algoritması kullanılır. Ancak bu yöntem, olasılık deneyi için madde kullanım sıklığı kontrol parametrelerini bulmak için zaman alan simülasyon çalışmalarına ihtiyaç duymamaktadır. Algoritma, gözlemlenen madde kullanım sıklığı oranlarına dayalı olarak, BOBUT'da maddelerin seçilmesi aşamasında, modele madde uygunluk sınırlandırmalarının eklenip eklenmeyeceğini belirler. Bu süreçte ilk olarak, bir maddenin uygun olup olmadığını belirlemek için bir olasılık deneyi yapılır. İkinci olarak, test uygulama modeline uygun olmayan sınırlandırmalar eklenir ve model çözülür. Daha sonra, uygunluk

sınırlandırmalarının eklenmesi uygulanamayan bir modele yol açarsa, bu sınırlandırmalar kaldırılır ve model çözülebilir bir hale getirilir (Veldkamp vd., 2010). Bir maddenin uygulamaya katılan $(i + 1)$. bireye uygun olma olasılığı şu şekilde ifade edilebilir (Huebner, 2012):

$$K_m^{(i+1)} = \begin{cases} 1 & \text{eğer } P^{(1\dots i)}(A_m) / K_m^i \leq r^{max} \\ r^{max} K_m^i / P^{(1\dots i)}(A_m) & \text{eğer } P^{(1\dots i)}(A_m) / K_m^i > r^{max} \end{cases}$$

Bu yöntemin çalışma prensibi şu şekildedir: Uygulamaya katılan her birey için, herhangi bir madde uygulanmadan önce uygun maddelerin bir alt kümesi oluşturulur. Her madde için, tek tip aralığa $(0, 1)$ ait rastgele bir sayı üretilir. Yalnızca bu sayı $K_m^{(i+1)}$ değerinden küçük olursa, o maddenin uygun olması demektir. BOBUT uygulaması sürecinde sadece bu uygun maddeler bireylere uygulanabilir. Süreç bu şekilde ilerleyerek, birkaç istisna dışında, tüm maddelerin kullanım sıklığı oranları r^{max} 'a eşit veya r^{max} 'ın altında olur (Barrada vd., 2009b).

MUİ yönteminde, r^{max} 'tan daha yüksek bir oranda kullanım sıklığına sahip maddelerin bile, küçük de olsa uygulanma şansı vardır. Bu nedenle bu yöntemde KY'nin öngörülebilirlik sorunu görülmemektedir (Barrada vd., 2009a; Huebner, 2012). MUİ yöntemi de KY gibi ön simülasyon gerektirmemekte ve K_m^i madde havuzundaki diğer maddelerden ve bireylerin yetenek dağılımından bağımsız olarak hesaplanabilmektedir (Huebner, 2012). Diğer bir yandan KY'den farklı olarak, MUİ yöntemi doğası gereği olasılıksaldır. Bu nedenle, en popüler maddelerden bazıları için maksimum kullanım sıklığı oranı (r^{max}) biraz ihlal edilebilir (Barrada vd., 2009a).

İlgili Araştırmalar

Kingsbury ve Zara (1991), istedik bir içerik planına karşılık gelen maddeleri seçmeyi sağlayan bir BOBUT uygulamasının geliştirilmesi için iki yöntemin karşılaştırmasını yapmışlardır. İki yetenek düzeyi dağılımı ve üç madde havuzu büyüklüğü altında 10.000 birey üzerinden küçük test tabanlı ve içerik ağırlıklandırma yapılan sınırlandırılmış BOBUT

uygulamasından elde edilen yetenek düzeyi kestirimlerini, ortalama mutlak hata, yanlılık değerleri açısından herhangi bir sınırlandırma yapılmayan BOBUT uygulamasından elde edilen yetenek düzeyi kestirimleri ile karşılaştırmışlardır. Araştırma sonucunda, yedi maddeden daha büyük her test uzunluğu için, içerik ağırlıklandırma yapılmayan BOBUT uygulamasının, içerik ağırlıklandırma yapılan BOBUT uygulamasına göre daha düşük bir ortalama mutlak hatayla sonuçlandığını bulmuşlardır. Küçük test tabanlı yaklaşımın ise en yüksek ortalama mutlak hata değerini verdiğini belirtmişlerdir. Bu sonuçlar göz önüne alındığında, bir BOBUT uygulamasının belirli içerik özelliklerini gerektirdiği durumlarda içerik ağırlıklandırma yapılan BOBUT uygulamalarının kullanılmasını önermişlerdir.

Newman (1995) yaptığı çalışmada, içerik ağırlıklandırma yapılmasının, BOBUT sertifika sınavı uygulaması üzerindeki etkilerini araştırmıştır. Kağıt-kalem testi olarak uygulanan önceki uygulamalarından elde edilen veri dağılımları ile simülasyon çalışmalarını yürütmüştür. BOBUT uygulamasının performansını, dört içerik ağırlıklandırma yöntemi ve on beş durdurma kuralı altında incelemiştir. Çalışmadan elde ettiği sonuçlara göre; uygun bir içerik ağırlıklandırma yöntemi kullanmanın, BOBUT uygulamalarının verimliliği veya doğruluğu üzerinde önemli bir olumsuz etkisinin olmadığı bulunmuştur. Ayrıca, yetenek kestiriminde oluşan yanlılığının, yetenek düzeyleri minimum geçme puanına yakın olan adaylar için geçme durumunun belirlenmesinde hafif sapmalara neden olduğunu da belirtmiştir.

Luecht vd., (1998), bir BOBUT uygulaması geliştirirken içerik ağırlıklandırma yapılmayan ve yapılan durumlarda meydana gelebilecek bazı değiş tokuşları deneysel olarak göstermişlerdir. BOBUT uygulamalarında içerik ağırlıklandırma göz ardı edildiğinde içerik geçerliliğinin bir puan ölçeğinde fiilen değişebileceğini vurgulamışlardır. Bu bağlamda yürütmüş oldukları çalışmada 2000 birey üzerinden 1000 maddelik bir havuzu her içerik alanında 500 madde olacak şekilde iki içerik alanına bölmüşlerdir. Sabit sonlandırma kuralı olarak test uzunluğunu 100 madde olacak şekilde uygulamışlardır. Çalışmayı içerik ağırlıklandırma yapılmayan ve yapılan her iki durumda her birey için en çok bilgi veren

maddelerin seçildiği iki farklı koşulda gerçekleştirmişlerdir. Elde edilen sonuçlara göre, BOBUT'da içerik ağırlıklandırma yapılmasının ölçme kesinliğinde bir miktar düşüşe yol açabileceğini belirtmişlerdir. İki simülasyon çalışmasından elde edilen bulguların, içerik ve istatistiksel değerlendirmeler arasında meydana gelebilecek bazı değiş tokuşlar olabileceğini gösterdiğini söylemişlerdir.

Chen (1999)'in yaptığı araştırmada, yetenek kestiriminin doğruluğu ve bir BOBUT'un başlangıç aşamalarında madde kullanım boyutu bakımından dört madde seçme yönteminin (MFI, sonsal dağılımlı MFI, sonsal dağılımlı Kullback-Leibler ve rastgele madde seçme yöntemleri) etkilerini karşılaştırmak amaçlanmıştır. Dört madde seçme yönteminin karşılaştırılması üç koşul altında gerçekleştirilmiştir: (a) madde seçme yöntemi olarak sadece madde bilgi fonksiyonunun kullanılması, (b) hem içerik ağırlıklandırmanın yapılıp hem de madde bilgi fonksiyonunun kullanılması ve (c) madde bilgi fonksiyonunun kullanıldığı durumda, içerik ağırlıklandırma ve madde kullanım sıklığı kontrolü gibi sınırlandırmaların da dâhil edilmesi. Çalışma sonucunda, test uzunluğunun 10 maddeden az olduğu durumlarda tek madde seçme yöntemi madde bilgi fonksiyonu olduğunda sonsal dağılımlı MFI ve sonsal dağılımlı Kullback-Leibler, MFI'den daha iyi performans gösterdiği bulunmuştur. Bununla birlikte, madde seçme yöntemi olarak içerik ağırlıklandırma ve madde kullanım sıklığı kontrolünün dahil edildiği madde bilgi fonksiyonu koşulunda MFI madde seçme yönteminin, sonsal dağılımlı MFI ve sonsal dağılımlı Kullback-Leibler yöntemlerine göre daha iyi sonuç verdiği görülmüştür. Test uzunluğunun 10 maddeden fazla olduğu durumlarda, çalışmada karşılaştırılan üç koşulda da dört madde seçme yönteminin benzer sonuçlar verdiği bulunmuştur.

Leung vd. (2003a), üç tabakalama tasarım yöntemini üç farklı içerik ağırlıklandırma yöntemi (SBOBUT, modifiye edilmiş SBOBUT ve modifiye edilmiş multinominal model (MMM)) ile çaprazlanan koşullar altında karşılaştırmışlardır. Bunun için 700 maddelik bir havuz üzerinden 5000 birey ile simülasyon çalışmalarını yürütmüşlerdir. BOBUT uygulamasını 35 madde ile sonlandırmışlardır. Madde havuzunu 4 temel içerik alanına

bölerek içerik ağırlıklandırma yapıma durumuna uygun hale getirmişlerdir. Çalışmada madde seçme tasarımı olarak üç tabakalama tasarım yöntemi ile üç farklı içerik ağırlıklandırma yönteminin çaprazlandığı koşulları belirlemişlerdir. Yetenek kestirim yöntemi olarak ise MLE yöntemini kullandıkları mevcut çalışmanın sonuçlarına göre RMSE ve korelasyon değerleri açısından çaprazlanan 9 madde seçme tasarımının da benzer sonuçlar sunduğunu belirtmişlerdir. Diğer bir yandan, madde örtüşme oranı açısından modifiye edilmiş SBOBUT ile MMM'in birbirine yakın değerler verdiğini, SBOBUT yönteminin ise diğerlerine göre daha kötü bir performans sergilediğini vurgulamışlardır.

Leung vd. (2003b), üç içerik ağırlıklandırma yöntemi olan SBOBUT, modifiye edilmiş SBOBUT ve modifiye edilmiş multinominal model (MMM) yöntemlerini çeşitli test uzunluğu ve istenen madde kullanım sıklığı oranı koşulları altında karşılaştırmışlardır. Bunun için 700 maddelik bir havuz üzerinden 5000 birey ile simülasyon çalışmalarını yürütmüşlerdir. BOBUT uygulamasını 16, 28 ve 40 maddeden oluşan farklı test uzunlukları ile sonlandırmışlardır. Madde havuzunu 4 temel içerik alanına bölerek içerik ağırlıklandırma yapıma durumuna uygun hale getirmişlerdir. Madde seçme yöntemi olarak MFI, madde kullanım sıklığı kontrol yöntemi olarak da Sympson-Hetter yöntemini çalışmaya dâhil etmişlerdir. Yetenek kestirim yöntemi olarak ise uygulamanın başlangıç aşamalarında EAP, daha sonraki aşamalarda MLE yöntemini kullanmışlardır. Mevcut çalışmanın sonuçlarına göre, üç içerik ağırlıklandırma yönteminin RMSE, yanlılık ve korelasyon değerleri açısından karşılaştırılabilir ölçme kesinliği ve doğruluğu sunduğunu belirtmişlerdir. Ayrıca, test uzunluğunun ve istenen madde kullanım sıklığı oranının ölçme performansını etkileyen iki önemli faktör olduğuna kanaat getirmişlerdir.

Barrada vd. (2009), madde kullanım sıklığı kontrol yöntemi olarak kullanılan Sympson-Hetter, KY ve MUİ yöntemlerinin BOBUT uygulamalarındaki işlevselliklerini detaylı olarak açıklamayı, her yöntemin teorik ve deneysel sınırlandırmalarını belirtmeyi ve performanslarını karşılaştırmayı amaçlamışlardır. 500 maddelik bir madde havuzu kullanarak 5000 birey üzerinden çalışmalarını gerçekleştirmişlerdir. Test uzunluğunu 25

madde olarak belirledikleri çalışmalarında, madde seçme yöntemi olarak MFI, yetenek kestirim yöntemi olarak ise EAP ve MLE yöntemlerini kullanmışlardır. Çalışmalarından elde ettikleri bulgulara göre, RMSE ve madde çakışma oranı değerleri bakımından üç madde kullanım sıklığı kontrol yönteminde de benzer sonuçlar elde edilmiştir. Ayrıca KY'in, çakışma oranını kontrol etmede en iyi yöntem olduğu ve ardından da MÜ yönteminin geldiği bulgular arasında yer almaktadır. Diğer bir yandan en kötü yöntemin Sympson-Hetter yöntemi olduğu görülmüştür. Çalışmalarından elde ettikleri genel bir sonuca göre, KY'nin sıralı örtüşme oranı sorunları yarattığı ve bu sebeple de bu üç yöntem arasında hem zaman kazandırması hem de madde kullanım sıklığı oranını karşılaması bakımından MÜ yönteminin kullanılmasını önermişlerdir.

Cheng ve Chang (2009) yaptıkları çalışmada, sınırlandırılma yapılmış bir BOBUT uygulamasında madde seçme yöntemi olarak buluşsal bir yaklaşım olan maksimum öncelik indeksi (MÖİ) yöntemini tanıtmışlardır. Bunun için 3PLM ile kalibre edilen 642 maddelik bir havuz üzerinden bir milyondan fazla sayıda bireye ait olan bir gerçek veri setinden rastgele çekilen 5000 birey ile simülasyon çalışmalarını yürütmüşlerdir. BOBUT uygulamasını 12 madde ile sonlandırmışlardır. Madde havuzu, üç temel içerik alanına ve her bir içerik alanı ise 4-10 sayıda içerik alanına bölünerek oluşturulmuştur. Bu şekilde içerik ağırlıklandırma yapılan toplam 15 farklı içerik alanı çalışmaya dâhil edilmiştir. Madde seçme yöntemi olarak MFI ve rastgele madde seçme yöntemlerini, içerik ağırlıklandırma yöntemi olarak da MÖİ yöntemi ve ağırlıklandırılmış sapmalar modelini kullanmışlardır. Yetenek kestirim yöntemi olarak ise uygulamanın başlangıç aşamalarında 0-1 yanıt örüntüsünü oluşturana kadar EAP, daha sonraki aşamalarda MLE yöntemini kullanmışlardır. Çalışmadan elde ettikleri sonuca göre, ölçme kesinliği açısından en iyi performansı MFI yönteminin ve en kötü performansı da rastgele madde seçme yönteminin verdiğini belirtmişlerdir. Ayrıca iki farklı içerik ağırlıklandırma yöntemi kullanılarak içerik ağırlıklandırma yapılan koşullarda, ölçme kesinliğinde bir miktar kayıp meydana geldiğini vurgulamışlardır. İçerik ağırlıklandırma

yöntemlerinde ise MÖİ yönteminin biraz daha küçük yanlılık ve RMSE değerlerini vermesinin yanısıra daha düşük korelasyon değerlerine de neden olduğunu belirtmişlerdir.

Cheng vd. (2009), a-tabakalama yöntemini daha pratik ve daha yaygın olarak uygulanabilir hale getirmek için, çoklu test sınırlandırmalarını karşılamamanın bir yolu olarak a-tabakalama yönteminde madde seçme sürecini ağırlıklandırmak için bir yöntem önermişlerdir. Bunun için 3PLM ile kalibre edilen 1000 maddelik bir havuz üzerinden 3000 birey ile simülasyon çalışmalarını yürütmüşlerdir. BOBUT uygulamasını 40 madde ile sonlandırmışlardır. Beş içerik ağırlıklandırma yapıldığı durumda kullanılan madde seçme yöntemlerinin oluşturduğu koşul ve içerik ağırlıklandırma yapılmayan iki kontrol koşulu (MFI ve rastgele madde seçme yönteminin kullanıldığı koşul) olmak üzere 7 koşul oluşturulmuştur. İçerik ağırlıklandırma yapılan beş deneysel koşulun, tabakalama içermeyen ve dört farklı tabakalama tasarımı içeren yöntemlerden oluştuğunu belirtmişlerdir. Çalışmadan elde ettikleri sonuca göre, a-tabakalama yönteminde madde seçme sürecini ağırlıklandırma mekanizmasının sınırlandırmaları başarılı bir şekilde ele aldığını, tabakalamanın madde kullanım sıklığı oranlarını dengelemeye büyük ölçüde yardımcı olduğunu ve kullanılan beş deneysel koşulun arasında artan a-tasarımının ölçme kesinliğini iyileştirdiğini vurgulamışlardır. Ayrıca kullanılan madde seçme yöntemleri arasında en iyi performansı içerik ağırlıklandırma yapılmayan koşullarda MFI yönteminin verdiğini de belirtmişlerdir.

Song (2010) araştırmasında, içerik ağırlıklandırmanın yapıldığı bir BOBUT uygulamasında tek boyutlu bir MTK modelini çok boyutlu verilere uydurmanın etkisini incelemiştir. İçerik ağırlıklandırma yapılmadığı durumda MFI madde seçme yönteminin kullanıldığı koşul ile içerik ağırlıklandırma yapıldığı durumda MFI yöntemi ile birlikte SBOBUT, modifiye edilmiş multinominal model (MMM) ve modifiye edilmiş SBOBUT olmak üzere üç içerik ağırlıklandırma yönteminin performanslarını, ölçme kesinliği ve madde havuzu kullanımı ile madde kullanım sıklığı kontrolü açısından değerlendirmiştir. Bunun için 3PLM ile kalibre edilen 400 maddelik bir havuzdan elde ettiği tek boyutlu verileri çok boyutlu

veri türüne uyarlayarak 2000 birey ile simülasyon çalışmalarını yürütmüştür. BOBUT uygulamasını 30 madde ile sonlandırmıştır. Madde havuzu, üç temel içerik alanına 160, 120 ve 120 madde olacak şekilde bölünerek oluşturulmuştur. Madde seçme yöntemi olarak MFI, içerik ağırlıklandırma yöntemi olarak da SBOBUT, modifiye edilmiş MMM ve modifiye edilmiş SBOBUT yöntemini kullanmıştır. Yetenek kestirim yöntemi olarak ise uygulamanın başlangıç aşamalarında 0-1 yanıt örüntüsünü oluşturana kadar ilk beş maddenin uygulanmasında EAP, daha sonraki aşamalarda MLE yöntemini kullanmıştır. Simülasyon sonuçlarına göre, özellikle içerik alanları farklı zorluk seviyelerine sahip olduğunda, genel olarak içerik ağırlıklandırma yöntemlerinin ölçme kesinliği açısından MFI yöntemine benzer sonuçlar verdiğini vurgulamıştır. Ancak, madde havuzu kullanımı ve madde kullanım sıklığı kontrolünde anlamlı bir fark bulunmadığını belirtmiştir. Son olarak, üç içerik ağırlıklandırma yönteminin çok benzer şekilde değerler verdiğini, ancak modifiye edilmiş MMM'in en verimli madde havuzu kullanımına sahip olduğuna değinmiştir.

Zheng vd. (2013), 36 maddeden oluşan Short Form Health Survey (SF-36) anketini BOBUT uygulaması formatına dönüştürmüşlerdir. Daha sonra içerik ağırlıklandırma yapılmayan ve yapılan koşullara, MFI ve rastgele madde seçme yöntemleri ile EAP yetenek kestirim yöntemini dâhil ederek, bu koşulların ölçme kesinliği ile madde havuzu kullanımı üzerindeki etkisini incelemeyi amaçlamışlardır. Bu amaçla dört farklı post-hoc simülasyon çalışması gerçekleştirmişlerdir. Araştırmanın sonucunda, SF-36 anketinin madde seçme yöntemi olarak MFI yönteminin kullanıldığı koşullarda elde edilen RMSE değerlerine göre, içerik ağırlıklandırma yapılmayan koşullar ile içerik ağırlıklandırma yapılan koşullar benzer düzeyde değerler vermiştir. Çok az bir farkla içerik ağırlıklandırma yapılmayan koşullarda daha düşük RMSE değerleri elde edilmiştir. Diğer bir yandan madde havuzu kullanımı konusunda da, madde kullanım sıklığı oranının içerik ağırlıklandırma yapıldığı durumda bilgisayar ortamında bireye uyarlanmış test uygulamaları formunda daha dengeli olarak dağıldığı gözlenmiştir.

He vd. (2014) çalışmalarında, madde havuzunda bazı sınırlandırılmalar yapılan BOBUT uygulamalarında kullanılmak üzere geliştirilen dört madde seçme yöntemini karşılaştırmışlardır. Bunun için 3PLM ile kalibre edilen 361 maddelik bir havuz üzerinden 10000 birey ile simülasyon çalışmalarını yürütmüşlerdir. BOBUT uygulamasını 20 madde ile sonlandırmışlardır. Madde havuzu, 18 temel içerik alanına bölünerek oluşturulmuştur. Çalışmada ağırlıklandırılmış sapma modeli (ASM), ağırlıklandırılmış ceza modeli (ACM), maksimum öncelik indeksi (MÖİ) ve gölge testi yaklaşımını (GTY) çalışmaya dahil etmişlerdir. Yetenek kestirim yöntemi olarak ise uygulamanın başlangıç aşamalarında 0-1 yanıt örüntüsünü oluşturulana kadar EAP, daha sonraki aşamalarda MLE yöntemini kullanmışlardır. Çalışmadan elde ettikleri sonuçlara göre, GTY'nın ölçme kesinliği ve sınırlandırmaların yönetimi açısından tüm diğer yöntemler arasında en iyi şekilde çalıştığını belirtmişlerdir. Diğer üç buluşsal yaklaşım için, alt sınır seviyesinde ölçme kesinliği ve sınırlandırmaların yönetimi açısından önemli ölçüde farklılık göstermediğini vurgulamışlardır. Bununla birlikte, ACM yönteminin genel sınırlandırmaların yönetiminde ASM veya MÖİ yönteminden önemli ölçüde daha iyi performans gösterdiğini bulmuşlardır.

Şahin ve Özbaşı (2017) yaptıkları araştırmada, Maksimum Fisher Bilgisi (MFI) ve Olabilirlik Ağırlıklı Bilgi Ölçütü (LWI) yöntemlerini karşılaştırarak, bilgisayar ortamında bireye uyarlanmış test uygulamalarında içerik ağırlıklandırma ve madde seçimi yöntemlerinin, yetenek kestirimine etkilerini ortaya koymayı amaçlamışlardır. Bu amaçla, Monte Carlo simülasyon çalışması aracılığıyla dört grup (250, 500, 750, 1000) ve 10 farklı içerik alanlı 500 maddelik bir madde havuzu oluşturmuşlardır. Bahsi geçen madde seçme yöntemleri dışında kalan tüm bileşenler sabitlenerek katılımcıların yetenek kestirimi yapılmıştır. Katılımcılar altı alt gruba bölünerek gerçek ve kestirilen yetenek düzeyleri ile kullanılan maddelerin ortalama sayısı karşılaştırılmıştır. Olabilirlik ağırlıklı bilgi ölçütü kullanıldığında, katılımcıların yetenek düzeyi arttıkça, tüm katılımcılar arasındaki korelasyonun istikrarlı bir şekilde azaldığı görülmüştür. Araştırma sonucunda, Maksimum Fisher Bilgisi yönteminin Olabilirlik Ağırlıklı Bilgi Ölçütüne göre daha az hataya sahip olduğu bulunmuştur.

Bölüm 3

Yöntem

Bu bölümde araştırmanın türü, simülasyon deseni, verilerin üretilmesi, simülasyon uygulamaları ve verilerin analizi ile ilgili detaylı bilgilere yer verilmiştir.

Araştırmanın Türü

Bu araştırmada, bir BOBUT uygulamasında iki farklı büyüklüğe sahip madde havuzundan elde edilen ölçme kesinliği kestirim değerlerinin; farklı madde seçme, farklı yetenek kestirim yöntemlerine göre içerik ağırlıklandırma ve madde kullanım sıklığı kontrolü gibi bazı madde havuzu sınırlandırmaları altında nasıl değiştiğinin incelenmesi amaçlanmıştır. Araştırmada kullanılan madde ve yetenek parametreleri simülatif olarak üretilmiş ve BOBUT uygulaması simülasyonlar yardımıyla gerçekleştirilmiştir. Bu yüzden bu araştırma verilerin simülatif olarak elde edildiği Monte Carlo simülasyon çalışması özelliğini taşımaktadır.

Monte Carlo simülasyonu, gerçek veriler mevcut olmadığında bireylerin bir BOBUT uygulamasındaki davranışlarını modellemek için iyi bir alternatiftir. Bu yöntemin bir dezavantajı, teorik dağılımların gerçek bireylerin dağılımlarını ne derecede karşılayabildiği ile genelleştirilebilirliğinin sınırlı olmasıdır. Bununla birlikte, bir BOBUT uygulamasının bireylerin davranışını modelleme yeteneği, gerçek veri simülasyon yönteminde olduğu gibi, bilinen birey-madde etkileşimlerinin bir alt kümesiyle sınırlı değildir (Newman, 1995).

Araştırma Deseni

Monte Carlo simülasyonu kapsamında BOBUT uygulaması için oluşturulan koşullar; farklı büyüklükte madde havuzu, içerik ağırlıklandırma yapılmadığı ve yapıldığı durumlarda kullanılan farklı madde seçme, farklı yetenek kestirim ve farklı madde kullanım sıklığı kontrol yöntemlerini içermektedir. Araştırmaya dâhil edilecek olan 2 farklı büyüklükte madde havuzu, içerik ağırlıklandırma yapılmayan ve yapılan 2 farklı durum, 3 madde seçme

yöntemi, 2 yetenek kestirim yöntemi ve 2 madde kullanım sıklığı kontrol yöntemi için toplam 48 koşul belirlenmiştir. Böylelikle 2 x 2 x 3 x 2 x 2 matrisi ile belirlenen simülasyon koşulları aşağıdaki tabloda yer almaktadır:

Tablo 1

BOBUT Simülasyon Koşulları

Madde Havuzu Büyüklüğü	İçerik Ağırlıklandırma Durumu	Madde Seçme Yöntemi	Yetenek Kestirim Yöntemi	Madde Kullanım Sıklığı Kontrol Yöntemi	Koşul No
320 Madde	İçerik Ağırlıklandırma Yok	MFI	EAP	KY	1
		MFI	EAP	MUİ	2
		MFI	MLE	KY	3
		MFI	MLE	MUİ	4
		MIS	EAP	KY	5
		MIS	EAP	MUİ	6
		MIS	MLE	KY	7
		MIS	MLE	MUİ	8
		GEK	EAP	KY	9
		GEK	EAP	MUİ	10
		GEK	MLE	KY	11
		GEK	MLE	MUİ	12
	İçerik Ağırlıklandırma Var	MFI	EAP	KY	13
		MFI	EAP	MUİ	14
		MFI	MLE	KY	15
		MFI	MLE	MUİ	16
		MIS	EAP	KY	17
		MIS	EAP	MUİ	18
		MIS	MLE	KY	19
		MIS	MLE	MUİ	20
		GEK	EAP	KY	21
		GEK	EAP	MUİ	22
		GEK	MLE	KY	23
		GEK	MLE	MUİ	24
1000 Madde	İçerik Ağırlıklandırma Yok	MFI	EAP	KY	25
		MFI	EAP	MUİ	26
		MFI	MLE	KY	27
		MFI	MLE	MUİ	28
		MIS	EAP	KY	29
		MIS	EAP	MUİ	30
		MIS	MLE	KY	31
		MIS	MLE	MUİ	32
		GEK	EAP	KY	33
		GEK	EAP	MUİ	34
		GEK	MLE	KY	35
		GEK	MLE	MUİ	36
	İçerik Ağırlıklandırma Var	MFI	EAP	KY	37
		MFI	EAP	MUİ	38
		MFI	MLE	KY	39
		MFI	MLE	MUİ	40
		MIS	EAP	KY	41
		MIS	EAP	MUİ	42
		MIS	MLE	KY	43
		MIS	MLE	MUİ	44
		GEK	EAP	KY	45
		GEK	EAP	MUİ	46
		GEK	MLE	KY	47
		GEK	MLE	MUİ	48

Tablo 1’de görüldüğü gibi tüm koşullar, buldukları desen içerisinde birbiriyle çaprazlanarak oluşturulmuştur. Örneğin 320 maddelik madde havuzu için; içerik ağırlıklandırma yapılmayan ve yapılan 2 farklı durum, 3 farklı madde seçme, 2 farklı yetenek kestirim ve 2 farklı madde kullanım sıklığı kontrol yöntemi olmak üzere 24 koşul belirlenmiştir. 1000 maddelik madde havuzu için de aynı koşullar belirlenerek toplamda 24 koşul ele alınmıştır. Böylelikle, çalışma kapsamında toplam 48 koşul karşılaştırılmıştır.

Tablo 1’de yer alan koşulların bileşenleri ilgili alanyazın incelenerek belirlenmiştir. Örneğin; Magis vd. (2018), 1PLM altında GEK ve MFI yöntemlerinin, 1PLM ve 2PLM altında ise GEK ve MIS yöntemlerinin eşdeğer olduğunu belirtmişlerdir. Bu sebeple, 3PLM altında değerlerin nasıl değiştiğini incelemek amacıyla MFI, MIS ve GEK yöntemleri madde seçme yöntemleri olarak belirlenmiştir. Yetenek kestirim yöntemi bileşeninde ise, tüm yetenek düzeylerinde EAP yönteminin MLE yöntemine göre daha düşük hata vermesi sebebiyle (Bock & Mislevy, 1982) ve MLE yöntemine göre sağladığı üstünlükleri karşılaştırabilmek adına MLE ve EAP yetenek kestirim yöntemleri kullanılmıştır. Bu yöntemlerin performanslarının, içerik ağırlıklandırma yapılmadığı ve yapıldığı durumlar ile madde kullanım sıklığı kontrol yöntemleri altında nasıl değiştiğini görmek adına da madde havuzu farklı içerik alanlarına ayrılmış ve 2 farklı madde kullanım sıklığı kontrol yöntemi kullanılmıştır.

Uygulamadaki koşullar için teste başlama kuralı olarak $\theta = 0.00$ belirlenmiş olup bütün koşullar için 156 replikasyon yapılmıştır. Pekmezci ve Avşar (2021), Monte Carlo simülasyon çalışmalarından daha doğru ve kesin sonuç alınabilmesi için replikasyon sayısının 156 ve üstü olmasını önermektedirler.

Verilerin Üretilmesi

Verilerin üretilmesi sürecinde ücretsiz ve açık kaynak kodlu istatistiksel bir dil olan R programlama dilinden faydalanılmıştır (R Core Team, 2022). Öncelikle 320 ve 1000 maddelik iki havuz ve 2000 bireye ilişkin yetenek parametreleri üretilmiştir. Daha sonra

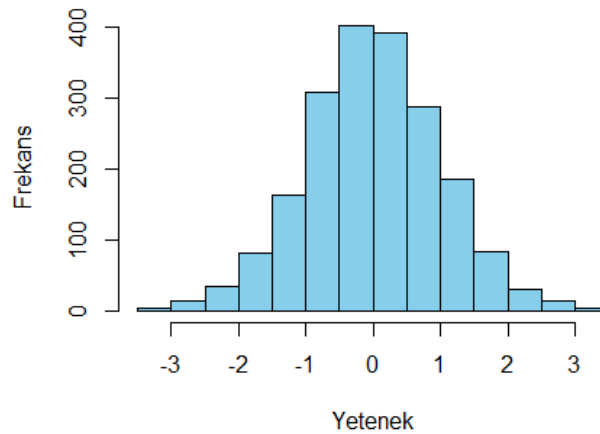
BOBUT uygulamaları için simülasyonlar oluşturulmuş; oluşturulan simülasyonlar aynı madde havuzları ve aynı bireyler üzerinden farklı koşullar altında karşılaştırılmıştır. Analizler catR (Magis vd., 2018) paketi ile Excel programı kullanılarak yapılmıştır.

Yetenek Parametrelerinin Üretilmesi

Araştırma kapsamında 2000 bireye ait yetenek parametreleri, standart normal dağılımdan $[N(0,1)]$ üretilmiştir. Uç değerler etkisinin göz ardı edilmesi için de yetenek parametreleri $(-3,+3)$ aralığında olacak şekilde sınırlandırılmıştır. Üretilen yetenek parametrelerinin dağılımı Şekil 4'te verilmiştir.

Şekil 4

Yetenek Parametrelerinin Dağılımı



Şekil 1'de görüldüğü üzere yetenek parametreleri $(-3, +3)$ aralığında kalacak şekilde normal dağılım göstermektedir. Ayrıca yetenek parametrelerine ait betimsel istatistikler hesaplanmış; ortalama, standart sapma, çarpıklık ve basıklık değerleri sırasıyla 0.002, 0.998, -0.028, -1.545 olarak elde edilmiştir.

Madde Havuzunun Oluşturulması

Madde havuzu büyüklüğünün ölçme kesinliği, test verimliliği ve içerik planlamanın yerine getirilmesi üzerinde büyük bir etkisi vardır (Wei & Lin, 2015; Babcock & Weiss, 2014; Davey, 2011; Dodd vd., 1993). BOBUT uygulamalarında farklı yetenek seviyelerindeki

bireyler için en uygun maddelerin seçilebilmesi için madde havuzunun yeterince büyük olması gerekir. Bu nedenle BOBUT uygulamalarında kullanılan madde havuzu, uygulamanın etkinliğini artırmak için bir dizi nitelikli maddeye sahip olmalıdır (Kalender, 2011). Davey (2011) BOBUT uygulamalarında madde havuzunun, sabit uzunluk sonlandırma kuralı ile gerçekleştirilen test uzunluğundan 10 kat daha fazla madde içermesi gerektiğini önermiştir. Ayrıca, bireylerin geleceği ile ilgili belirleyici olan sınavlardaki bir başarı testi için, içerik ağırlıklandırma yapıldığı takdirde daha büyük bir madde havuzunun (500'den fazla madde) istenebileceğini de belirtmiştir.

Ayala (2009) bir BOBUT uygulaması için gerekli madde havuzu büyüklüğünün kullanılan test uzunluğuna bağlı olduğunu belirtmiştir. Bu temel kurala göre bir madde havuzunda yer alan madde sayısı test uzunluğunun en az 8 ila 12 katı olması gerektiğidir. Örneğin, ortalama 25 maddelik bir BOBUT uygulaması için madde havuzunun 200 ila 300 maddeye sahip olması gerektiği ifade edilmiştir. Stocking (1994) ve Way (1998) ise, madde havuzunun test uzunluğunun en az 12 katı kadar sayıda maddeye sahip olmasını önermektedir. Bu çalışma kapsamında, literatürde yer alan bilgilere dayanarak farklı büyüklükteki madde havuzlarından elde edilen sonuçların karşılaştırılması amacıyla iki farklı büyüklükte madde havuzu oluşturulmuştur.

BOBUT uygulaması için kullanılacak olan iki farklı büyüklükte madde havuzu, R programıyla 3PLM'ye göre 320 ve 1000 madde olacak şekilde oluşturulmuştur. Flaugher (2000)'e göre BOBUT uygulamaları için nitelikli bir madde havuzu, yüksek ayrımcılığa (a 1'den büyük), dikdörtgen bir zorluk dağılımına ve düşük tahminli (c'nin 2'den küçük) maddelere sahip olmalıdır. Bu bağlamda iki farklı büyüklükteki madde havuzunda kullanılacak maddelerin tüm parametreleri tek biçimli (uniform) dağılıma sahip olacak şekilde; a parametresi için minimum değer 0.80 ile maksimum değer 2.00, b parametresi için minimum değer -3.00 ile maksimum değer 3.00 ve c parametresi için minimum değer 0.05 ile maksimum değer 0.20 olarak belirlenmiştir. Oluşturulan 320 ve 1000 maddelik havuzlardaki maddelere ait betimsel istatistikler Tablo 2 ve Tablo 3'te verilmiştir:

Tablo 2*320 Maddelik Havuza Ait Betimsel İstatistikler*

	a Parametresi	b Parametresi	c Parametresi
K	320	320	320
Ortalama	1,421	-0,069	0,124
Standart Sapma	0,348	1,654	0,044
Minimum	1,004	-2,921	0,050
Maksimum	1,995	2,956	0,199
Basıklık	-1,220	1,862	1.753
Çarpıklık	0,024	0,132	-0,038

Tablo 3*1000 Maddelik Havuza Ait Betimsel İstatistikler*

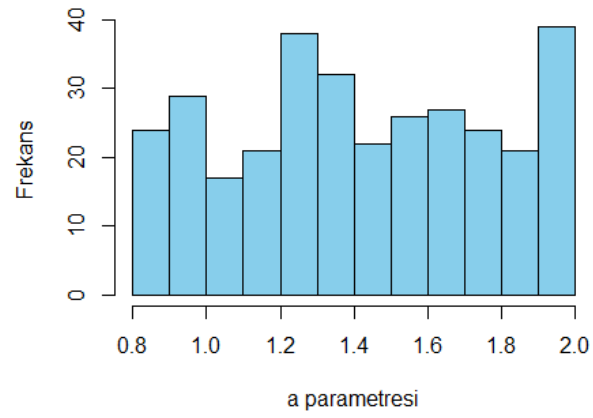
	a Parametresi	b Parametresi	c Parametresi
K	1000	1000	1000
Ortalama	1,388	-0,001	0,123
Standart Sapma	0,355	1,730	0,043
Minimum	0,800	-2,992	0,050
Maksimum	1,998	2,999	0,199
Basıklık	1,739	1,815	1,783
Çarpıklık	0,040	0,038	0,066

Tablolarda a, b ve c parametrelerine ait betimsel istatistik değerleri verilmiştir. Minimum ve maksimum değerlerin tüm parametreler için belirlenen aralıklar içerisinde kaldığı tablolardan anlaşılmaktadır. Madde havuzlarına ait madde parametrelerine ilişkin dağılımlar Şekil 5 ve Şekil 6'da verilmiştir:

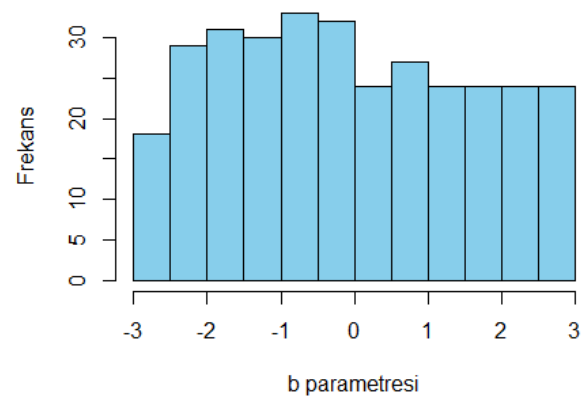
Şekil 5

320 Maddelik Havuza Ait Madde Parametre Dağılımları

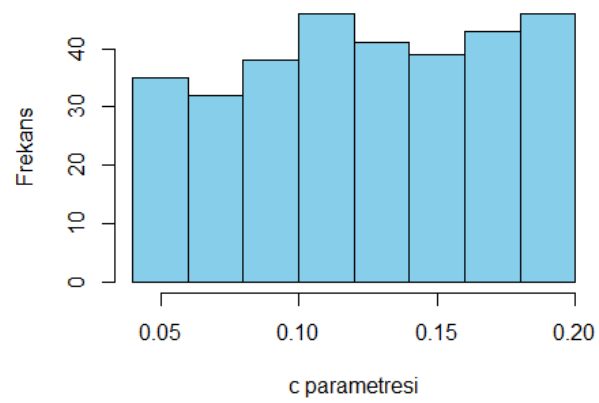
Şekil 5a



Şekil 5b



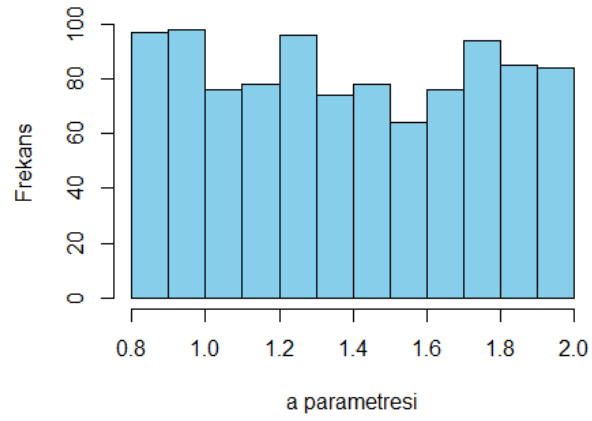
Şekil 5c



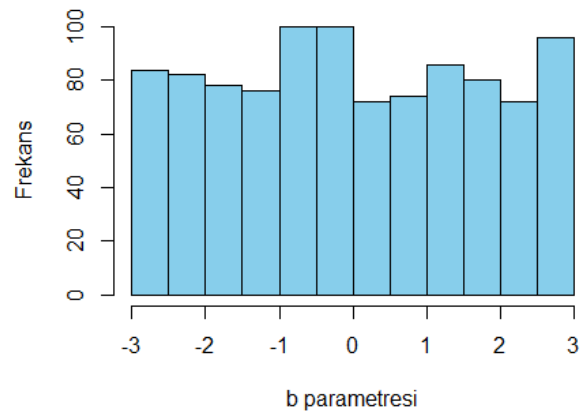
Şekil 6

1000 Maddelik Havuza Ait Madde Parametre Dağılımları

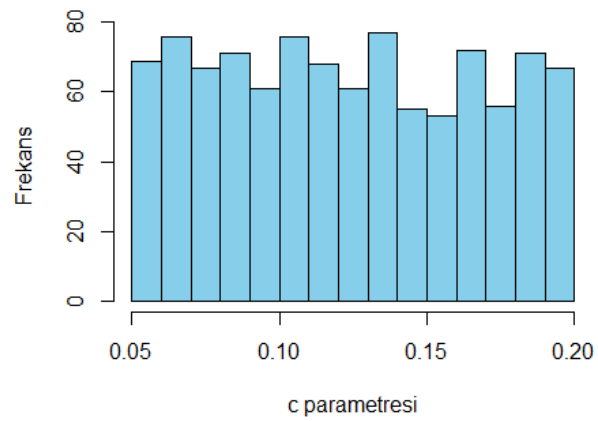
Şekil 6a



Şekil 6b



Şekil 6c



Madde havuzları, içerik ağırlıklandırma yapılabilmesi için 4 farklı içerik alanına ayrılmıştır. İçerik alanları, madde havuzlarında yer alan toplam madde sayısına bağlı olarak rastgele belirlenen 0.2, 0.4, 0.3 ve 0.1 oranlarına göre oluşturulmuştur. Buna göre, 320 maddelik madde havuzunda 1. içerik alanında 64, 2. içerik alanında 128, 3. içerik alanında 96 ve 4. içerik alanında 32 madde yer almaktadır. 1000 maddelik madde havuzunda ise 1. içerik alanında 200, 2. içerik alanında 400, 3. içerik alanında 300 ve 4. içerik alanında 100 madde yer almaktadır. Bu içerik alanlarına ait madde parametrelerinin ortalama ve standart sapma değerleri Tablo 4 ve Tablo 5'te, madde havuzlarına ait test bilgi fonksiyonları ise EK-A'da sunulmuştur.

Tablo 4

320 Maddelik Havuzda İçerik Alanları Boyunca Parametre Dağılımları

İçerik Alanı	Madde Sayısı	Ayırdedicilik (a)		Güçlük (b)		Şans Başarısı (c)	
		Ortalama	Standart Sapma	Ortalama	Standart Sapma	Ortalama	Standart Sapma
İçerik Alanı 1	64	1,420	0,341	-0,235	1,790	0,121	0,046
İçerik Alanı 2	128	1,410	0,346	-0,090	1,580	0,120	0,045
İçerik Alanı 3	96	1,430	0,352	0,060	1,670	0,132	0,045
İçerik Alanı 4	32	1,410	0,372	-0,049	1,650	0,128	0,039

Tablo 5

1000 Maddelik Havuzda İçerik Alanları Boyunca Parametre Dağılımları

İçerik Alanı	Madde Sayısı	Ayırdedicilik (a)		Güçlük (b)		Şans Başarısı (c)	
		Ortalama	Standart Sapma	Ortalama	Standart Sapma	Ortalama	Standart Sapma
İçerik Alanı 1	200	1,410	0,350	-0,044	1,750	0,125	0,042
İçerik Alanı 2	400	1,360	0,355	0,104	1,720	0,124	0,043
İçerik Alanı 3	300	1,390	0,358	-0,055	1,730	0,125	0,045
İçerik Alanı 4	100	1,440	0,359	-0,183	1,740	0,114	0,046

Verilerin Analizi

BOBUT uygulamaları, bireyin yetenek düzeyini kestirme sürecinde yapılan bu kestirimlerin istenilen ölçme kesinliği düzeyine ulaşmasını da gerektirmektedir. Hata kareleri ortalamasının karekökü (Root Mean Squared Error – RMSE; yanlışlık ve standart hata değerlerinin hesaba katılarak, parametre değerlerinde iyileştirmenin mutlak doğruluğunun

bir ölçüsü); yanlılık (bias; kestirilen yetenek düzeyinin gerçek yetenek düzeyinden sistematik sapmasının bir ölçüsü); gerçek ve kestirilen yetenek düzeyleri arasındaki ilişkiye bağlı olarak elde edilen korelasyon ve standart hata (yetenek kestiriminin standart hatasının bir ölçüsü), BOBUT uygulamalarında yetenek kestirimlerine ait ölçme kesinliğini değerlendirmek için yaygın olarak kullanılan istatistiklerdir (Babcock & Weiss, 2014; Piromsombat, 2014; Risk, 2015; Wang & Vispoel, 1998; Wang & Zhang, 2017; Yang, 2016). Bu çalışmada da benzer çalışmalarda olduğu gibi, bu değerler ölçme kesinliği kestirim değerleri olarak kullanılmış olup, araştırmaya dâhil edilen her bir koşulun ölçme kesinliği kestirim değerleri üzerindeki etkisinin nasıl değiştiğinin incelenmesi amacıyla; RMSE, yanlılık (bias), gerçek ve kestirilen yetenek düzeyleri arasındaki ilişkiye bağlı olarak elde edilen korelasyon değerleri hesaplanmıştır. 156 replikasyon sonucunda elde edilen bu değerlerin ortalaması alınarak koşulların değerlendirilmesi yapılmıştır.

Hata kareleri ortalamasının karekökü (RMSE); hem yanlılığın hem de standart hata değişkenliğinin bir fonksiyonudur. RMSE değeri ne kadar düşük olursa, bir BOBUT'da yapılan yetenek kestirim düzeyi bireyin gerçek yetenek düzeyine o kadar yakın olur. Wang ve Vispoel (1998), RMSE'yi kestirimlere ait yanlılık ve standart hata değerlerinin karekökünün alınması ile elde edilen iki bileşene sahip, toplam kestirim hatasının ölçüsü olarak tanımlamıştır. Yanlılık ise, kestirilen yetenek düzeyinin gerçek yetenek düzeyinden sistematik olarak sapmasının bir ölçüsü olarak tanımlanmıştır (Risk, 2015). Kestirilen yetenek düzeylerinin, gerçek yetenek düzeylerinin üstünde mi yoksa altında mı bir değer verdiğini göstermektedir. Yanlılık değerinin pozitif olması, kestirilen yetenek düzeyinin gerçek yetenek düzeyinin altında kaldığını gösterirken; negatif olması ise kestirilen yetenek düzeyinin gerçek yetenek düzeyinin üstünde olduğunu göstermektedir. 0'a yakın olması da daha doğru bir kestirim yapıldığına işarettir (Liu, 2019). N , toplam birey sayısını, θ_j , j . bireyin gerçek yetenek düzeyini ve $\hat{\theta}_j$, j . bireyin kestirilen yetenek düzeyini göstermek üzere RMSE ve yanlılık (bias) değerlerine ait formüller şu şekildedir:

$$RMSE = \sqrt{\frac{\sum_{j=1}^N (\hat{\theta}_j - \theta_j)^2}{N}}$$

$$Yanlılık = \frac{\sum_{j=1}^N (\hat{\theta}_j - \theta_j)}{N}$$

Ölçme kesinliği kestirim değerlerinden olan korelasyon ise; bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki ilişki kestiriminin doğruluğunu Pearson momentler çarpımı korelasyonu ile tanımlar. Hesaplanan korelasyon katsayıları ne kadar yüksekse tahmin o kadar doğru olur. Ayrıca elde edilen sonuçlarla daha güvenilir kararlar verilmiş ve dolayısıyla daha iyi bir madde seçme yöntemi belirlenmiş olur (Leung vd., 2002).

Koşullara ait ölçme kesinliği kestirim değerleri karşılaştırıldıktan sonra, koşullar arasındaki farkların anlamlılık düzeyine ulaşip ulaşmadığını belirlemek amacıyla faktöriyel ANOVA yapılmıştır. Faktöriyel ANOVA, iki veya daha fazla bağımsız değişken ile tek bir bağımlı değişkenin olduğu durumlarda kullanılmaktadır. Genel olarak bağımlı değişkenin, iki veya daha fazla bağımsız değişkenden nasıl etkilendiğini ve bağımlı değişken üzerinde bağımsız değişkenlerin arasında bir etkileşim etkisinin olup olmadığını anlamak amacıyla tercih edilmektedir. Bu çalışma kapsamında RMSE, yanlılık ve korelasyon değerlerinin ayrı ayrı bağımlı değişken, çalışmaya ait simülasyon koşulları olan toplam 48 koşulun ise bağımsız değişken olarak belirlendiği faktöriyel ANOVA analizi her bir koşul için gerçekleştirilmiştir. Ayrıca farklılıkların hangi gruplardan kaynaklandığını tespit etmek için Post-Hoc karşılaştırmaları yapılmıştır. Bu çalışma için $p < 0.05$ istatistiksel anlamlılık düzeyi kullanılmıştır.

Anlamlılık testleri büyük oranda örneklem büyüklüğüne bağlıdır. Örneğin, örneklem boyutunun büyük olduğu bir durumda, önemsiz olan etkiler bile çok önemliymiş gibi görünen p değerlerine sahip olabilmekte ve bu etkiler olduğundan daha önemli gibi görünebilmektedir. Kısacası, anlamlılık testlerinden elde edilen p değerleri, hem örneklem büyüklüğünü hem de incelenen etkilerin büyüklüğünü yansıtır (Levine & Hullet, 2002). Bu nedenle, gruplar arasındaki farklılığın ne kadar büyük olduğunu belirlemek için, etki

büyükülüğünün göstergesi olan eta kare (η^2) değerinin hesaplanması gerekebilir. Khalilzadeh ve Tasci (2017), endüstrilerin arz ve talep tarafının farklı yönlerinde biriken büyük verilerin artan kullanılabilirliğinin ve ilgisinin, akademik yayınlar için de büyük örneklemelerin kullanılmasına neden olduğunu belirtmişlerdir. Bununla beraber, büyük örneklemelerin istatistiksel anlamlılık değerinin elde edilmesini garantilediğini ve bu nedenle de etki büyüklüğü ölçümleri kullanılarak, uygulanabilir anlamlılığın rapor edilmesinin gerektiğine de değinmişlerdir. Aynı zamanda, etki büyüklükleri deneysel çalışmaların en önemli çıktısıdır (Lakens, 2013). Bu nedenlerden dolayı mevcut çalışmada, ANOVA analizleri sonuçlarını yorumlamada p değerinin yanında etki büyüklüğünün belirtkesi olan eta kare (η^2) değeri de hesaplanmıştır. Etki büyüklüğünü değerlendirirken yaygın olarak kullanılan yorumlama, Cohen (1988) tarafından önerilen ölçütlere dayalı olarak etki büyüklüklerini küçük ($d = 0,010$), orta ($d = 0,059$) ve büyük ($d = 0,138$) olarak ifade etmektedir.

Simülasyon Geçerliği

Simülasyon çalışmalarının teori geliştirmeye etkili bir şekilde katkıda bulunduğu bazı araştırmacılar tarafından iddia edilmektedir. Örneğin Zott (2003), özellikle zorlayıcı deneysel veri sınırlandırmaları söz konusu olduğunda simülasyon çalışmalarının yapılar arasındaki karmaşık teorik ilişkiler hakkında üstün bilgiler sağlayabildiğini belirtmiştir. Diğer bir yandan, sözel teorilerin temelinde yatan varsayımları ve teorik mantığı belirlemenin analitik olarak kesin bir yolunu da sağlayabildiği vurgulanmıştır (Carroll & Harrison, 1998). Bunlarla birlikte, simülasyon çalışmalarının özellikle zaman içinde geliştikçe altta yatan çoklu organizasyonel ve stratejik süreçler arasındaki etkileşimlerin sonuçlarını açıkça ortaya çıkarabildiğine dikkat çekilmiştir (Repenning, 2002). Bu bakış açılarından hareketle simülasyon çalışmalarının, mevcut teorinin keskin bir şekilde belirlenmesi ve genişletilmesi için güçlü bir yöntem olabileceği üzerinde durulmuştur.

Simülasyon çalışmalarının birçok avantajı olmasına rağmen bazı dezavantajları da vardır. Bu dezavantajlar, bir sistemin modellenmesi ve analizi ile doğrudan ilişkili değildir,

daha çok simülasyon çalışmalarıyla ilgili beklenen durumlarla ilişkilidir. Bu dezavantajlar aşağıda şu şekilde sıralanmıştır (Chung, 2004):

- Simülasyonun girdi verileri yanlış olduğunda doğru sonuçlar veremez.
- Simülasyon çalışmaları, karmaşık problemlere kolay cevaplar veremez.
- Simülasyon esnasında bir sorun meydana gelirse bu sorun kendi başına çözülemez.

Bahsedilen bu dezavantajlar çerçevesinde simülasyon çalışmalarında görülen diğer bir sınırlılık geçerlik problemidir. Simülasyonu geliştiren ve uygulayan kişinin, gerçek sistemin makul bir temsili olan bir modeli yaratmaya çalışması gerekmektedir. Ancak çeşitli nedenlerle, özenle oluşturulmuş modeller bile gerçeği tam olarak yansıtmayabilir. Bu durum, simülasyonu uygulayan kişinin modeldeki hataları elimine ettiğine ne kadar kanaat getirirse getirsin, modelin yine de başarılı olamayabileceğini gösterebilir (Chung, 2004). Ayrıca bilimsel yaklaşımlar, doğrulanma zorunluluğuyla diğer yöntemlerden ayrılmakta ve bu durum da akademide yayınlanan bulguların geçerliğinin incelenmesi için sonuçların çoğaltılması veya tahrif edilmeye çalışılması anlamına gelmektedir (Townsend & Johnson, 2008).

Çalışmalarda simülasyon yöntemi kullanılması ne kadar heyecan verici olsa da, bu durumun popülerleşmesi ciddi metodolojik sorunlara yol açmıştır. Bu bakımdan, simülasyon çalışmalarının ve bulgularının geçerli olarak tanımlanabileceği yolların keşfedilmesi gerekmektedir. Simülasyon çalışmaları, sosyal bilimcilere deney yapma olanağı sağladığı için çalışmalardaki öneminin gözden kaçırılmaması gerektiği vurgulanmıştır. Buradaki zorluk, simülasyon çalışmalarının geçerli nedensel çıkarımların çizilebileceği şekilde en iyi nasıl tasarlanacağı ve yorumlanacağıdır. Bu bağlamda simülasyon çalışmalarında geçerliliğe yönelik önemli tehditler oluşabileceği ve araştırmacıların bunları en aza indirmek için aktif önlemler almadığı durumlarda, simülasyon çalışmalarının sonuçlarının beklenen yönde gerçekleşmeyebileceği vurgulanmıştır. Ayrıca, geçerli sonuçların elde edilmesini

kolaylařtırmak adına replikasyon yapmanın gerekliliđi üzerinde durulmuřtur (Townslley & Johnson, 2008).

Yapılan bu alıřmada, simülasyon geerliđi iin ilk olarak girilen parametre deđerlerine göre üretilen madde parametre deđerlerinin verilen girdi deđerler aralıđında deđiřip deđiřmediđine, güven aralıđı deđerlerine bakılarak karar verilmiřtir. Yapılan alıřma sonucunda deđerlerin girdi deđer aralıklarında olduđu görülmüřtür. Buna yönelik tablolar EK-1 ve EK-2'de verilmiřtir.

Simülasyonun geerliđine iliřkin kanıt sunmak amacıyla ikinci olarak, kořullara göre yapılan kestirimlerin RMSE ve yanlılık deđerleri hesaplanmıřtır. Buna göre simülasyonlar sonrasında elde edilen 1-0 verilerinden kestirilen madde ve yetenek parametrelerinin bařlangı madde ve yetenek parametrelerine olan benzeřikliđi incelenmiřtir. Daha sonrasında, madde ve yetenek parametrelerinin RMSE ve yanlılık deđerlerinin her bir kořula göre nasıl deđiřtiđine bakılmıřtır. Kestirilen madde parametrelerine ait RMSE ve yanlılık deđerlerinin ortalama sonuçları EK-D1 ile EK-D2'de, kestirilen yetenek parametrelerine ait RMSE ve yanlılık deđerlerinin ortalama sonuçları ise EK-E1 ile EK-E2'de tablolar halinde verilmiřtir.

Bölüm 4

Bulgular, Yorumlar ve Tartışma

BOBUT uygulaması simülasyonlarından elde edilen sonuçlara bu bölümde değinilmiş, her bir alt probleme yönelik elde edilen bulgulara alt başlıklarda yer verilmiştir.

Birinci Alt Probleme İlişkin Bulgular ve Yorum

1. BOBUT uygulamasında, madde havuzu büyüklüğü 320 madde olan havuzdan elde edilen ölçme kesinliği kestirim değerleri; içerik ağırlıklandırma yapılmadığı ve yapıldığı durumda, madde seçme yöntemine (MFI, MIS, GEK), yetenek kestirim yöntemine (EAP, MLE) ve madde kullanım sıklığı kontrol yöntemine (KY, MUI) göre nasıl değişmektedir?

Çalışmada madde havuzu büyüklüğü 320 madde olan havuzdan elde edilen RMSE, yanlılık ve korelasyon değerlerinin; madde seçme, yetenek kestirim, madde kullanım sıklığı kontrol yöntemine ve içerik ağırlıklandırma yapılmama durumuna göre nasıl değiştiğini ve koşulların birbirlerinden nasıl farklılaştığını belirlemek için faktöriyel ANOVA yapılmıştır. Faktöriyel ANOVA analizi sonucu elde edilen bulgular Tablo 6'da verilmiştir.

Tablo 6 incelendiğinde; madde seçme yönteminin, RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2=0,071$) orta düzeyde olduğu görülmektedir. 0,071 değeri ile en küçük etkiye sahip faktör madde seçme yöntemi olmuştur. Farklı madde seçme yöntemi kullanılarak oluşturulan koşulların birbirinden nasıl farklılaştığını görmek için Bonferroni yöntemi kullanılarak Post-Hoc analizi yapılmıştır. Buna göre üç madde seçme yönteminin bulunduğu koşullardan elde edilen RMSE değerlerinin istatistiksel olarak anlamlı farklılaştığı tespit edilmiştir. Yetenek kestirim yöntemlerinin (EAP ve MLE), RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2=0,434$) yüksek düzeyde olduğu görülmektedir. 0,434 değeri ile en büyük etkiye sahip faktör yetenek kestirim yöntemi olmuştur. EAP yetenek kestirim yöntemi kullanılarak elde edilen RMSE değerlerinin, MLE yetenek kestirim yöntemi kullanılarak elde edilen RMSE değerlerinden anlamlı olarak daha

düşük olduğu bulunmuştur. Madde kullanım sıklığı kontrol yönteminin RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün de ($\eta^2=0,083$) orta düzeyde olduğu tablodan anlaşılmaktadır. İçerik ağırlıklandırma yönteminin ise, RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2=0,107$) orta düzeyde olduğu yine tablodan görülmektedir. Madde seçme ile yetenek kestirim yöntemi etkisinin ($\eta^2=0,047$), madde seçme ile madde kullanım sıklığı kontrol yöntemi etkisinin ($\eta^2=0,069$), yetenek kestirim ile madde kullanım sıklığı kontrol yöntemi etkisinin ($\eta^2=0,059$), yetenek kestirim yöntemi ile içerik ağırlıklandırma yapılma etkisinin ($\eta^2=0,043$) ve madde kullanım sıklığı kontrol yöntemi ile içerik ağırlıklandırma yapılma etkisinin ($\eta^2=0,017$) orta düzeyde olduğu tablodan anlaşılmaktadır. Dört faktörün birbirleriyle olan karşılıklı etkileşimleri incelendiğinde ise tüm etkileşimlerin etki büyüklüğünün orta düzeyde ($\eta^2=0,064$) olduğu görülmektedir.

Tablo 6

Madde Havuzu Büyüklüğü 320 Madde Olan Havuzdan Elde Edilen Ölçme Kesinliği

Kestirim Değerlerine Ait ANOVA Sonuçları

Simülasyon Koşulları	sd	RMSE		Yanlılık		Korelasyon	
		F	Eta Kare (η^2)	F	Eta Kare (η^2)	F	Eta Kare (η^2)
Madde Seçme Yöntemi (MSY)	2	141,315*	0,071	72,090*	0,038	116,756*	0,059
Yetenek Kestirim Yöntemi (YKY)	1	2850,003*	0,434	1725,499*	0,317	2,368	0,001
Madde Kullanım Sıklığı Kontrol Yöntemi (MKS)	1	335,235*	0,083	0,011	0,000	261,557*	0,066
İçerik Ağırlıklandırma (İA)	1	446,939*	0,107	0,060	0,000	314,382*	0,078
MSY x KYK	2	92,337*	0,047	1,832	0,001	0,450	0,000
MSY x MKS	2	138,003*	0,069	0,204	0,000	87,060*	0,045
MSY x İA	2	0,535	0,000	6,602*	0,004	5,226	0,003
YKY x MKS	1	232,329*	0,059	84,291*	0,022	42,538	0,011
YKY x İA	1	167,553*	0,043	128,043*	0,033	12,974*	0,003
MKS x İA	1	62,971*	0,017	3,582	0,001	74,541*	0,020
MSY x KYK x MKS	2	174,060*	0,086	124,841*	0,063	3,350	0,002
MSY x KYK x İA	2	125,291*	0,063	150,101*	0,075	8,129*	0,004
MSY x MKS x İA	2	14,518*	0,008	3,936*	0,002	21,086*	0,011
YKY x MKS x İA	1	140,671*	0,036	186,369*	0,048	8,700	0,002
MSY x KYK x MKS x İA	1	127,378*	0,064	148,866*	0,074	13,885*	0,007

$p < 0,05$

Yanlılık değerleri için ANOVA tablosu incelendiğinde, madde seçme yöntemi ve yetenek kestirim yöntemi faktörlerinin oluşturduğu farkın istatistiksel olarak anlamlı olduğu bulunurken, madde kullanım sıklığı kontrol yöntemi ve içerik ağırlıklandırmanın yanlılık üzerinde oluşturduğu farkların istatistiksel olarak anlamlı olmadığı görülmüştür. Bu dört faktörün karşılıklı olarak birbirleriyle olan etkileşimlerinin etkisine bakıldığında ise madde seçme yöntemi ile yetenek kestirim yöntemi etkileşimi, madde seçme yöntemi ile madde kullanım sıklığı kontrol yöntemi etkileşimi ve içerik ağırlıklandırma ile madde kullanım sıklığı kontrol yöntemi etkileşimi anlamlı bulunmazken diğer etkileşimlerin yanlılık değeri üzerinde oluşturduğu farkın istatistiksel olarak anlamlı olduğu bulunmuştur. Etki büyüklükleri incelendiğinde yanlılık değerlerinin farklılaşması üzerinde madde seçme yönteminin etki büyüklüğünün ($\eta^2=0,038$) orta düzeyde olduğu, yetenek kestirim yönteminin etki büyüklüğünün ($\eta^2=0,317$) yüksek düzeyde olduğu görülmektedir. Madde seçme ile içerik ağırlıklandırma yapıma etkisinin ($\eta^2=0,004$) düşük düzeyde, yetenek kestirim yöntemi ile madde kullanım sıklığı kontrol yöntemi etkisinin ($\eta^2=0,022$) ve yetenek kestirim yöntemi ile içerik ağırlıklandırma yapıma etkisinin ($\eta^2=0,033$) orta düzeyde olduğu tablodan anlaşılmaktadır. Dört faktörün birbirleriyle olan karşılıklı etkileşimleri incelendiğinde ise tüm etkileşimlerin etki büyüklüğünün orta düzeyde ($\eta^2=0,074$) olduğu görülmektedir.

Korelasyon değerleri için ANOVA tablosu incelendiğinde, madde seçme yöntemi, madde kullanım sıklığı kontrol yöntemi ve içerik ağırlıklandırma faktörlerinin oluşturduğu farkın istatistiksel olarak anlamlı olduğu bulunurken, yetenek kestirim yönteminin korelasyon üzerinde oluşturduğu farkların istatistiksel olarak anlamlı olmadığı görülmüştür. Bu dört faktörden karşılıklı olarak birbirleriyle olan ikili etkileşimlerinin etkisine bakıldığında ise; madde seçme yöntemi ile madde kullanım sıklığı kontrol yöntemi etkileşiminin, yetenek kestirim yöntemi ile içerik ağırlıklandırma etkileşiminin ve madde kullanım sıklığı kontrolü ile içerik ağırlıklandırma etkileşiminin etkisi anlamlı bulunmuştur. Dört faktörün birbirleriyle olan karşılıklı etkileşimleri incelendiğinde ise tüm etkileşimlerin etki büyüklüğünün çok düşük düzeyde olduğu tablodan anlaşılmaktadır. Etki büyüklükleri incelendiğinde

korelasyon değerlerinin farklılaşması üzerinde madde seçme yönteminin etki büyüklüğünün ($\eta^2=0,059$), madde kullanım sıklığı kontrol yönteminin etki büyüklüğünün ($\eta^2=0,066$) ve içerik ağırlıklandırma yapılmasının etki büyüklüğünün ($\eta^2=0,078$) orta düzeyde olduğu görülmektedir. Madde seçme ile madde kullanım sıklığı kontrol yöntemi etkisi ($\eta^2=0,045$) ve madde kullanım sıklığı kontrol yöntemi ile içerik ağırlıklandırma yapılma etkisinin ($\eta^2=0,020$) orta düzeyde olduğu, yetenek kestirim yöntemi ile içerik ağırlıklandırma yapılma etkisinin ($\eta^2=0,003$) düşük düzeyde olduğu tablodan anlaşılmaktadır. Dört faktörün birbirleriyle olan karşılıklı etkileşimleri incelendiğinde ise tüm etkileşimlerin etki büyüklüğünün düşük düzeyde ($\eta^2=0,007$) olduğu görülmektedir.

Birinci alt probleme ilişkin yapılan analizler sonucunda, içerik ağırlıklandırma yapılmadığı ve yapıldığı durumda; üç farklı madde seçme, iki farklı yetenek kestirim ve iki farklı madde kullanım sıklığı kontrol yöntemine göre belirlenen koşullardan elde edilen ölçme kesinliği kestirim değerleri Tablo 7'de verilmiştir. Ölçme kesinliği kestirim değerlerine ait grafikler ise sırasıyla Şekil 7, Şekil 8 ve Şekil 9'da sunulmuştur.

0,261 ile 0,281 arasında değişen RMSE değerlerinin en düşük olduğu koşulun, içerik ağırlıklandırmanın yapılmadığı; madde seçme yönteminin MFI, yetenek kestirim yönteminin EAP ve madde kullanım sıklığı kontrol yönteminin MUİ koşulu olduğu görülmüştür. En yüksek RMSE değerinin elde edildiği koşulun ise içerik ağırlıklandırmanın yapıldığı; madde seçme yönteminin GEK ve yetenek kestirim yönteminin MLE koşulu olduğu görülmüştür. Madde havuzu büyüklüğünün 320 madde olduğu tüm koşullar altında; içerik ağırlıklandırma yapılmadığı ve yapıldığı durumda, yetenek kestirim EAP, madde kullanım sıklığı kontrol yönteminin KY olduğu ile içerik ağırlıklandırma yapılmadığı durumda, yetenek kestirim MLE, madde kullanım sıklığı kontrol yönteminin KY olduğu koşullar hariç diğer tüm koşullarda GEK yönteminin kullanılmasının ortalama hata değerini artırdığı görülmektedir. Kullanılan madde seçme yöntemleri arasında, bahsedilen koşullar hariç en yüksek RMSE değerleri GEK olduğunda hesaplanmıştır. Madde seçme yönteminin MIS olduğu sekiz koşulda ise dört koşulda RMSE değerinin MFI yöntemine göre azaldığı görülürken, diğer dört koşulda

arttıđı gözlemlenmiřtir. Madde kullanım sıklıđı kontrol yöntemlerine bakılacak olursa, genel olarak KY'in kullanıldıđı kořullara göre MUİ yönteminin kullanıldıđı kořullarda daha düşük RMSE deđerleri elde edilmiřtir.

Tablo 7

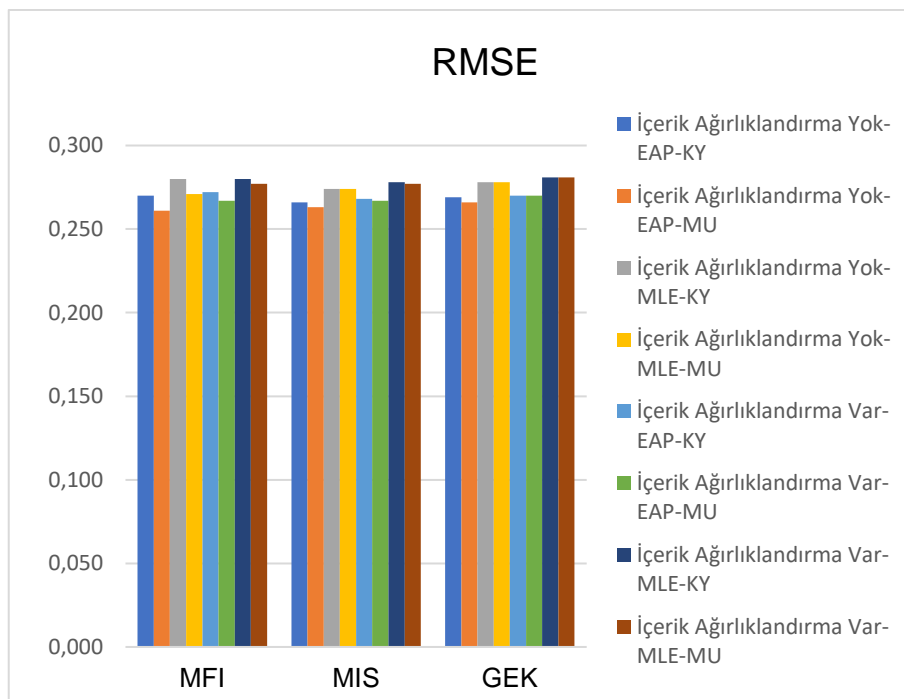
Madde Havuzu Büyüklüğü 320 Madde Olan Havuzdan Elde Edilen Ölçme Kesinliği Kestirim Değerleri

Madde Seçme Yöntemi	Yetenek Kestirim Yöntemi	Madde Kullanım Sıklığı Kontrol Yöntemi	İçerik Ağırlıklandırma Yok			İçerik Ağırlıklandırma Var		
			RMSE	Yanlılık	Korelasyon	RMSE	Yanlılık	Korelasyon
MFI	EAP	Kısıtlayıcı Yöntem	0,270	0,000	0,963	0,272	-0,001	0,962
		Madde Uygunluk İndeksi	0,261	0,001	0,965	0,267	-0,001	0,964
	MLE	Kısıtlayıcı Yöntem	0,280	0,009	0,963	0,280	0,008	0,964
		Madde Uygunluk İndeksi	0,271	0,008	0,965	0,277	0,008	0,964
MIS	EAP	Kısıtlayıcı Yöntem	0,266	0,000	0,964	0,268	0,000	0,963
		Madde Uygunluk İndeksi	0,263	-0,001	0,965	0,267	0,001	0,964
	MLE	Kısıtlayıcı Yöntem	0,274	0,009	0,964	0,278	0,008	0,963
		Madde Uygunluk İndeksi	0,274	0,007	0,964	0,277	0,009	0,964
GEK	EAP	Kısıtlayıcı Yöntem	0,269	-0,001	0,963	0,270	-0,001	0,963
		Madde Uygunluk İndeksi	0,266	0,000	0,964	0,270	0,000	0,963
	MLE	Kısıtlayıcı Yöntem	0,278	0,015	0,963	0,281	0,013	0,963
		Madde Uygunluk İndeksi	0,278	0,014	0,963	0,281	0,013	0,963

Yetenek kestirim yönteminin EAP olduğu durumda içerik ağırlıklandırmanın hem yapılmadığı hem de yapıldığı tüm koşullarda ortalama hata değeri daha düşük olarak hesaplanmıştır. Sekiz koşul altında da madde seçme yöntemi MIS olduğu durumdan madde seçme yöntemi GEK yöntemine geçildiği durumdaki RMSE değerlerine ait grafikler benzerlik göstermektedir (Şekil 7). İçerik ağırlıklandırma yapılmadığı durumda; MIS madde seçme, MLE yetenek kestirim, madde kullanım sıklığı kontrol yöntemi olarak ise hem KY hem de MUİ yönteminin kullanıldığı koşul ve GEK madde seçme, MLE yetenek kestirim, madde kullanım sıklığı kontrol yöntemi olarak yine hem KY hem de MUİ yönteminin kullanıldığı koşullardan elde edilen RMSE değerleri aynıdır. Bu sebeple bu koşullara ait grafikler binişiktir. Bu durum, içerik ağırlıklandırma yapılmadığı ve yetenek kestirim yönteminin MLE olduğu koşullarda, madde seçme yöntemlerinden MIS ile GEK yönteminden elde edilen RMSE değerlerinin madde kullanım sıklığı yöntemine göre değişmediğini göstermektedir. Diğer bir yandan tabloya göre içerik ağırlıklandırma yapılmasının genel olarak tüm koşullarda RMSE değerini artırdığı görülmektedir.

Şekil 7

Madde Havuzu Büyüklüğü 320 Madde Olan Havuza Ait RMSE Değerleri

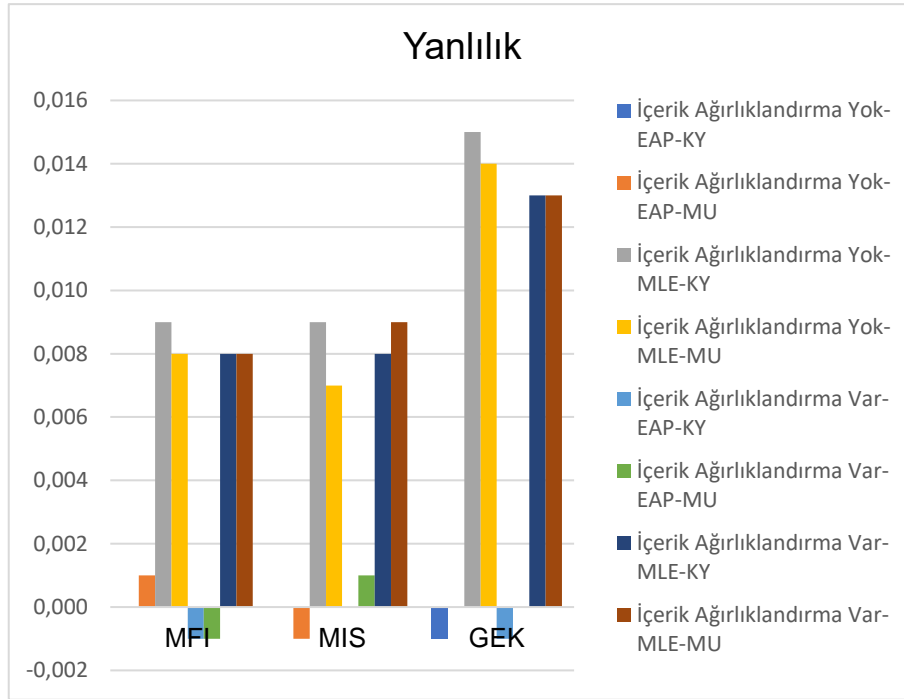


Tablo 7’de yer alan ortalama yanlılık deęerleri incelendięinde genel olarak tm koşullar için yanlılık deęerlerinin dşk olduęu ve -0,001 ile 0,015 arasında deęiştiięi grlmektedir. 0,000 yanlılık deęerinin, ięerik aęırlıklandırma yapılmadıęı durumda; madde seęme ynteminin MFI, yetenek kestirim ynteminin EAP, madde kullanım sıklıęı kontrol ynteminin KY olduęu koşul ile ięerik aęırlıklandırmanın hem yapılmadıęı hem de yapıldıęı durumda; madde seęme ynteminin MIS, yetenek kestirim ynteminin EAP, madde kullanım sıklıęı kontrol ynteminin KY olduęu koşul ve ięerik aęırlıklandırmanın hem yapıldıęı hem de yapılmadıęı durumda; madde seęme ynteminin GEK, yetenek kestirim ynteminin EAP, madde kullanım sıklıęı kontrol ynteminin MUİ olduęu koşullardan elde edildięi grlmştr. Buna gre bu koşullar için yansız hesaplamalar yapıldıęı sylenebilir. Yanlılık deęerinin sifıra en uzak olduęu koşul; ięerik aęırlıklandırmanın yapılmadıęı, madde seęme ynteminin GEK, yetenek kestirim ynteminin MLE ve madde kullanım sıklıęı kontrol yntemi olarak KY’nin kullanıldıęı koşul olmuştur. En yksek yanlılık deęerine sahip koşulların GEK madde seęme yntemine ait koşullarda, yetenek kestirim yntemi olarak MLE ynteminin kullanıldıęı koşulların olduęu gzlemlenmiştir. Bu baęlamda yetenek kestirim yntemi olarak MLE yntemi kullanılmasının yanlılık deęerini artırdıęı grlmştr. İęerik aęırlıklandırma yapıldıęı ve yapılmadıęı durumlarda koşullara ait yanlılık deęerlerinin bu durumdan etkilenmedięi sonucuna varılmıştır. Ayrıca madde kullanım sıklıęı kontrol yntemlerinden KY ile MUİ ynteminin yanlılık zerinde herhangi bir etkisinin olmadıęı gzlemlenen bulgular arasındadır.

Şekil 8’deki grafik incelendięinde yetenek kestirim yntemi olarak MLE ynteminin kullanıldıęı koşullarda grafiklerin benzer yapıda ve MLE ynteminin kullanıldıęı drt farklı koşulun ortalama yanlılık deęerlerinin birbirine ęok yakın olduęu grlmektedir. Ancak yetenek kestirimi yntemi olarak EAP ynteminin kullanıldıęı koşullardan elde edilen deęerler için benzer bir rnt gze ęarpmamıştır.

Şekil 8

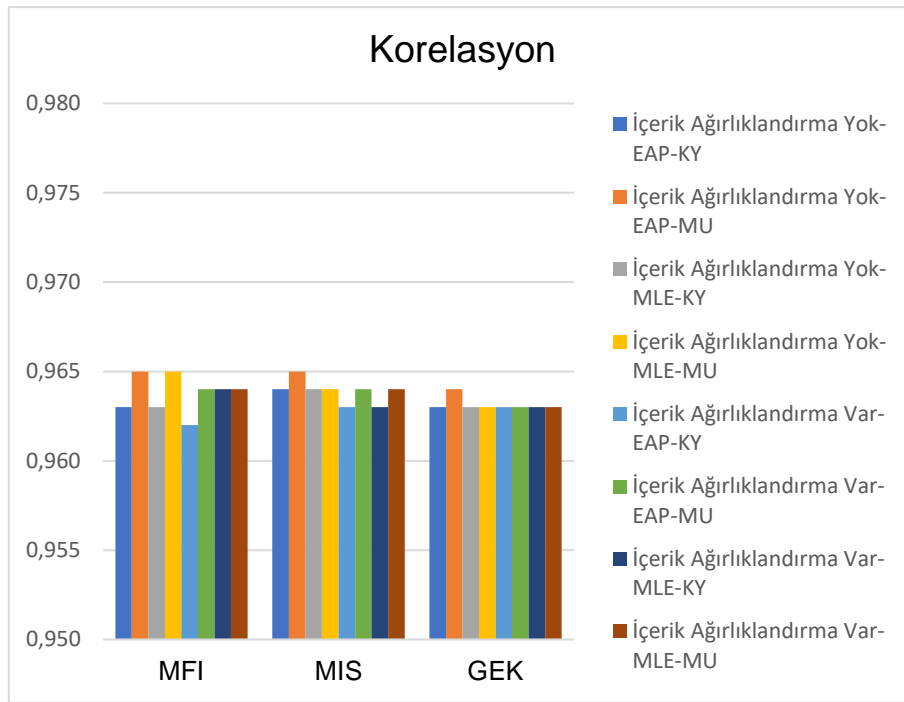
Madde Havuzu Büyüklüğü 320 Madde Olan Havuza Ait Yanlılık Değerleri



Son olarak Tablo 7'de yer alan korelasyon değerleri incelendiğinde en yüksek değere; içerik ağırlıklandırmanın yapılmadığı durumda madde seçme yönteminin MFI, yetenek kestirim yönteminin EAP ve madde kullanım sıklığı kontrol yöntemi olarak da MUİ yönteminin olduğu koşullarda ulaşılmıştır. Aynı değer, içerik ağırlıklandırmanın yapılmadığı durumda madde seçme yönteminin MIS, yetenek kestirim yönteminin EAP ve madde kullanım sıklığı kontrol yönteminin de MUİ yönteminin olduğu koşulda da hesaplanmıştır. En düşük korelasyon değerinin ise; içerik ağırlıklandırma yapıldığı durumda, madde seçme yönteminin MFI, yetenek kestirim yönteminin EAP ve madde kullanım sıklığı kontrol yönteminin KY olduğu koşulda elde edildiği tablodan anlaşılmaktadır. Ayrıca madde kullanım sıklığı kontrol yöntemlerinden KY ile MUİ yönteminin korelasyon üzerinde herhangi bir etkisinin olup olmadığına bakılacak olursa, MUİ yönteminin KY'ye göre daha yüksek değerler verdiği gözlemlenen bulgular arasındadır.

Şekil 9

Madde Havuzu Büyüklüğü 320 Madde Olan Havuza Ait Korelasyon Değerleri



İkinci Alt Probleme İlişkin Bulgular ve Yorum

2. BOBUT uygulamasında, madde havuzu büyüklüğü 1000 madde olan havuzdan elde edilen ölçme kesinliği kestirim değerleri; içerik ağırlıklandırma yapılmadığı ve yapıldığı durumda, madde seçme yöntemine (MFI, MIS, GEK), yetenek kestirim yöntemine (EAP, MLE) ve madde kullanım sıklığı kontrol yöntemine (KY, MU) göre nasıl değişmektedir?

Çalışmada madde havuzu büyüklüğü 1000 madde olan havuzdan elde edilen RMSE, yanlılık ve korelasyon değerlerinin; madde seçme, yetenek kestirim, madde kullanım sıklığı kontrol yöntemine ve içerik ağırlıklandırma yapıma durumuna göre nasıl değiştiğini ve koşulların birbirlerinden nasıl farklılaştığını belirlemek için ANOVA yapılmıştır. ANOVA'dan elde edilen bulgular Tablo 8'de verilmiştir.

Tablo 8 incelendiğinde madde seçme yönteminin, RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2=0,967$) yüksek düzeyde olduğu görülmektedir. 0,967

değeri ile en büyük etkiye sahip faktör madde seçme yöntemi olmuştur. Farklı madde seçme yönteminin kullanıldığı koşulların birbirinden nasıl farklılaştığını görmek için Bonferroni yöntemi kullanılarak Post-Hoc analizi yapılmıştır. Buna göre MFI yöntemi ile edilen RMSE değerleri MIS ve GEK yöntemleri ile elde edilen RMSE değerlerinden anlamlı derecede düşüktür. Yetenek kestirim yöntemleri olan EAP ve MLE'nin, RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2=0,809$) yüksek düzeyde olduğu görülmektedir. EAP yetenek kestirim yöntemi kullanılarak elde edilen RMSE değerleri, MLE yöntemi kullanılarak elde edilen RMSE değerlerinden anlamlı olarak daha düşük olduğu bulunmuştur. Madde kullanım sıklığı kontrol yönteminin RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün de ($\eta^2=0,000$) önemsiz düzeyde olduğu görülürken, içerik ağırlıklandırma yönteminin RMSE değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2=0,041$) orta düzeyde olduğu görülmektedir. Dört faktörün birbiri ile karşılıklı etkileşimlerinden doğan farkların etki büyüklükleri incelendiğinde ise, yetenek kestirim yöntemi ile madde seçme yöntemi etkileşiminin ($\eta^2=0,312$) yüksek düzeyde etkiye ve madde seçme yöntemi ile içerik ağırlıklandırma etkileşiminin ($\eta^2=0,047$) de orta düzeyde etkiye sahip olduğu belirlenmiştir. Diğer etkileşimlerin ise istatistiksel olarak anlamlı fark oluşturmadığı tablodan anlaşılmaktadır.

Yanlılık değerleri için ANOVA tablosu incelendiğinde, madde seçme yöntemi ve yetenek kestirim yöntemi faktörlerinin oluşturduğu farkın istatistiksel olarak anlamlı olduğu bulunurken, madde kullanım sıklığı kontrol yöntemi ve içerik ağırlıklandırmanın yanlılık üzerinde oluşturduğu farkların istatistiksel olarak anlamlı olmadığı görülmüştür. Bu dört faktörün karşılıklı olarak birbirleriyle olan etkileşimlerinin etkisine bakıldığında ise madde seçme yöntemi ve yetenek kestirim yöntemi etkileşimi anlamlı bulunurken diğer etkileşimlerin yanlılık değeri üzerinde oluşturduğu farkın istatistiksel olarak anlamlı olmadığı belirlenmiştir. Etki büyüklükleri incelendiğinde madde seçme yönteminin, yanlılık değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2=0,088$) orta düzeyde olduğu, yetenek kestirim yönteminin etki büyüklüğünün ($\eta^2=0,406$) yüksek düzeyde olduğu

görülmektedir. Madde seçme ve yetenek kestirim yöntemlerinin etkileşiminden doğan farkın etki büyüklüğünün ($\eta^2=0,101$) orta düzeyde olduğu tablodan anlaşılmaktadır.

Tablo 8

Madde Havuzu Büyüklüğü 1000 Madde Olan Havuzdan Elde Edilen Ölçme Kesinliği Kestirim Değerlerine Ait ANOVA Sonuçları

Simülasyon Koşulları	sd	RMSE		Yanlılık		Korelasyon	
		F	Eta Kare (η^2)	F	Eta Kare (η^2)	F	Eta Kare (η^2)
Madde Seçme Yöntemi (MSY)	2	53889,320*	0,967	180,420*	0,088	42259,250*	0,958
Yetenek Kestirim Yöntemi (YKY)	1	15748,240*	0,809	2537,487*	0,406	4000,739*	0,518
Madde Kullanım Sıklığı Kontrol Yöntemi (MKS)	1	0,035	0,000	0,011	0,000	2,780	0,000
İçerik Ağırlıklandırma (İA)	1	159,377*	0,041	1,444	0,000	113,510*	0,030
MSY x YKY	2	842,858*	0,312	208,090*	0,101	391,733*	0,174
MSY x MKS	2	4,054	0,002	0,907	0,000	2,298	0,001
MSY x İA	2	91,100*	0,047	0,647	0,000	51,548*	0,027
YKY x MKS	1	0,521	0,000	0,045	0,000	0,625	0,000
YKY x İA	1	0,015	0,000	0,021	0,000	0,001	0,000
MKS x İA	1	0,689	0,000	0,654	0,000	0,343	0,000
MSY x YKY x MKS	2	0,355	0,000	1,228	0,001	0,058	0,000
MSY x YKY x İA	2	0,084	0,000	1,329	0,001	0,244	0,000
MSY x MKS x İA	2	0,564	0,000	2,862	0,002	0,126	0,000
YKY x MKS x İA	1	0,810	0,000	0,073	0,000	2,406	0,001
MSY x YKY x MKS x İA	2	0,763	0,000	1,669	0,001	0,798	0,000

$p < 0,05$

Korelasyon değerleri için ANOVA tablosu incelendiğinde, madde seçme yöntemi, yetenek kestirim yöntemi ve içerik ağırlıklandırma faktörlerinin oluşturduğu farkın istatistiksel olarak anlamlı olduğu bulunurken, madde kullanım sıklığı kontrol yönteminin korelasyon üzerinde oluşturduğu farkların istatistiksel olarak anlamlı olmadığı görülmüştür. Bu dört faktörün karşılıklı olarak birbirleriyle olan etkileşimlerinin etkisine bakıldığında ise madde seçme yöntemi ve yetenek kestirim yöntemi etkileşimi ile madde seçme yöntemi ve içerik ağırlıklandırma etkileşim etkisinin anlamlı olduğu diğer etkileşimlerin anlamlı olmadığı belirlenmiştir. Etki büyüklükleri incelendiğinde madde seçme yönteminin, korelasyon değerlerinin farklılaşması üzerindeki etki büyüklüğünün ($\eta^2=0,958$) yüksek düzeyde olduğu,

yetenek kestirim yönteminin etki büyüklüğünün ($\eta^2=0,518$) yüksek düzeyde olduğu görülmektedir. Madde seçme ve yetenek kestirim yöntemlerinin etkileşiminden doğan farkın etki büyüklüğünün ($\eta^2=0,174$) orta düzeyde olduğu ve madde seçme yöntemi ve içerik ağırlıklandırma yapıma etkileşiminden doğan farkın etki büyüklüğünün ise ($\eta^2=0,027$) orta düzeyde olduğu tablodan anlaşılmaktadır.

İkinci alt probleme ilişkin yapılan analizler sonucunda, içerik ağırlıklandırma yapılmadığı ve yapıldığı durumda, üç farklı madde seçme, iki farklı yetenek kestirim ve iki farklı madde kullanım sıklığı kontrol yöntemine göre belirlenen koşullardan elde edilen ölçme kesinliği kestirim değerleri Tablo 9'da verilmiştir. Ölçme kesinliği kestirim değerlerine ait grafikler ise sırasıyla Şekil 10, Şekil 11 ve Şekil 12'de sunulmuştur.

0,212 ile 0,279 arasında değişen RMSE değerlerinin en düşük olduğu koşulun; içerik ağırlıklandırmanın yapılmadığı, madde seçme yönteminin MFI, yetenek kestirim yönteminin EAP, madde kullanım sıklığı kontrol yöntemi olarak ise her iki yöntemin kullanıldığı koşul olduğu görülmüştür. En yüksek RMSE değerinin elde edildiği koşulun ise; içerik ağırlıklandırmanın yapıldığı, madde seçme yönteminin GEK, yetenek kestirim yönteminin MLE, madde kullanım sıklığı kontrol yönteminin MUİ yönteminin kullanıldığı koşul olduğu görülmüştür. Madde havuzu büyüklüğünün 1000 madde olduğu tüm koşullarda GEK yönteminin kullanılmasının ortalama hata değerini artırdığı görülmektedir. Kullanılan madde seçme yöntemleri arasında, sekiz koşul altında da en yüksek RMSE değeri GEK olduğunda hesaplanmıştır. Madde seçme yönteminin MIS olduğu sekiz koşulda da RMSE değerinin MFI yöntemine göre arttığı gözlemlenmiştir. Diğer bir yandan tabloya göre içerik ağırlıklandırma yapılmasının genel olarak tüm koşullarda RMSE değerini artırdığı görülmektedir.

Tablo 9

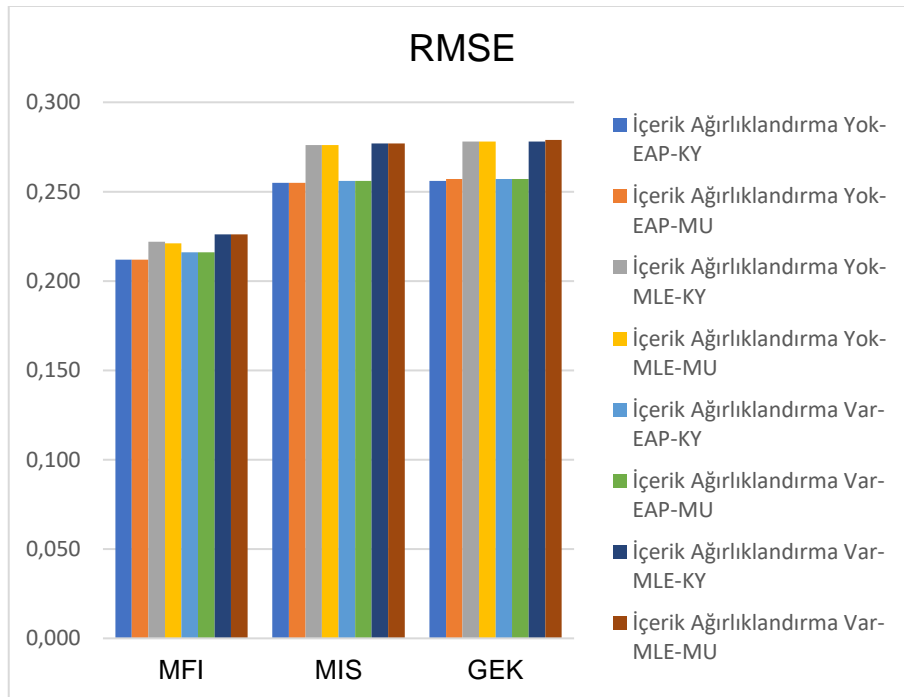
Madde Havuzu Büyüklüğü 1000 Madde Olan Havuzdan Elde Edilen Ölçme Kesinliği Kestirim Değerleri

Madde Seçme Yöntemi	Yetenek Kestirim Yöntemi	Madde Kullanım Sıklığı Kontrol Yöntemi	İçerik Ağırlıklandırma Yok			İçerik Ağırlıklandırma Var		
			RMSE	Yanlılık	Korelasyon	RMSE	Yanlılık	Korelasyon
MFI	EAP	Kısıtlayıcı Yöntem	0,212	0,000	0,977	0,216	0,000	0,976
		Madde Uygunluk İndeksi	0,212	0,000	0,977	0,216	0,000	0,976
	MLE	Kısıtlayıcı Yöntem	0,222	0,006	0,976	0,226	0,006	0,975
		Madde Uygunluk İndeksi	0,221	0,004	0,976	0,226	0,006	0,975
MIS	EAP	Kısıtlayıcı Yöntem	0,255	0,000	0,967	0,256	0,000	0,967
		Madde Uygunluk İndeksi	0,255	0,000	0,967	0,256	0,000	0,967
	MLE	Kısıtlayıcı Yöntem	0,276	0,008	0,964	0,277	0,007	0,964
		Madde Uygunluk İndeksi	0,276	0,009	0,964	0,277	0,008	0,964
GEK	EAP	Kısıtlayıcı Yöntem	0,256	0,000	0,967	0,257	0,000	0,966
		Madde Uygunluk İndeksi	0,257	0,000	0,966	0,257	-0,001	0,966
	MLE	Kısıtlayıcı Yöntem	0,278	0,013	0,963	0,278	0,014	0,963
		Madde Uygunluk İndeksi	0,278	0,015	0,963	0,279	0,013	0,963

Yetenek kestirim yönteminin EAP olduğu durumda hem içerik ağırlıklandırmanın yapılmadığı hem de yapıldığı tüm koşullarda ortalama hata değeri daha düşük olarak hesaplanmıştır. Sekiz koşul altında da üç farklı madde seçme yöntemine ait RMSE değerlerine ait grafikler benzerlik göstermektedir (Şekil 10). Yetenek kestirim yönteminin EAP ve MLE olduğu koşullarda, madde kullanım sıklığı kontrol yöntemlerinin hem KY hem de MU olduğu, içerik ağırlıklandırma yapıldığı ve yapılmadığı durumlardan elde edilen grafikler binişiktir. Bu durum, 1000 maddelik madde havuzu ile gerçekleştirilen BOBUT uygulamasında EAP ve MLE yetenek kestirim yöntemleri koşullarında içerik ağırlıklandırma yapılmadığı ve yapıldığı durumda elde edilen ortalama hata değerlerinin birbirine çok yakın değerler olduğunu göstermektedir.

Şekil 10

Madde Havuzu Büyüklüğü 1000 Madde Olan Havuza Ait RMSE Değerleri



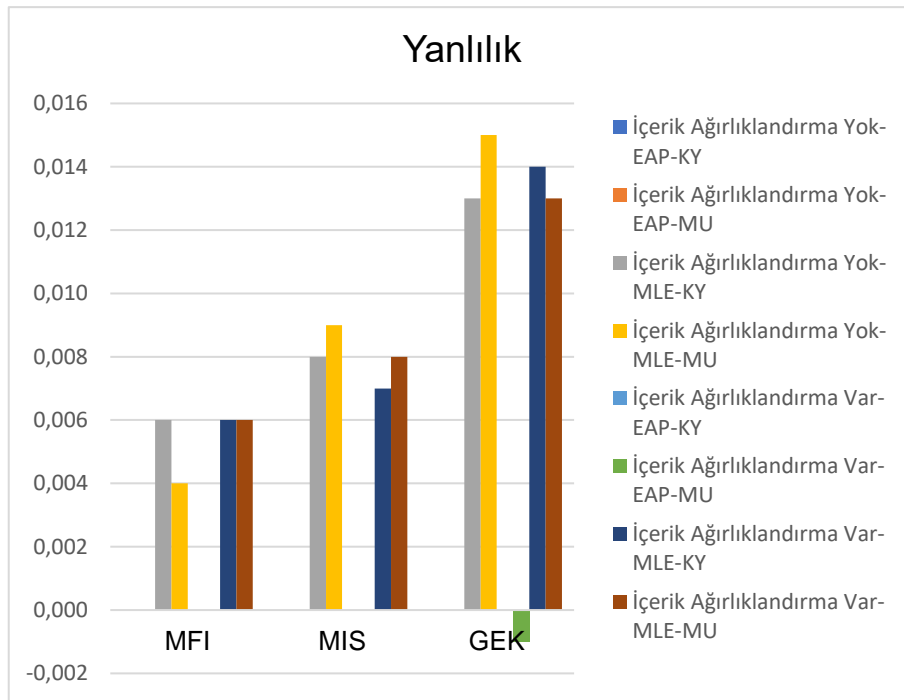
Tablo 9'da yer alan ortalama yanlılık değerleri incelendiğinde genel olarak tüm koşullar için yanlılık değerlerinin düşük olduğu ve -0,001 ile 0,015 arasında değiştiği görülmektedir. 0,000 yanlılık değerinin içerik ağırlıklandırmanın yapıldığı durumda; madde seçme yönteminin MIS, yetenek kestirim yönteminin EAP, MU madde kullanım sıklığı

kontrol yönteminin kullanıldığı koşul hariç diğer tüm EAP koşullarında hesaplandığı görülmüştür. Buna göre bu koşullar için yansız hesaplamalar yapıldığı söylenebilir. Yanlılık değerinin sıfıra en uzak olduğu koşul ise içerik ağırlıklandırmanın yapılmadığı, madde seçme yönteminin GEK, MLE yetenek kestirim ve madde kullanım sıklığı kontrol yöntemi olarak MUİ yönteminin kullanıldığı koşul olmuştur. En yüksek yanlılık değerine sahip koşulların GEK madde seçme yöntemine ait koşullarda, yetenek kestirim yöntemi olarak MLE yönteminin kullanıldığı koşulların olduğu gözlemlenmiştir. Ayrıca diğer madde seçme yöntemlerinin olduğu koşullarda da MLE yöntemi EAP yöntemine göre daha yüksek yanlılık değerleri ile sonuçlanmıştır. Bu bağlamda yetenek kestirim yöntemi olarak MLE yöntemi kullanılmasının yanlılık değerini artırdığı görülmüştür. İçerik ağırlıklandırma yapıldığı ve yapılmadığı durumlarda koşullara ait yanlılık değerlerinin bu durumdan etkilenmediği sonucuna varılmıştır.

Şekil 11'deki grafik incelendiğinde yetenek kestirim yöntemi olarak MLE yönteminin kullanıldığı koşullarda grafiklerin benzer yapıda ve MLE yönteminin kullanıldığı dört farklı koşulun ortalama yanlılık değerlerinin birbirine çok yakın olduğu görülmektedir. Hatta üç farklı madde seçme yönteminin dâhil edildiği koşullarda; içerik ağırlıklandırma yapılmadığı, MLE yetenek kestirim ve KY madde kullanım sıklığı kontrol yöntemlerinin kullanıldığı koşul ile içerik ağırlıklandırmanın yapıldığı, MLE yetenek kestirim ve MUİ madde kullanım sıklığı kontrol yöntemlerinin kullanıldığı koşuldan elde edilen yanlılık değerleri aynıdır. Bu yüzden bu koşulların grafikleri binişiktir. Aynı şekilde yetenek kestirim yöntemi olarak EAP olduğu tüm koşulların da grafiklerinin binişik olduğu Şekil 11'den anlaşılmaktadır.

Şekil 11

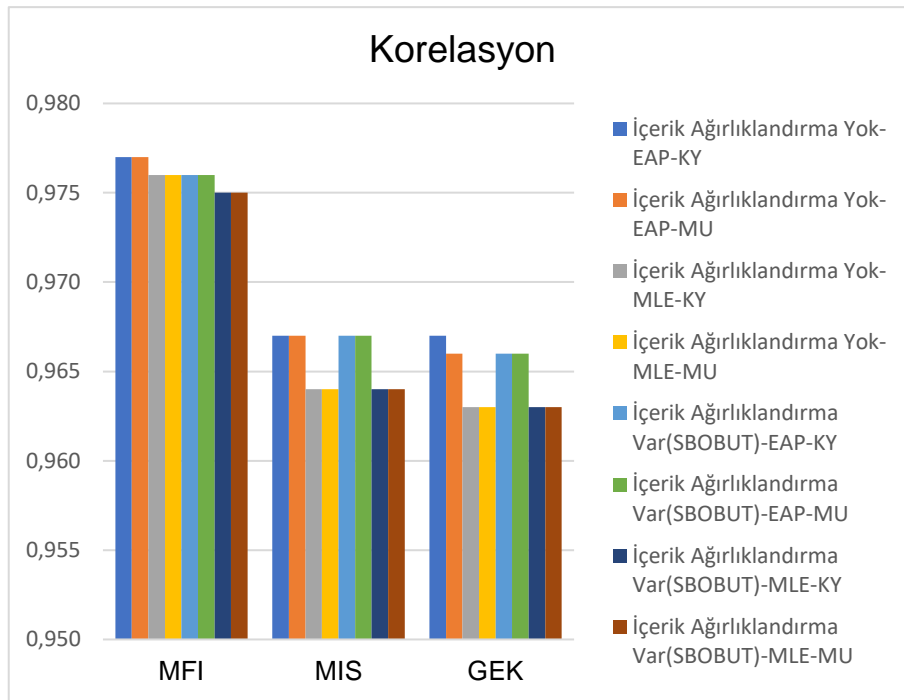
Madde Havuzu Büyüklüğü 1000 Madde Olan Havuza Ait Yanlılık Değerleri



Son olarak Tablo 9'da yer alan korelasyon değerleri incelendiğinde ise en yüksek değere; içerik ağırlıklandırmanın yapılmadığı durumda madde seçme yönteminin MFI, yetenek kestirim yönteminin EAP ve her iki madde kullanım sıklığı kontrol yönteminin olduğu koşullarda ulaşılmıştır. En düşük korelasyon değerinin ise; içerik ağırlıklandırma yapıldığı durumda, madde seçme yönteminin GEK, yetenek kestirim yönteminin MLE ve her iki madde kullanım sıklığı kontrol yönteminin olduğu koşullardan elde edildiği tablodan anlaşılmaktadır. Yetenek kestirim yönteminin EAP ve MLE olduğu koşullarda, madde kullanım sıklığı kontrol yöntemlerinin hem KY hem de MU olduğu, içerik ağırlıklandırma yapıldığı ve yapılmadığı durumlardan elde edilen grafikler binişiktir. Bu durum, 1000 maddelik madde havuzu ile gerçekleştirilen BOBUT uygulamasında EAP ve MLE yetenek kestirim yöntemleri koşullarında içerik ağırlıklandırma yapılmadığı ve yapıldığı durumda elde edilen korelasyon değerlerinin birbirine çok yakın değerler olduğunu göstermektedir.

Şekil 12

Madde Havuzu Büyüklüğü 1000 Madde Olan Havuza Ait Korelasyon Değerleri



Tartışma

Bu çalışmada, BOBUT uygulamalarında ölçme kesinliği kestirim değerlerinin; farklı madde seçme, farklı yetenek kestirim yöntemlerine göre içerik ağırlıklandırma ve madde kullanım sıklığı kontrolü gibi bazı madde havuzu sınırlandırmaları altında nasıl değiştiğini incelemeyi amaçlanmıştır. Bu incelemenin önemli olduğu düşünülmektedir çünkü içerik ağırlıklandırma ve madde havuzunun farklı içerik alanlarına bölünmesi, puanlara dayalı yapılan değerlendirmelerin geçerliği için bir gerekliliktir. Gerçek uygulamalarda, madde havuzları genellikle birden çok farklı içerik alanlarından maddelerden oluşmaktadır ve BOBUT uygulamalarında içerik kontrolü ve düzenlemelerinin yapılması ile uğraşmak kolay olmayabilir (Wise & Kingsbury, 2015). Diğer bir yandan, madde havuzlarının farklı içerik alanlarına bölünerek yapılan içerik ağırlıklandırma sınırlandırmasının BOBUT uygulamalarına dâhil edilmemesi, testin uygulamaya katılan bireylerde oluşturması gereken adalet algısı ve puanlara dayalı yapılan değerlendirmeler üzerinde olası olumsuz etkilere sahip olabilir. Bu nedenden dolayı, özellikle sonuçları öğrencinin geleceği ile ilgili belirleyici

olacak sınavlarda içerik ağırlıklandırma yapılmasının gerekliliği göz önünde bulundurulması gereken durumlardan biri olabilir.

Bu çalışmadan elde edilen temel bulgu, madde havuzunda pratik sınırlandırmalardan olan içerik ağırlıklandırma yapılması durumunun, ölçme kesinliği kestirim değerlerinden RMSE değerlerinde az da olsa bir artış meydana gelmesine neden olmasıdır. Çalışmanın bu bulgusu Kingsbury ve Zara (1991), Newman (1995), Luecht ve Nunchester (1998), Chen (1999), Cheng ve Chang (2009), Cheng vd. (2009), Song (2010) ve Zheng vd. (2013)'nin çalışmalarından elde edilen bulgularla paralellik göstermektedir. Sözü edilen çalışmalar kapsamında, farklı koşullar içeren BOBUT uygulamalarında içerik ağırlıklandırma yapılmasının ölçme kesinliği kestirim değerleri üzerindeki etkisi incelenmiştir. Buna göre, madde havuzunda içerik ağırlıklandırma yöntemi kullanılmasının ölçme kesinliği üzerinde olumsuz bir etkiye sebep olmadığına yanısıra sadece çok az da olsa hata değerinde bir miktar artış meydana getirdiği sözü edilen araştırmacılar tarafından belirtilmiştir. Hata değerinde meydana gelen bu artışın, göze alınmaya değer olup olmadığı testin amacı ile ilişkilidir. Özellikle öğrencinin geleceği ile ilgili belirleyici sonuçları olan sınavlara ait içerik geçerliliği ve her bireye aynı içerik dağılımından soru gelmesi gibi kritik öneme sahip olan durumlarda hata değerindeki bu artış göz ardı edilebilir. Madde havuzunun farklı içerik alanlarına ayrılarak içerik ağırlıklandırma yapılması, ölçme kesinliği değerleri bir değiş tokuş durumu olarak ifade edilebilir. Madde havuzunda pratik sınırlandırmalardan olan içerik ağırlıklandırmanın diğer ölçme kesinliği kestirim değerleri üzerinde bir etkisinin görülmediği bahsi geçen çalışmalarda elde edilen bulgular arasındadır.

Çalışmada elde edilen bir diğer bulgu ise, çaprazlanan simülasyon koşulları gözden geçirildiğinde, BOBUT uygulamalarında madde havuzunda içerik ağırlıklandırma yapıldığı zaman, madde seçme yöntemleri arasında en iyi performans gösteren yöntemin MFI yöntemi olduğu görülmüştür. Çalışmanın bu bulgusu Chen (1999), Cheng ve Chang (2009), Zheng vd. (2013) ve Şahin (2017)'nin çalışmalarından elde edilen bulgularla paralellik

göstermektedir. Chen (1999) çalışmasında, madde seçme yöntemi olarak içerik ağırlıklandırma ve madde kullanım sıklığı kontrolünün dâhil edildiği madde bilgi fonksiyonu koşulunda MFI madde seçme yönteminin, sonsal dağılımlı MFI ve sonsal dağılımlı Kullback-Leibler yöntemlerine göre daha iyi sonuç verdiğini belirtmiştir. Zheng (2013), yaptığı çalışmada SF-36 anketinin madde seçme yöntemi olarak MFI yönteminin kullanıldığı koşullarda elde edilen RMSE değerlerinin, diğer koşullardan elde edilen RMSE değerlerine göre daha düşük olduğunu vurgulamıştır. Çaprazlanan simülasyon koşullarında içerik ağırlıklandırma yapılmadığı durumlarda da yine madde seçme yöntemleri arasında en iyi performans gösteren yöntemin MFI yöntemi olduğu görülmüş olup, bu bulgu Cheng vd. (2009)'nin çalışmasından elde edilen bulgularla benzerlik göstermektedir. Cheng vd. (2009), a-tabakalama yönteminde madde seçme sürecini ağırlıklandırma mekanizmasının sınırlandırmaları başarılı bir şekilde ele aldığını, tabakalamanın madde kullanım sıklığı oranlarını dengelemeye büyük ölçüde yardımcı olduğunu ve kullanılan beş deneysel koşulun arasında artan a-tasarımının ölçme kesinliğini iyileştirdiğini vurgulamışlardır. Ayrıca kullanılan madde seçme yöntemleri arasında en iyi performansı içerik ağırlıklandırma yapılmayan koşullarda MFI yönteminin verdiğini de belirtmişlerdir. Song (2010) çalışmasında, MFI yönteminin daha düşük yanlılık ve RMSE değerleri ile sonuçlanma eğiliminde olduğunu söylemiştir.

Diğer bir yandan, çalışma sonucunda elde edilen başka bir bulguya göre aynı sonlandırma kuralının kullanıldığı tüm koşullarda, madde havuzu büyüklüğünün artması ile ölçme kesinliği kestirim değerleri için elde edilen RMSE ve yanlılık değerlerinin genel olarak daha düşük elde edildiği bulunmuştur. Bu bulgu Eroğlu (2013)'nin çalışmasındaki sonuç ile benzerlik göstermektedir. Ayrıca yetenek kestirim yöntemi olarak EAP yönteminin kullanılmasının ölçme kesinliği kestirim değerlerinden olan RMSE ve yanlılık değerlerini düşürebileceği çalışmadan elde edilen bulgular arasında olup, Bulut ve Kan (2012) ile Eroğlu (2013)'ün çalışmalarından elde edilen bulgularla benzerlik taşımaktadır.

Çalışmadan elde edilen bulgulara göre bazı koşullarda ölçme kesinliği kestirim değerleri arasında bir uyumsuzluk olduğu görülmüştür. Beklenen durum en düşük RMSE ve yanlılık değeri ile en yüksek korelasyon değerinin aynı koşulda görülmesi gerektiğidir. Fakat bazı koşullarda bu değerler aynı paralellikte değildir. İleri zamanlarda yapılacak olan araştırmalarda, çalışma bu yönleriyle tekrarlanarak bu durumun nedeni tartışılabilir.

Özetlenecek olursa, BOBUT uygulamalarında içerik ağırlıklandırma ve madde kullanım sıklığı kontrol yöntemleri gibi madde havuzunda yapılan pratik sınırlandırmalar, uygulamaya katılan bireyleri maddeleri uygun olmayan zorlukta almaya zorlayabileceğinden, BOBUT uygulamalarının ölçme kesinliği ve test verimliliği etkilenebilir (Reese vd., 1999). Çalışma bu yönüyle alanyazınla paralellik göstermektedir.

Bölüm 5

Sonuç ve Öneriler

Bu bölümde bir önceki bölümde yer alan bulgu ve yorumlara dayanarak elde edilen sonuçlara ve önerilere yer verilmiştir.

Sonuçlar

Bu çalışma kapsamında, BOBUT uygulamalarında içerik ağırlıklandırma yapılmadığı ve yapıldığı durumda; farklı madde seçme, farklı yetenek kestirim ve farklı madde kullanım sıklığı kontrol yöntemlerinin birlikte kullanılmasının, gerçek ve kestirilen yetenek düzeyleri üzerinden hesaplanan ölçme kesinliği kestirim değerleri (RMSE, yanlılık ve korelasyon) üzerindeki etkisinin incelenmesi amaçlanmıştır. Bu amaç doğrultusunda da bir önceki bölümde yer alan bulgu ve yorumlara dayalı olarak elde edilen sonuçlar alt problem başlıkları altında sunulmuştur.

Birinci Alt Probleme İlişkin Sonuçlar

Çalışmanın birinci alt probleminde, BOBUT uygulamasında madde havuzu büyüklüğü 320 madde olan havuzdan elde edilen ölçme kesinliği kestirim değerlerinin; içerik ağırlıklandırma yapılmadığı ve yapıldığı durumda, madde seçme yöntemine, yetenek kestirim yöntemine ve madde kullanım sıklığı kontrol yöntemine göre nasıl değiştiğine ilişkin sonuçlar maddeler halinde aşağıda sunulmuştur.

1. İçerik ağırlıklandırma yapılmadığı ve yapıldığı durumda, yetenek kestirim EAP, madde kullanım sıklığı kontrol yönteminin KY olduğu ile içerik ağırlıklandırma yapılmadığı durumda, yetenek kestirim MLE, madde kullanım sıklığı kontrol yönteminin KY olduğu koşullar hariç diğer tüm koşullarda GEK yönteminin kullanılmasının ortalama hata değerini artırdığı görülmüştür.
2. 40 maddelik test uzunluğunun sabit olduğu her koşulda bir önceki maddede bahsedilen koşullar hariç, GEK yönteminin kullanıldığı koşulların ortalama hata

değerlerinin, diğer madde seçme yöntemlerinin kullanıldığı koşullardaki ortalama hata değerlerinden daha yüksek olduğu saptanmıştır.

3. Yetenek kestirim yöntemi olarak EAP seçildiğinde tüm koşullar için ortalama hata değeri daha düşük olarak elde edilmiştir.
4. Madde havuzu büyüklüğü 320 madde olan bu çalışmanın tüm koşulları altında, içerik ağırlıklandırma yapılmasının ortalama hatayı bir miktar artırdığı görülmüştür.
5. İçerik ağırlıklandırmanın hem yapılmadığı hem de yapıldığı durumda yer alan tüm koşullarda yetenek kestirim yönteminin MLE olmasının yanlılık değerini artırdığı görülmüştür.
6. En yüksek yanlılık değerine sahip koşulların GEK madde seçme yöntemi ile yetenek kestirim yöntemi olarak MLE yönteminin kullanıldığı koşulların olduğu gözlemlenmiştir.
7. İçerik ağırlıklandırmanın yapılmadığı ve yapıldığı durumlarda; GEK haricindeki diğer madde seçme yöntemlerinin kullanıldığı tüm koşulların ortalama yanlılık değerleri açısından birbirinden farklılaşmadığı bulunmuştur.
8. İçerik ağırlıklandırmanın yapılmadığı ve yapıldığı durumlarda; farklı madde seçme, farklı yetenek kestirim yöntemleri ile birlikte kullanılan madde kullanım sıklığı kontrol yöntemlerinden KY ile MUİ yöntemlerinin ortalama yanlılık değeri açısından birbirinden farklılaşmadığı bulunurken; ortalama hata değeri açısından MUİ yöntemi ile çoğu koşulda KY'ye göre daha düşük değerler elde edildiği saptanmıştır.
9. İçerik ağırlıklandırma yapılmasının yanlılık değerleri üzerinde bir etkisinin olmadığı sonucuna varılmıştır.
10. En yüksek korelasyon değerine içerik ağırlıklandırma yapılmadığı durumda; madde seçme yönteminin MFI, yetenek kestirim yönteminin EAP ve madde kullanım sıklığı kontrol yönteminin MUİ olduğu koşullar ile içerik ağırlıklandırmanın yapılmadığı durumda; madde seçme yönteminin MIS,

yetenek kestirim yönteminin EAP ve madde kullanım sıklığı kontrol yönteminin MUİ olduğu koşulda ulaşılmıştır.

11. En düşük korelasyon değerine ise içerik ağırlıklandırma yapıldığı durumda; madde seçme yönteminin MFI, yetenek kestirim yönteminin EAP ve madde kullanım sıklığı kontrol yönteminin KY olduğu koşulda ulaşılmıştır.

İkinci Alt Probleme İlişkin Sonuçlar

Çalışmanın ikinci alt probleminde, BOBUT uygulamasında madde havuzu büyüklüğü 1000 madde olan havuzdan elde edilen ölçme kesinliği kestirim değerlerinin; içerik ağırlıklandırma yapılmadığı ve yapıldığı durumda, madde seçme yöntemine, yetenek kestirim yöntemine ve madde kullanım sıklığı kontrol yöntemine göre nasıl değiştiğine ilişkin sonuçlar maddeler halinde aşağıda sunulmuştur.

1. Madde havuzu büyüklüğünün 1000 madde olduğu tüm koşullarda MFI yönteminin ortalama hata değerini düşürdüğü, GEK yönteminin kullanılmasının ise ortalama hata değerini artırdığı bulunmuştur.
2. MIS yönteminin kullanıldığı koşulların ortalama hata değerlerinin, MFI madde seçme yönteminin kullanıldığı koşullardaki ortalama hata değerlerinden daha yüksek, GEK madde seçme yöntemlerinin kullanıldığı koşullardaki ortalama hata değerlerinden daha düşük olduğu saptanmıştır.
3. Yetenek kestirim yöntemi olarak EAP seçildiğinde tüm koşullar için ortalama hata değeri daha düşük olarak elde edilmiştir.
4. Madde havuzu büyüklüğü 1000 madde olan bu çalışmanın tüm koşulları altında, içerik ağırlıklandırma yapılmasının ortalama hatayı bir miktar artırdığı görülmüştür.
5. İçerik ağırlıklandırmanın hem yapılmadığı hem de yapıldığı durumda yer alan tüm koşullarda yetenek kestirim yönteminin MLE olmasının yanlışlık değerini artırdığı görülmüştür.

6. En yüksek yanlılık değerine sahip koşulların GEK madde seçme yöntemi ile yetenek kestirim yöntemi olarak MLE yönteminin kullanıldığı koşulların olduğu gözlemlenmiştir.
7. İçerik ağırlıklandırmanın yapılmadığı ve yapıldığı durumlarda; MLE yönteminin olduğu koşullar hariç üç madde seçme yönteminin kullanıldığı tüm koşulların ortalama yanlılık değerleri açısından birbirinden farklılaşmadığı bulunmuştur.
8. İçerik ağırlıklandırmanın yapılmadığı ve yapıldığı durumlarda; farklı madde seçme, farklı yetenek kestirim yöntemleri ile birlikte kullanılan madde kullanım sıklığı kontrol yöntemlerinden KY ile MUİ yöntemlerinin ortalama yanlılık değeri ve ortalama hata değeri açısından birbirinden farklılaşmadığı görülmüştür.
9. İçerik ağırlıklandırma yapılmasının yanlılık değerleri üzerinde bir etkisinin olmadığı sonucuna varılmıştır.
10. En yüksek korelasyon değerine içerik ağırlıklandırmanın yapılmadığı durumda, madde seçme yönteminin MFI, yetenek kestirim yönteminin EAP ve her iki madde kullanım sıklığı kontrol yönteminin olduğu koşullarda ulaşılmıştır.
11. En düşük korelasyon değerine ise içerik ağırlıklandırma yapıldığı durumda, madde seçme yönteminin GEK, yetenek kestirim yönteminin MLE ve her iki madde kullanım sıklığı kontrol yönteminin olduğu koşullarda ulaşılmıştır.
12. Madde havuzu büyüklüğünün 1000 madde olduğu tüm koşullarda MFI yönteminin korelasyon değerini artırdığı, GEK yönteminin kullanılmasının ise korelasyon değerini düşürdüğü bulunmuştur.

Öneriler

Bu kısımda, çalışmadan elde edilen sonuçlara dayalı olarak geliştirilen uygulayıcılara ve araştırmacılara yönelik önerilere yer verilmiştir.

Uygulayıcılara Yönelik Öneriler

- Bu çalışmada ele alınan iki farklı madde havuzu büyüklüğü, ölçme kesinliği kestirim değerleri açısından değerlendirildiğinde; madde havuzu büyüklüğü 1000 madde olan havuzda, daha düşük ortalama hata değeri, daha yüksek korelasyon ve daha yansız hesaplamaların elde edildiği görülmüştür. Bu sebeple BOBUT uygulamaları geliştirilirken daha fazla sayıda madde içeren daha büyük madde havuzlarının tercih edilmesi önerilebilir.
- Çalışma sonucunda, madde havuzu büyüklüğü 1000 madde olan havuzda, MFI madde seçme yönteminin en kullanışlı yöntem olduğu görülmüştür. Dolayısıyla yeterli büyüklükte olan madde havuzlarında MFI madde seçme yönteminin tercih edilmesi önerilebilir.
- Yetenek kestirim yöntemi olarak EAP yönteminin her koşulda MLE yetenek kestirim yönteminden daha az hata ve daha az yanlılık ile kestirimler yaptığı sonucuna ulaşılmıştır. Bu yüzden çalışmada kullanılan veri seti de göz önünde bulundurularak yetenek kestirim yöntemi olarak EAP yönteminin kullanılması önerilebilir.
- Literatürde yer alan çalışma sonuçlarıyla da desteklenen içerik ağırlıklandırma yapılmasının ortalama hata değerini bir miktar artırdığı görülse de; BOBUT uygulamasında adillik olması hususunda içerik ağırlıklandırma yapılmasının kritik öneme sahip olduğu durumlar düşünüldüğünde, içerik ağırlıklandırma yapılmasının tercih edilmesi önerilebilir.
- Çalışmaya ait tüm sonuçlar genel bir çerçeveden değerlendirildiğinde, içerik ağırlıklandırma ve madde kullanım sıklığı kontrolü gibi madde havuzunda bazı sınırlandırmaların dahil edildiği BOBUT uygulamalarında; RMSE, yanlılık ve korelasyon değerleri bakımından daha iyi performans gösteren MFI madde seçme yöntemi ile EAP yetenek kestirim yönteminin birlikte kullanılması önerilebilir.

Arařtırmacılara Yönelik Öneriler

- Çalışma kapsamında MTK modeli kullanılmış, iki kategorili puanlanan maddelerden oluşan madde havuzunun parametreleri 3PLM kullanılarak kestirilmiştir. Sonraki çalışmalarda, çok kategorili maddeler veya karma maddelerin de kullanıldığı BOBUT uygulamaları gerçekleştirilebilir.
- Bu çalışmada, normal dağılımdan üretilen bireylerin gerçek yetenek düzeyleri ile farklı koşullarda kestirilen yetenek düzeylerinden elde edilen ölçme kesinliği kestirim değerleri karşılaştırılmıştır. Bahsi geçen bu farklı koşulların, farklı yetenek grupları için nasıl değiştiği araştırılabilir.
- Çalışmada kullanılan madde havuzu, arařtırmacı tarafından belirlenen madde parametre değerlerine göre oluşturulmuştur. Farklı madde parametre değerleri ve dağılımları ile madde havuzu oluşturulup çalışma tekrarlanabilir.
- Bu çalışmada iki farklı madde havuzu büyüklüğü ele alınmıştır. Çok sayıda daha farklı madde havuzu büyüklüğüne sahip madde havuzları ile benzer bir çalışma gerçekleştirilebilir.
- Çalışma kapsamında başlama kuralı olarak yetenek düzeyi "0" ($\theta = 0$) olarak belirlenmiştir. Farklı başlama kuralları çalışmaya dahil edilebilir.
- Bu çalışmanın bir sınırlılığı, içerik ağırlıklandırma yöntemi olarak sadece SBOBUT yönteminin kullanılmış olmasıdır. İleride gerçekleştirilecek olan çalışmalarda diğer içerik ağırlıklandırma yöntemleri çalışmaya dahil edilerek yöntem karşılaştırması yapılabilir.
- Bu çalışmada madde seçme yöntemlerinden sadece MFI, MIS ve GEK kullanılmıştır. Gerçekleştirilecek olan benzer çalışmalarda diğer madde seçme yöntemleri de ele alınabilir.
- Çalışma kapsamında madde kullanım sıklığı kontrolü yöntemi olarak KY ve MUİ yöntemleri kullanılmıştır. Gerçekleştirilecek olan benzer çalışmalarda diğer

madde kullanım sıklığı kontrol yöntemlerinin BOBUT uygulamalarındaki performansı karşılaştırılabilir.

- Bu çalışmada, sabit sonlandırma uzunluğu kuralı kullanılarak 40 madde ile uygulama gerçekleştirilmiştir. Literatür incelendiğinde genel olarak madde sayılarının 25, 35, 40 ve 60 arasında değiştiği görülmüş olup, bu çalışma için de benzer bir değer (40) seçilmiştir (Barrada vd., 2009; Cheng & Chang, 2007; Leung vd., 2002; Leung vd., 2003). Farklı test uzunlukları ya da farklı sonlandırma kuralları ile oluşturulacak simülasyon koşullarının sonuçları, üstün ve zayıf yönleri de değerlendirilebilir.
- Bu çalışmada simülatif veri ile gerçekleştirilen koşullar, BOBUT uygulamasının amacına uygun olarak gerçek veri seti ile de gerçekleştirilebilir.

Kaynaklar

- Babcock, B. ve Weiss, D. J. (2009). Termination criteria in computerized adaptive tests: Variable-length CATs are not biased. *Proceedings of the 2009 GMAC conference on computerized adaptive testing* (14. Cilt).
- Babcock, B. ve Weiss, D. J. (2014). Termination criteria in computerized adaptive tests: Do variable-length CATs provide efficient and effective measurement? *Journal of Computerized Adaptive Testing*, 1, 1–5. <https://doi.org/10.7333/1212-0101001>
- Barış-Pekmezci, F. ve Şengül-Avşar, A. (2021). A guide for more accurate and precise estimations in simulative unidimensional IRT models. *International Journal of Assessment Tools in Education*, 8(2), 423–453. <https://doi.org/10.21449/ijate.790289>
- Barrada, J. R., Abad F. J. ve Veldkamp, B. P. (2009a). Comparison of methods for controlling maximum exposure rates in computerized adaptive testing. *Psicothema*, 21(2), 313-320.
- Barrada, J. R., Mazuela, P. ve Olea, J. (2006). Maximum information stratification method for controlling item exposure in computerized adaptive testing. *Psicothema*, 18(1), 156-159.
- Barrada, J. R., Veldkamp, B. P. ve Olea, J. (2009b). Multiple maximum exposure rates in computerized adaptive testing. *Applied Psychological Measurement*, 33(1), 58–73. <https://doi.org/10.1177/0146621608315329>
- Bejar, I. I. ve Weiss, D. J. (1979). *Computer programs for scoring test data with item characteristic curve models* (Rapor No: 79-1). University of Minnesota, Department of Psychology.
- Bergstrom B. A. ve Lunz, M. E. (1999). CAT for certification and licensure. F. Drasgow ve J. B. Olson-Bunchanan (Ed.), *Innovations in computerized assessment* içinde (s. 67-91). Lawrence Erlbaum Associates.

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. F. M. Lord ve M. R. Novick (Ed.), *Statistical theories of mental test scores* içinde (s. 397-472). Addison-Wesley.
- Bjorner, J. B., Kosinski, M. ve Ware, J. E. (2005). Computerized adaptive testing and item banking. P. Fayers ve R. Hayes (Ed.), *Assessing quality of life in clinical trials: Methods and practice* içinde (s. 95-112). Oxford University Press.
- Bock, D. R. ve Mislevy, R. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6(4), 431-444.
<https://doi.org/10.1177/014662168200600405>
- 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.
- Carroll, G. R. ve Harrison, J. R. (1998). Organizational demography and culture: Insights from a formal model and simulation. *Administrative Science Quarterly*, 43, 637-667.
<https://doi.org/10.2307/2393678>
- Chang, H. H. ve Ying, Z. (1999). A-stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, 23(3), 211-222.
<https://doi.org/10.1177/01466219922031338>
- Chang, H. H. (2004). Understanding computerized adaptive testing: From Robbins - Monro to Lord and beyond. D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* içinde (s. 117-133). Sage.
- Chen, S. Y. (1999). *Comparison of item selection rules including precision, content, and exposure considerations at the early stages of computerized adaptive testing* (Doktora tezi). The University of Iowa.

- Chen, S. Y. ve Ankenman, 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. <https://doi.org/10.1111/j.1745-3984.2004.tb01112.x>
- Chen, S. Y., Lei, P. W. ve Liao, W. H. (2008). Controlling item exposure and test overlap on the fly in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, 61(2), 471-492. <https://doi.org/10.1348/000711006X129553>
- Cheng, Y., Chang, H. H. ve Yi, Q. (2007). Two-phase item selection procedure for flexible content balancing in CAT. *Applied Psychological Measurement*, 31(6), 467–482. <https://doi.org/10.1177/0146621606292933>
- Cheng, Y., Chang, H. H., Douglas, J. ve Guo, F. (2009). Constraint-weighted a-stratification for computerized adaptive testing with nonstatistical constraints. *Educational and Psychological Measurement*, 69(1), 35–49. <https://doi.org/10.1177/0013164408322030>
- Cheng, Y. ve Chang, H. H. (2009). The maximum priority index method for severely constrained item selection in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, 62(2), 369-383. <https://doi.org/10.1348/000711008X304376>
- Chien, Y. (2017). Weighted penalty model for content balancing in CAT. *International Journal of Intelligent Technologies and Applied Statistics*, 10(3), 129-144, <https://doi.org/10.6148/IJITAS.2017.1003.01>
- Choi, S. W., Grady, M. W. ve Dodd, B. G. (2010). A new stopping rule for computerized adaptive testing. *Educational and Psychological Measurement*, 71(1), 37–53.
- Choi S. W. ve Swartz R. J. (2009). Comparison of CAT item selection criteria for polytomous items. *Applied Psychological Measurement*, 32, 419–440. <https://doi.org/10.1177/0146621608327801>

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Baskı). Routledge.
- Cohen, J. ve Albright, L. (2014). *Smarter Balanced adaptive item selection algorithm design report*. <http://www.smarterapp.org/specs/AdaptiveAlgorithm.html>
- Conejo, R., Guzmán, E., Millán, E., Trella, M., PérezdelaCruz, J. L. ve Ríos, A. (2004). SIETTE: A web-based tool for adaptive testing. *J. Artif. Intell. Educ*, 14, 29–61
- Crocker, L. ve Algina, J. (1986). Introduction to item response theory. L. Crocker ve J. Algina (Ed.), *Introduction to classical and modern test theory*. Cengage Learning.
- Davey, T. (2011). *A guide to computer adaptive testing systems*. Council of Chief State School Officers.
- Davey, T. ve Parshall, C. G. (1995). *New algorithms for item selection and exposure control with computerized adaptive testing*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Davey, T. ve Nering, N. (2002). Controlling item exposure and maintaining item security. C. N. Mills, M. T. Potenza, J. J. Fremer ve W. C. Ward, (Ed.), *Computer-based testing: Building the foundation for future assessments* (s. 165-191). Erlbaum.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. The Guilford Press.
- DeMars, C. (2010). *Item response theory*. Oxford University Press, Inc.
- Dodd, B. G., Koch, W. R. ve De Ayala, R. J. (1989). Operational characteristics of adaptive testing procedures using the graded response model. *Applied Psychological Measurement*, 13, 129–143. <https://doi.org/10.1177/014662168901300202>
- Dodd, B. G., Koch, W. R. ve De Ayala, R. J. (1993). Computerized adaptive testing using the partial credit model: Effects of item pool characteristics and different stopping rules. *Educational and Psychological Measurement*, 53(1), 61-77. <https://doi.org/10.1177/0013164493053001005>

- Embretson, S. E. ve Reise, S. P. (2000). *Item response theory for psychologist*. Lawrence Erlbaum Associates Publishers
- Erođlu, M. G. (2013). *Bireyselleřtirilmiř bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının ölçme kesinliđi ve test uzunluđu açısından karşılařtırılması* (Doktora tezi). Hacettepe Üniversitesi, Ankara.
- Evans, J. J. (2010). *Comparability of examinee proficiency scores on computer adaptive tests using real and simulated data* (Doktora tezi). Rutgers University-Graduate School, New Brunswick.
- Flaugher, R. (2000). Item pools. H. Wainer (Ed.), *A primer* (2. Baskı.). Lawrence Erlbaum Associates.
- Folk, V. G. ve Smith, R. L. (2002). Models for delivery of CBTs. Mills, Potenza, Framer ve Ward (Ed.), *Computer-based testing: Building the foundations for future assessments* içinde (s. 41-66). Lawrence Erlbaum Associates.
- Georgiadou, E., Triantafillou, E. ve Economides, 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), 4-38.
- Gu, L. ve Reckase, M. D. (2007). *Designing optimal item pools for computerized adaptive tests with sympon-hetter exposure control*. Paper presented at the Item Exposure Paper session at the 2007 GMAC Conference on Computerized Adaptive Testing. <http://iacat.org/sites/default/files/biblio/cat07gu-reckase.pdf>
- Gialluca, K. A. ve Weiss, D. J. (1979). *Efficiency of an adaptive inter-subset branching strategy in the measurement of classroom achievement* (Rapor No. 79-6). University of Minnesota. <http://www.psych.umn.edu/psylabs/catcentral/>
- Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L. ve Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of*

Educational Measurement, 21(4), 347–360. <https://doi.org/10.1111/j.1745-3984.1984.tb01039.x>

Hambleton, R. (2005). *Preface to linear models to optimal test design by van der Linden*. W. J. Springer.

Hambleton, R. K. ve Swaminathan, H. (1985). *Item response theory: Principles and applications*. Kluwer Nijhoff Publishing.

Hambleton, R. K., Swaminathan, H. ve Rogers, H. J. (1991). *Fundamentals of item response theory* (1. Baskı). SAGE Publications, Inc.

Han, K. (2012). SimulCAT: Windows application that simulates computerized adaptive test administration. *Applied Psychological Measurement*, 36, 64-66. <https://doi.org/10.1177/0146621611414407>

Han, K. T. (2018). Components of the item selection algorithm in computerized adaptive testing. *Journal of Educational Evaluation for Health Professions*, 15(7), <https://doi.org/10.3352/jeehp.2018.15.7>

Hau, K. T. ve Chang, H. H. (2001). Item selection in computerized adaptive testing: Should more discriminating items be used first? *Journal of Educational Measurement*, 38(3), 249–266. <https://doi.org/10.1111/j.1745-3984.2001.tb01126.x>

He, W. (2010). *Optimal item pool design for a highly constrained computerized adaptive test* (Doktora tezi). Michigan State University.

He, W., Diao, Q. ve Hauser, C. (2014). A comparison of four item-selection methods for severely constrained CATs. *Educational and Psychological Measurement*, 74(4), 677-696. <https://doi.org/10.1177/0013164413517503>

He, W. ve Reckase, M. D. (2013). Item pool design for an operational variable length computerized adaptive test. *Educational and Psychological Measurement*, 74(3), 473-494. <https://doi.org/10.1177/0013164413509629>

- Huebner, A. (2012). Item overexposure in computerized classification tests using sequential item selection. *Practical Assessment, Research, and Evaluation*, 17(12). <https://doi.org/10.7275/nr1c-yv82>
- Jensem, C. J. (1977). Bayesian tailored testing and the influence of item bank characteristics. *Applied Psychological Measurement*, 1, 111-120. <https://doi.org/10.1177/014662167700100117>
- Jodoin, M. G. (2003). *Psychometric properties of several computer-based test designs with ideal and constrained item pools* (Doktora tezi). University of Massachusetts Amherst.
- Kalender, İ. (2011). *Effects of different computerized adaptive testing strategies on recovery of ability* (Doktora tezi). MiddleEast Technical University, Ankara.
- Khalilzadeh, J. & Tasci, A. D. A. (2017). Large sample size, significance level, and the effect size: Solutions to perils of using big data for academic research. *Tourism Management*, 62, 89-96. <https://doi.org/10.1016/j.tourman.2017.03.026>
- Kingsbury, G. G. ve Zara, A. R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, 359-375. https://doi.org/10.1207/s15324818ame0204_6
- Kingsbury, G. G. ve Zara, A. R. (1991). A comparison of procedures for content-sensitive item selection in computerized adaptive tests. *Applied Measurement in Education*, 4(3), 241–261. https://doi.org/10.1207/s15324818ame0403_4
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4(863), 1-12. <https://doi.org/10.3389/fpsyg.2013.00863>

- Leung, C.-K. , Chang, H.-H. & Hau, K.-T. (2002). Item selection in computerized adaptive testing: Improving the a-stratified design with the Sympton-Hetter algorithm. *Applied Psychological Measurement*, 376-392. <https://doi.org/10.1177/014662102237795>
- Leung, C. K., Chang, H. H. ve Hau, K. T. (2003a). Incorporation of content balancing requirements in stratification designs for computerized adaptive testing. *Educational and Psychological Measurement*, 63(2), 257-270. <https://doi.org/10.1177/0013164403251326>
- Leung, C. K., Chang, H. H. ve Hau, K. T. (2003b). Computerized adaptive testing: A comparison of three content balancing methods. *Journal of Technology, Learning, and Assessment*, 2(5), 3-15.
- Levine, T. R.ve Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research*, 28(4), 612-625. <https://doi.org/10.1111/j.1468-2958.2002.tb00828.x>
- Lin, C. (2011). Item selection criteria with practical constraints for computerized classification testing. *Applied Psychological Measurement* 71(1), 20-36. <https://doi.org/10.1177/0013164410387336>
- Lin, H. (2012). *Item selection methods in multidimensional computerized adaptive testing adopting polytomously-scored items under multidimensional generalized partial credit model* (Doktora tezi). University of Illinois at Urbana-Champaign.
- Liu, X. (2019). *Optimizing design of incorporating off-grade items for constrained computerized adaptive testing in K-12 assessment* (Doktora tezi). The University of Iowa.
- Lord, F.M. (1977). A broad-range tailored test of verbal ability. *Applied Psychological Measurement*, 1, 95-100. <https://doi.org/10.1177/014662167700100115>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates.

- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.
- Lord, F. M. ve Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- Luecht, R. M., De Champlain, A. ve Nungester, R. J. (1998). Maintaining content validity in computerized adaptive testing. *Advances in Health Sciences Education*, 3, 29-41. <https://doi.org/10.1023/A:1009789314011>
- Lunz, M. E. ve Stahl, J. A. (1998). *Patterns of item exposure using a randomized CAT algorithm*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Magis, D. ve Raiche, G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software*, 48(8), 1–31. <https://doi.org/10.18637/jss.v048.i08>
- Magis, D., Raiche, G. ve Barrada, J. R. (2018). Package 'catR': Generation of IRT response patterns under computerized adaptive testing. <https://cran.r-project.org/web/packages/catR/catR.pdf>
- Magis, D., Yan, D. ve von Davier, A. (2017). *Computerized adaptive and multistage testing with R: Using packages catR and mstR*. Springer International Publishing.
- Mahmud, J. (2017). Item response theory: A basic concept. *Educational Research and Reviews*, 12(5), 258-266. <https://doi.org/10.5897/ERR2017.3147>
- Mao, L. (2014). *Designing p-optimal item pools for multidimensional computerized adaptive testing* (Doktora tezi). Michigan State University.
- Maurelli, V. ve Weiss, D. J. (1981). *Factors influencing the psychometric characteristics of an adaptive testing strategy for test batteries*. (Rapor No. 81-4). University of Minnesota. <http://www.psych.umn.edu/psylabs/catcentral/>

- Meijer, R. R. ve Tendeiro, J. N. (2018). Unidimensional item response theory. P. Irwing, T. Booth ve D. J. Hugh (Ed.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* içinde (s. 413-433). Wiley. <https://doi.org/10.1002/9781118489772.ch15>
- Mills, C. N. ve Stocking, M.L. (1996). Practical issues in large-scale Computerized Adaptive Testing. *Applied Measurement in Education*, 9(4), 287-304. https://doi.org/10.1207/s15324818ame0904_1
- Millman, J. ve Arter, J. A. (1984). Issues in item banking. *Journal of Educational Measurement*, 21, 315-330. <https://doi.org/10.1111/j.1745-3984.1984.tb01037.x>
- Newman, L. S. (1995). *Content validity of a computerized adaptive licensing and certification examination: A comparison of content-balancing methods* (Doktora tezi). Temple University.
- Nguyen, T. H., Han, H. R., Kim, M. T. ve Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *The Patient: Patient-Centered Outcomes Research*, 7(1), 23-35.
- Orcutt, V. L. (2002). *Computerized adaptive testing: Some issues in development*. Paper presented at the annual meeting of the Educational Research Exchange, University of North Texas.
- Paap, M. C. S., Kroeze, K. A., Terwee, C. A., van der Palen, J. ve Veldkamp, B. P. (2017). Item usage in a multidimensional computerized adaptive test (MCAT) measuring health-related quality of life. *Quality of Life Research*, 26(11), 2909-2918.
- Piromsombat, C. (2014). *Differential item functioning in computerized adaptive testing: Can CAT self-adjust enough?* (Doktora tezi). University of Minnesota.
- Reckase, M. D. (2010). Designing item pools to optimize the functioning of a computerized adaptive test. *Psychological Test and Assessment Modeling*, 127-141.

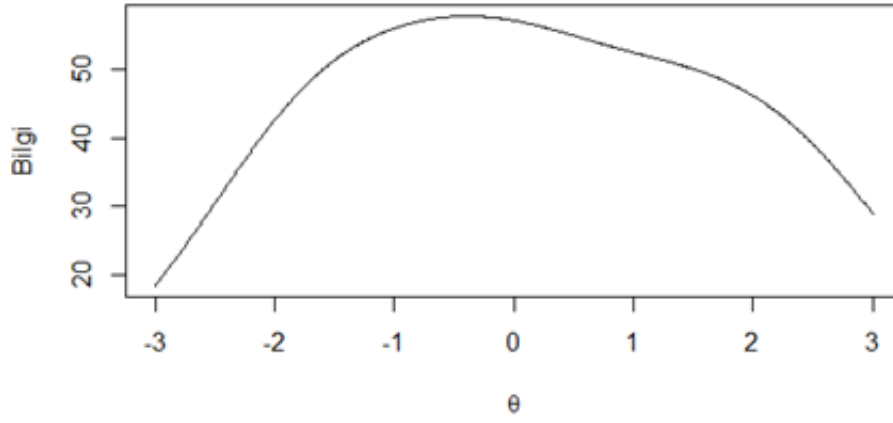
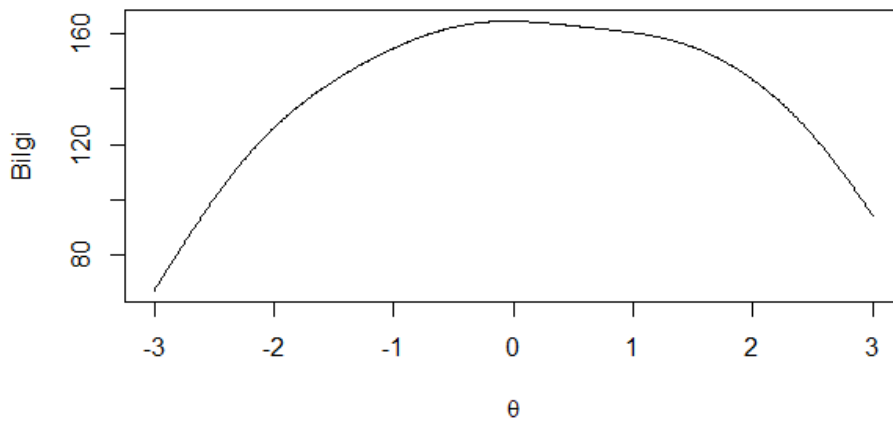
- Reese, L. M., Schnipke, D. L. ve Luebke, S. W. (1999). *Incorporating content constraints into a multi-stage adaptive testlet design*. Law School Admission Council Computerized Testing Report.
- Repenning, N. (2002). A simulation-based approach to understanding the dynamics of innovation implementation. *Organization Science*, 13, 109–127. <https://doi.org/10.1287/orsc.13.2.109.535>
- Revuelta, J. ve Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 35, 311-327. <https://doi.org/10.1111/j.1745-3984.1998.tb00541.x>
- Risk, N. M. (2015). *The impact of item parameter drift in computer adaptive testing (CAT)* (Doktora tezi). The University of Illinois at Chicago.
- Rudner, L. M. (1998). *An on-line, interactive computer adaptive testing mini tutorial*. <http://echo.edres.org:8080/scripts/cat/catdemo.htm>
- Segall, D. O. (2004). Computerized adaptive testing. K. Kempf-Lenard (Ed.), *The Encyclopedia of social measurement* içinde (s. 429-438). Academic Press.
- Sireci, S. (2003). Computerized adaptive testing: An introduction. Wall ve Walz (Ed.), *Measuring up: Assessment issues for teachers, counselors and administrators* içinde (s. 684-694). CAPS Press.
- Song, T. (2010). *The effect of fitting a unidimensional irt model to multidimensional data in content-balanced computerized adaptive testing* (Doktora tezi). Michigan State University.
- Stafford, R. E, Runyon, C. R., Casabianca, J. M. ve Dodd, B. G. (2019). Comparing computer adaptive testing stopping rules under the generalized partial-credit model. *Behavior Research Methods*, 51,1305–1320.
- Stocking, M. L. (1994). Three practical issues for modern adaptive testing item pools 1. *ETS Research Report Series*, 1994(1), i-34. <https://doi.org/10.3758/s13428-018-1068-x>

- Stocking, M. L. ve Lewis, C. (1995). *A new method of controlling item exposure in computerized adaptive testing* (Rapor No. ETS-RR-95-25).
- Stocking, M. L. ve Lewis, C. (1998). Controlling item exposure conditional on ability in computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, 23, 57–75. <https://doi.org/10.3102/10769986023001057>
- Stocking, M. L. ve Swanson, L. (1993). A method for severely constrained item selection in adaptive testing. *Applied Psychological Measurement*, 17(3), 277–292. <https://doi.org/10.1177/014662169301700308>
- Stone, E. ve Davey, T. (2011). Computer adaptive testing for students with disabilities: A review of the literature. *ETS Research Report Series*, 2011(2). <https://doi.org/10.1002/j.2333-8504.2011.tb02268.x>
- Sympson, J. ve Hetter, R. (1985). Controlling item exposure rates in computerized adaptive testing. *Proceedings of the 27th annual meeting of the Military Testing Association* (s. 973–977), Navy Personnel Research and Development Center.
- 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, 21-36. <http://dx.doi.org/10.14689/ejer.2017.69.2>
- Şenel, S. (2021). *Bilgisayar ortamında bireye uyarlanmış testler* (2. Baskı). Pegem Akademi.
- Thissen, D. ve Mislevy, R. J. (2000). Testing algorithms. H. Wainer, (Ed.), *Computerized adaptive testing: A primer* (s. 101-133), Lawrence Erlbaum Associates.
- Thompson, N. A. ve Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research and Evaluation*, 16(1), 1-9. <https://doi.org/10.1007/BF02294775>

- van der Linden, W. J. (1998). Bayesian item selection criteria for adaptive testing. *Psychometrika*, 63, 201-216. <https://doi.org/10.1007/BF02294775>
- van der Linden, W. J., Ariel, A. ve Veldkamp, B. P. (2006). Assembling a CAT item pool as a set of linear tests. *Journal of Educational and Behavioral Statistics*, 31, 81- 99.
- van der Linden, W. J. ve Veldkamp, B. P. (2004). Constraining item exposure in computerized adaptive testing with shadow tests. *Journal of Educational and Behavioral Statistics*, 29, 273-291.
- Veldkamp, B. P. ve van der Linden, W. J. (2000). Designing item pools for computerized adaptive testing. van der Linden, W. J. ve Glas, G. A. (Ed.), *Computerized adaptive testing: Theory and practice*. Springer. https://doi.org/10.1007/0-306-47531-6_8
- Wainer, H. (2000). Introduction and history. H. Wainer (Ed.), *Computerized adaptive testing: A primer* (2. Baskı). Lawrence Erlbaum Associates.
- Wainer, H. ve Mislevy, R. J. (2000). Item response theory, item calibration, and proficiency estimation. H. Wainer (Ed.), *Computerized adaptive testing: A primer* (2. Baskı.). Lawrence Erlbaum Associates.
- Wang, K. (2017). *A fair comparison of the performance of computerized adaptive testing and multistage adaptive testing* (Doktora tezi). Michigan State University.
- Wang, T. & Vispoel, W. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(2), 109-135. <https://doi.org/10.1111/j.1745-3984.1998.tb00530.x>
- Wang, S. ve Zhang, L. (2017). *Effects of ignoring discrimination parameter in CAT item selection on student scores*. Presented at the National Council on Measurement in Education (NCME) conferences, San Antonio, TX.
- Way, W. D. (1998). Protecting the integrity of computerized testing item pools. *Educational Measurement: Issues and Practice*, 17(4), 17-27. <https://doi.org/10.1111/j.1745-3992.1998.tb00632.x>

- Wei, H. ve Lin, J. (2015). Using out-of-level items in computerized adaptive testing. *International Journal of Testing*, 15(1), 50–70. <https://doi.org/10.1080/15305058.2014.979492>
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, 6, 473-492. <https://doi.org/10.1177/014662168200600408>
- Weiss, D. J. (1983). Latent trait theory and adaptive testing. David J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (s. 5-7). Academic Press.
- Weiss, D. J. ve Kingsbury, G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361-375. <https://doi.org/10.1111/j.1745-3984.1984.tb01040.x>
- Wise, S. L. (1997). *Overview of practical issues in a CAT program*. Paper presented at the annual meeting of the National Council on Measurement in Education (NCME), Chicago, IL.
- Wise, S. L., Kingsbury, G. G. ve Webb, N. L. (2015). Evaluating content alignment in computerized adaptive testing. *Educational Measurement: Issues and Practice*, 34(4), 41-48. <https://doi.org/10.1111/emip.12094>
- Yan, D., Von Davier, A. A. ve Lewis, C. (Eds.). (2016). *Computerized multistage testing: Theory and applications*. CRC Press.
- Yang, L. (2016). *Enhancing item pool utilization when designing multistage computerized adaptive tests* (Doktora tezi). Michigan State University.
- Yao, L. (2019). Item Selection methods for computer adaptive testing with passages. *front. Psychol.* 10, 240. <https://doi.org/10.3389/fpsyg.2019.0024>

- Yi, Q., Zhang, J. ve Chang, H.-H. (2006). Assessing CAT test security severity. *Applied Psychological Measurement*, 30(1), 62-63. <https://doi.org/10.1177/0146621605280354>
- Zanon, C., Hutz, C. S., Yoo, H. ve Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica*, 29, 18. <https://doi.org/10.1186/s41155-016-0040-x>
- Zheng, Y., Chang, C. H. ve Chang, H.-H. (2013). Content-balancing strategy in bifactor computerized adaptive patient-reported outcome measurement. *Quality of Life Research*, 22, 491-499. <https://doi.org/10.1007/s11136-012-0179-6>
- Zott, C. (2003). Dynamic capabilities and the emergence of intra-industry differential firm performance: Insights from a simulation study. *Strategic Management Journal*, 24, 97–12. <https://doi.org/10.1002/smj.288>

EK-A: 320 ve 1000 Maddelik Havuzlara Ait Test Bilgi Fonksiyonları**320 maddelik madde havuzuna ait test bilgi fonksiyonu****1000 maddelik madde havuzuna ait test bilgi fonksiyonu**

EK-B: 320 ve 1000 Maddelik Havuzdan Rastgele Seçilen 10 Maddeye Ait a, b ve c**Değerleri**

Sayı	a	b	c
1	1,597	-1,407	0,186
2	1,175	-0,767	0,095
3	1,287	0,437	0,190
4	1,995	2,449	0,080
5	1,826	-1,790	0,169
6	1,944	2,390	0,084
7	1,775	2,668	0,055
8	1,739	0,965	0,179
9	1,121	0,775	0,153
10	1,715	-2,629	0,191

Sayı	a	b	c
1	1,003	-0,262	0,054
2	0,844	-0,505	0,060
3	1,088	1,022	0,078
4	1,895	1,541	0,170
5	0,802	2,470	0,175
6	1,073	1,920	0,159
7	1,002	-2,438	0,092
8	0,980	-1,933	0,146
9	1,578	0,853	0,194
10	1,383	2,246	0,131

EK-C: Rastgele Seçilen Bir Bireyin Cevaplaması Gereken 40 Maddenin İçerik Alanlarına Göre Dağılımı

```

Expected proportions of items per subgroup:
Bolum1 Bolum2 Bolum3 Bolum4
0.2    0.4    0.3    0.1

Adaptive test details:
Nr      1      2      3      4      5      6      7      8      9      10     11     12     13     14     15     16     17     18     19     20
Item   146    138    206    163    152    234    131    255    58    269    157    207    33    239    100    67    282    303    110    108
Resp.   1      0      1      1      1      0      0      1      1      0      0      1      1      1      0      1      1      0      0      0
Est.    0.438 -0.128 0.172 0.311 0.548 0.364 0.188 0.293 0.399 0.301 0.173 0.279 0.332 0.365 0.297 0.342 0.391 0.311 0.255 0.157
SE      0.441* 0.584* 0.578* 0.551* 0.51* 0.501* 0.481* 0.456* 0.431* 0.412* 0.361 0.342 0.332 0.324 0.313 0.305 0.298 0.293 0.285 0.28

Nr      21     22     23     24     25     26     27     28     29     30     31     32     33     34     35     36     37     38     39     40
Item   310    179    53     258    48     240    295    235    301    137    83     320    284    190    245    51     120    273    97     136
Resp.   1      0      1      1      1      1      1      1      0      1      1      0      1      0      1      1      0      1      1      0
Est.    0.18   0.14   0.169 0.204 0.236 0.26   0.291 0.306 0.283 0.322 0.335 0.306 0.318 0.281 0.296 0.317 0.292 0.302 0.32 0.292
SE      0.274 0.269 0.262 0.258 0.253 0.249 0.246 0.243 0.239 0.236 0.234 0.23 0.228 0.225 0.223 0.221 0.217 0.215 0.214 0.211

(*: Exact SE)

Satisfied stopping rule:
Length of test

Final results:
Length of adaptive test: 40 items
Final ability estimator: Expected a posteriori (EAP) estimator
Final prior distribution: N(0,1) prior
Final standard error: asymptotic SE (ASE)
Type of ASE formula: classic formula
Final ability estimate (SE): 0.292 (0.211)
95% confidence interval: [-0.121,0.705]

Proportions of items per subgroup (expected and observed)
at the end of the test:
Bolum1 Bolum2 Bolum3 Bolum4
Exp. 0.2 0.4 0.3 0.1
Obs. 0.2 0.4 0.3 0.1

Items administered per subgroup:
Bolum1: 33, 138, 235, 255, 273, 284, 303, 310
Bolum2: 48, 100, 108, 120, 131, 136, 137, 152, 179, 206, 207, 245, 269, 282, 295, 320
Bolum3: 51, 53, 58, 67, 83, 97, 110, 146, 157, 234, 240, 301
Bolum4: 163, 190, 239, 258

```

EK-Ç: 320 Maddelik Havuza Ait a, b ve c Parametrelerinin Güven Aralığı Değerleri

	a	b	c
Ortalama	1,421	-0,070	0,125
Serbestlik Derecesi	0,348	1,655	0,045
Minimum Değer	0,802	-2,922	0,050
1. Dörtte Birlik	1,142	-1,456	0,086
2. Dörtte Birlik	1,396	-0,136	0,123
3. Dörtte Birlik	1,722	1,337	0,163
Maksimum Değer	1,995	2,956	0,200
Fark	0,580	2,793	0,077

Madde Parametreleri	"a" parametresi	"b" parametresi	"c" parametresi
Ortalama (μ)	1,421	-0,070	0,125
Varyans (δ)	0,121	2,729	0,002
Standart Sapma (δ^2)	0,348	1,652	0,045
Standart Hata (se)	0,036	0,193	0,006
Güven aralığının olasılık değeri	0,025	0,025	0,025
Serbestlik derecesi	319,000	319,000	319,000
Z tablo değeri	1,960	1,960	1,960
%95 Güven Aralığı Alt Limit	1,350	-0,449	0,113
%95 Güven Aralığı Üst Limit	1,492	0,309	0,137
Güven Aralığı Genişliği	0,141	0,758	0,025
Güven aralığı gösterimi	$1,35 \leq CI \leq 1,49$	$-0,45 \leq CI \leq 0,31$	$0,11 \leq CI \leq 0,14$

EK-D: 1000 Maddelik Havuza Ait a, b ve c Parametrelerinin Güven Aralığı Değerleri

	a	b	c
Ortalama	1,388	-0,002	0,123
Serbestlik Derecesi	0,356	1,730	0,044
Minimum Değer	0,801	-2,992	0,050
1. Dörtte Birlik	1,073	-1,452	0,085
2. Dörtte Birlik	1,377	-0,100	0,122
3. Dörtte Birlik	1,709	1,484	0,163
Maksimum Değer	1,999	3,000	0,200
Fark	0,636	2,936	0,078

Madde Parametreleri	"a" parametresi	"b" parametresi	"c" parametresi
Ortalama (μ)	1,388	-0,002	0,123
Varyans (δ)	0,126	2,991	0,002
Standart Sapma (δ^2)	0,355	1,729	0,045
Standart Hata (se)	0,037	0,202	0,006
Güven aralığının olasılık değeri	0,025	0,025	0,025
Serbestlik derecesi	319,000	319,000	319,000
Z tablo değeri	1,960	1,960	1,960
%95 Güven Aralığı Alt Limit	1,316	-0,399	0,111
%95 Güven Aralığı Üst Limit	1,460	0,395	0,135
Güven Aralığı Genişliği	0,144	0,793	0,025
Güven aralığı gösterimi	$1,32 \leq CI \leq 1,46$	$-0,4 \leq CI \leq 0,39$	$0,11 \leq CI \leq 0,14$

EK-E: Simülasyon Geçerliği İçin Kestirimi Yapılan Madde Parametrelerinin RMSE ve Yanlılık Değerleri (320 Madde Havuz Büyüklüğü)

Madde Seçme Yöntemi	Yetenek Kestirim Yöntemi	Madde Kullanım Sıklığı Kontrol Yöntemi	İçerik Ağırlıklandırma Yok						İçerik Ağırlıklandırma Var					
			a		b		c		a		b		c	
			RMSE	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık
MFI	EAP	KY	4,828	-0,563	4,462	2,652	0,288	0,132	7,232	-0,126	5,160	-0,254	0,315	0,159
		MUİ	4,727	-1,405	5,455	-1,387	0,302	0,144	6,654	-1,405	6,957	-0,450	0,289	0,132
	MLE	KY	5,164	-0,798	4,112	-0,157	0,293	0,140	6,654	-1,405	4,957	-0,450	0,289	0,132
		MUİ	6,837	-1,708	5,618	-0,096	0,319	0,135	9,108	1,213	3,833	0,117	0,284	0,117
MIS	EAP	KY	4,714	0,471	7,067	0,264	0,275	0,170	5,340	-1,035	3,390	-0,307	0,327	0,222
		MUİ	4,776	-0,623	5,480	0,010	0,299	0,166	3,714	-1,277	4,274	2,735	0,295	0,147
	MLE	KY	7,451	-0,403	6,578	2,006	0,316	0,162	7,435	0,180	4,901	1,970	0,323	0,197
		MUİ	6,280	0,709	4,983	2,528	0,325	0,173	4,151	-0,863	3,143	-1,290	0,281	0,162
GEK	EAP	KY	7,100	0,114	6,712	2,645	0,295	0,192	4,922	-1,172	4,112	0,760	0,279	0,112
		MUİ	4,036	0,899	5,907	1,082	0,275	0,109	10,665	1,553	3,897	0,669	0,337	0,209
	MLE	KY	7,060	0,922	2,175	0,251	0,298	0,161	7,170	-1,259	3,788	-0,646	0,284	0,163
		MUİ	6,327	0,127	4,936	-0,480	0,264	0,110	4,406	0,807	4,855	1,967	0,267	0,098

EK-F: Simülasyon Geçerliği İçin Kestirimi Yapılan Madde Parametrelerinin RMSE ve Yanlılık Değerleri (1000 Madde Havuz Büyüklüğü)

Madde Seçme Yöntemi	Yetenek Kestirim Yöntemi	Madde Kullanım Sıklığı Kontrol Yöntemi	İçerik Ağırlıklandırma Yok						İçerik Ağırlıklandırma Var					
			a		b		c		a		b		c	
			RMSE	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık
MFI	EAP	KY	3,649	-1,150	8,189	2,902	0,244	0,114	7,409	0,377	5,737	0,483	0,327	0,222
		MUİ	4,970	-1,731	9,918	2,797	0,284	0,120	4,713	-2,280	3,227	0,612	0,308	0,153
	MLE	KY	7,058	-2,006	8,487	-1,025	0,310	0,207	7,103	-0,064	5,032	-0,809	0,273	0,144
		MUİ	4,702	-0,939	8,465	-2,507	0,275	0,110	7,960	-0,918	4,571	0,760	0,321	0,154
MIS	EAP	KY	8,605	1,118	3,279	0,160	0,297	0,159	11,704	0,927	2,370	-0,099	0,346	0,201
		MUİ	8,229	0,362	4,606	-0,732	0,315	0,141	9,207	-0,473	5,598	0,464	0,259	0,122
	MLE	KY	9,446	1,038	4,366	1,193	0,304	0,140	9,139	-1,483	7,365	-2,577	0,300	0,156
		MUİ	8,763	-4,656	5,328	0,529	0,353	0,280	9,080	-1,510	7,361	-2,593	0,294	0,155
GEK	EAP	KY	7,361	-1,133	1,882	0,440	0,288	0,114	9,299	-1,232	1,841	-0,395	0,299	0,182
		MUİ	6,656	-0,306	1,851	0,081	0,247	0,068	8,935	0,449	3,087	0,340	0,311	0,163
	MLE	KY	6,113	-1,816	1,805	-0,300	0,276	0,139	12,407	0,534	2,431	-0,093	0,279	0,169
		MUİ	9,114	-1,014	1,652	-0,453	0,280	0,117	11,287	1,790	2,736	-0,654	0,275	0,146

EK-G: Simülasyon Geçerliği İçin Kestirimi Yapılan Yetenek Parametrelerinin RMSE ve Yanlılık Değerleri (320 Madde Havuzu

Büyüklüğü)

Madde Seçme Yöntemi	Yetenek Kestirim Yöntemi	Madde Kullanım Sıklığı Kontrol Yöntemi	İçerik Ağırlıklandırma Yok		İçerik Ağırlıklandırma Var	
			RMSE	Yanlılık	RMSE	Yanlılık
MFI	EAP	KY	0,934	0,205	1,318	0,004
		MUİ	1,343	-0,274	1,044	-0,253
	MLE	KY	1,479	0,043	1,517	-0,398
		MUİ	1,563	-0,313	1,568	-0,104
MIS	EAP	KY	1,040	-0,081	1,074	-0,127
		MUİ	1,220	-0,061	1,063	0,031
	MLE	KY	1,788	0,564	1,926	0,353
		MUİ	1,327	0,095	1,582	-0,393
GEK	EAP	KY	1,045	0,004	1,144	0,002
		MUİ	1,212	-0,478	1,144	0,002
	MLE	KY	1,500	0,013	1,582	-0,393
		MUİ	1,678	-0,459	1,582	-0,393

EK-Ğ: Simülasyon Geçerliği İçin Kestirimi Yapılan Yetenek Parametrelerinin RMSE ve Yanlılık Değerleri (1000 Madde Havuzu

Büyüklüğü)

Madde Seçme Yöntemi	Yetenek Kestirim Yöntemi	Madde Kullanım Sıklığı Kontrol Yöntemi	İçerik Ağırlıklandırma Yok		İçerik Ağırlıklandırma Var	
			RMSE	Yanlılık	RMSE	Yanlılık
MFI	EAP	KY	1,327	-0,160	1,247	-0,230
		MUİ	1,381	-0,126	1,259	-0,160
	MLE	KY	1,497	-0,202	1,832	0,156
		MUİ	1,635	-0,034	1,572	-0,118
MIS	EAP	KY	1,063	0,141	1,252	-0,329
		MUİ	1,433	-0,276	1,223	-0,380
	MLE	KY	1,703	-0,363	1,466	-0,532
		MUİ	1,532	-0,198	1,798	-0,356
GEK	EAP	KY	1,263	-0,084	1,212	-0,170
		MUİ	1,070	0,160	1,353	0,003
	MLE	KY	1,733	-0,413	1,305	-0,037
		MUİ	0,765	-0,079	1,440	0,201

EK-H: Araştırma Etik Komisyon İzin Muafiyeti Formu/ Araştırma Etik Komisyonu

Onay Bildirimi

	Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması/Araştırma Etik Komisyon İzin Muafiyeti Formu	F46
16 / 03 / 2023		
Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Ana Bilim Dalı Başkanlığına		
Tez/Araştırma Başlığı	BOBUT Uygulamalarında Farklı Koşullardan Elde Edilen Ölçme Kesinliği Kestirim Değerlerinin Karşılaştırılması	
Yukarıda başlığı/konusu verilen tez/araştırma çalışmam,		
<ol style="list-style-type: none"> 1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır. 2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir. 3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir. 4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırmalar niteliğinde değildir. 5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir. 		
Çalışmada kullanacağım veriler:		
<input type="checkbox"/> Kamusal erişime açık (buraya yazınız):		
<input type="checkbox"/> Özel izin ve onaya tabi (buraya yazınız):		
<input checked="" type="checkbox"/> Üretilmiş veri (buraya yazınız): RStudio yazılımında bulunan catR paketi kullanılarak elde edilen veri setinde çalışılıyor.		
<input type="checkbox"/> Diğer (buraya yazınız):		
Yükseköğretim Kurumları Etik Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.		
Gereğini saygılarımla arz ederim.		
Melike ÖZER TAYMUR <small>(Araştırmacı Adı Soyadı, İmzası)</small>		
Araştırmacı Bilgileri		
Adı Soyadı	Melike ÖZER TAYMUR	
Öğrenci ise No	N13249842	
Ana Bilim Dalı	Eğitim Bilimleri	
Programı	Eğitimde Ölçme ve Değerlendirme	
Statüsü	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr. <input type="checkbox"/> Diğer	
Danışman Görüşü ve Onayı*		
Prof. Dr. Duygu ANIL <small>(İmza)</small> <small>(Danışmanın unvanı, Adı ve Soyadı)</small>		
*Tez ve tezden üretilen yayınlarda gerekli		
Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü. Beytepe Yerleşkesi, 06800, Çankaya / ANKARA Telefon: 0(312) 297 85 72 Belgegeçer: 0(312) 297 85 66 e-Ağ: http://ebe.hacettepe.edu.tr/ e-Posta: ebe@hacettepe.edu.tr		

EK-I: Etik Beyanı

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- * tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- * görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- * başkalarının eserlerinden yararlanması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- * atıfta bulunduğum eserlerin bütününe kaynak olarak gösterdiğimi,
- * kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- * bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

25/07/2023

Melike ÖZER TAYMUR

EK-İ: Doktora Tez Çalışması Orijinallik Raporu

HACETTEPE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

...../...../.....

Tez Başlığı: BOBUT Uygulamalarında Farklı Koşullardan Elde Edilen Ölçme Kesinliği Kestirim Değerlerinin Karşılaştırılması

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
17/07/2023	134	138091	23/06/2023	%11	2132549876

Uygulanan filtreler:

- Kaynaklar hariç
- Alıntılar dâhil
- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esaslarını inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Melike ÖZER TAYMUR

Öğrenci No.: N13249842

Ana Bilim Dalı: Eğitim Bilimleri

İmza

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

Prof. Dr. Duygu ANIL

EK-J: Dissertation Originality Report

HACETTEPE UNIVERSITY
Graduate School of Educational Sciences
To The Department of Educational Sciences

...../...../.....

Thesis Title: Comparison of Measurement Precision Values Obtained From Different Conditions in Computerized Adaptive Testing

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
17/07/2023	134	138091	23/06/2023	%11	2132549876

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Name Lastname: Melike ÖZER TAYMUR

Student No.: N13249842

Department: Educational Sciences

Program: Measurement and Evaluation in Education

Status: Masters Ph.D. Integrated Ph.D.

Signature

ADVISOR APPROVAL

APPROVED
Prof. Dr. Duygu ANIL

EK-K: Yayınlanma ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezimin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına ilişkin Yönerge**" kapsamında tezimin aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü/ Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren ... ay ertelenmiştir. ⁽²⁾
- Teziminle ilgili gizlilik kararı verilmiştir. ⁽³⁾

25/07/2023

(imza)

Melike ÖZER TAYMUR

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına ilişkin Yönerge"

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6.2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3 şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlerle ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
- Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir
- *Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

