



HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

İNGİLİZCE SEVİYE BELİRLEME SINAV SONUÇLARI ÜZERİNDE BİLGİSAYARDA
BİREYSELLEŞTİRİLMİŞ SINIFLAMA TESTİ YAKLAŞIMININ UYGULANMASI

DEMET ALKAN

Doktora Tezi

Ankara, 2023

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

Daha ileriye... En İyiyeye...



HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

İNGİLİZCE SEVİYE BELİRLEME SINAV SONUÇLARI ÜZERİNDE BİLGİSAYARDA
BİREYSELLEŞTİRİLMİŞ SINIFLAMA TESTİ YAKLAŞIMININ UYGULANMASI

APPLICATION OF COMPUTERIZED ADAPTIVE CLASSIFICATION TEST APPROACH ON
ENGLISH PLACEMENT TEST RESULTS

DEMET ALKAN

Doktora Tezi

Ankara, 2023

Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Demet ALKAN'ın hazırladıđı “İngilizce Seviye Belirleme Sınav Sonuçları Üzerinde Bilgisayarda Bireyselleştirilmiş Sınıflama Testi Yaklaşımının Uygulanması” başlıklı bu çalışma j¼rimiz tarafından Eđitim Bilimleri Ana Bilim Dalı, **Eđitimde Ölçme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiştir.

J¼ri Başkanı	Prof. Dr. Selahattin GELBAL	İmza
J¼ri Üyesi (Danışman)	Prof. Dr. Nuri DOĐAN	İmza
J¼ri Üyesi	Doç. Dr. Levent YAKAR	İmza
J¼ri Üyesi	Doç.Dr. Celal Deha DOĐAN	İmza
J¼ri Üyesi	Doç. Dr. Sevda ÇETİN	İmza

Bu tez Hacettepe Üniversitesi Lisansüstü Eđitim, Öğretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından / / tarihinde uygun gör¼lmüş ve Enstitü Yönetim Kurulunca / / tarihi itibarıyla kabul edilmiştir.

Prof. Dr. İsmail Hakkı MİRİCİ
Eđitim Bilimleri Enstitüsü M¼d¼r¼

Öz

Bu arařtırmada Bilkent Üniversitesi Yabancı Diller Yüksekokulunda öğrenim gören 919 bireye uygulanan İngilizce seviye belirleme sınavına ait tek boyutlu 256 okuma maddelerinin cevap örüntüleri üzerinde sınıflamanın iki, üç ve dört kategoride yapıldığı durumlarda BBST yaklaşımının uygulanabilirliğinin araştırılması amaçlanmıştır. Arařtırmada yetenek kestirim yöntemi olarak Maksimum Olabilirlik Kestirimi (MOK), Ağırlıklandırılmış Olabilirlik Kestirimi (AOK), Beklenen Sonsal Dağılım (BSD) yöntemleri kullanıldığında, ortalama test uzunluğu (OTU), ortalama sınıflama doğruluğu (OSD), gerçek yeteneklerle kestirilen yetenekler arası korelasyon (r), kestirimin standart hatasını gösteren RMSE, ortalama mutlak hata (OMH) ve yanlılık değerlerinin Ardışık olasılık Oran Testi (AOOT), Genelleştirilmiş Olabilirlik Oran (GOO) ve Güven Aralığı (GA) sınıflama kriterleri ile Kestirilen Yetenek temelli Maksimum Fisher Bilgisi (MFB-KY) ve Kullback-Leibler Bilgisi (KLB-KY), Kesme Noktası Temelli Maksimum Fisher Bilgisi (MFB-KN) ve Kullback-Leibler Bilgisi (KLB-KN) madde seçme yöntemlerine göre iki, üç, dört kategorili sınıflamada nasıl değiştiği incelenmiştir. Arařtırmada R yazılımı kullanılarak gerçek veriye dayanan simülasyon için toplam 108 koşul oluşturulmuştur. Arařtırma sonunda iki, üç ve dört kategorili sınıflamalarda test etkililiği (OTU, OSD) için GA sınıflama kriterinin, ölçme kesinliği (korelasyon, yanlılık, RMSE, OMH) için AOOT sınıflama kriterinin yüksek performans gösterdiği sonucuna ulařılmıştır. GOO sınıflama kriterinin dört kategorili sınıflamada MOK ve BSD yetenek kestirimleri ve MFB madde seçme yöntemleriyle test etkililiği için başarılı olduğu anlaşılmıştır.

Anahtar sözcükler: bilgisayarda bireyselleştirilmiş sınıflama testi, sınıflama kategori sayısı, sınıflama kriterleri, madde seçme yöntemleri, yetenek kestirim yöntemleri

Abstract

In this study, it was aimed to investigate the applicability of the CACT approach when the classification is made in two, three, or four categories on the response patterns of 256 one-dimensional reading items of the English placement test administered to 919 individuals studying at Bilkent University School of Foreign Languages. In the study, when Maximum Likelihood Estimation (MLE), Weighted Likelihood Estimation (WLE), Expected a Posteriori (EAP) methods were used as ability estimation methods, average test length (ATL), average classification accuracy (ACA), correlation between real and estimated thetas (r), Root Mean Square Error (RMSE) indicating the standard error of estimation, mean absolute error (MAE) and bias values of Sequential Probability Ratio Test (SPRT), Generalized Likelihood Ratio (GLR) and Confidence Interval (CI) classification criteria and how they change in two- three, and four-category classification according to Estimated Ability-based Maximum Fisher Information (MFI-EB) and Kullback-Leibler Information (KLI-EB), cut-point based Maximum Fisher Information (MFI-CB) and Kullback-Leibler Information (KLI-CB) item selection methods were examined. In the study, a total of 108 conditions were created for the simulation based on real data using R software. At the end of the study, it was concluded that the CI classification criterion showed high performance for test effectiveness (ATL, ACA), and the WLE classification criterion showed high performance for measurement precision (correlation, bias, RMSE, MAE) in two, three and four category classifications. The GLR classification criterion was found to be successful for test effectiveness in four-category classification with MLE and EAP ability estimations and MFI item selection methods.

Keywords: ability estimation methods, classification criteria, computerized adaptive classification test, number of classification categories, item selection methods

Teşekkür

Doktora sürecimin her aşamasında desteğini hiç esirgemeyen, tüm süreç boyunca en büyük motivasyonla ilerlememi sağlayan, yol gösteren değerli danışmanım Prof.Dr. Nuri DOĞAN'a sonsuz saygılarımla teşekkürlerimi sunarım.

Doktora sürecimin ilk gününden itibaren varlığını hep hissettiğim değerli hocam Prof.Dr. Selahattin GELBAL'a en büyük saygılarımla teşekkürlerimi sunarım.

Tez jürimde yer alan ve tezim için beni Tübitak projesine dahil eden değerli hocam Doç.Dr. Levent YAKAR'a ve Tübitak'a saygılarımla teşekkürlerimi sunarım.

Tez Jürimde yer alan Doç Dr. Deha DOĞAN ve Doç Dr.Sevda ÇETİN hocalarıma tezimin gelişimine sundukları katkılar için sonsuz teşekkür ederim.

Tez sürecimde aradığım her zaman yardım eden, arkadaşım Dr. Mahmut Sami YİĞİTER'e çok teşekkür ediyorum. Desteğini her konuda esirgemeyen arkadaşım Dr.Güzide ŞEN'e, veri toplama sürecindeki desteği ile arkadaşım Dr. Sebahat GÖREN'e çok teşekkür ederim.

Bugünlere gelmemdeki en büyük fedakarlık Annem ve Babamın. Evlatları olmaktan gurur duyuyorum. Köyünün ilk okuyan kızı ve bana da okumanın güzelliklerini gösteren annem Pakize ALKAN, merhametini örnek aldığım, bilgili olmanın, saygının, sevginin, güçlü olmanın bendeki en güzel örneği babam Haydar ALKAN'a en büyük sevgilerimle teşekkür ederim.

İçindekiler

Kabul ve Onay.....	ii
Öz	iii
Abstract	iv
Teşekkür	v
Tablolar Dizini	viii
Şekiller Dizini	ix
Simgeler ve Kısaltmalar	x
Bölüm 1 Giriş	1
Problem Durumu	1
Araştırmanın Amacı ve Önemi.....	4
Araştırma Problemi	7
Sınırlılıklar	8
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar	9
Bilgisayarlı Bireyselleştirilmiş Sınıflama Testi	9
Psikometrik Model	12
Madde Havuzu	20
Başlama Noktası	21
Madde Seçimi	21
YeteneK Kestirimi	23
Sınıflama Kriterleri	26
İlgili Araştırmalar	32
Bölüm 3 Yöntem	42
Araştırma Yöntemi	42
Verilerin Elde Edilmesi	42
BBST simülasyon Koşulları	52
Verilerin Analizi	52

Bölüm 4 Bulgular, Yorumlar ve Tartışma	54
Birinci Alt Probleme Ait Bulgu Ve Yorumlar	54
İkinci Alt Probleme Ait Bulgu Ve Yorumlar	67
Üçüncü Alt Probleme Ait Bulgu Ve Yorumlar	78
Dördüncü Alt Probleme Ait Bulgu Ve Yorumlar	87
Beşinci Alt Probleme Ait Bulgu Ve Yorumlar	89
Altıncı Alt Probleme Ait Bulgu Ve Yorumlar	91
Yedinci Alt Probleme Ait Bulgu Ve Yorumlar	93
Sekizinci alt Probleme Ait Bulgu Ve Yorumlar	94
Bölüm 5 Sonuç ve Öneriler	95
Sonuçlar	95
Öneriler	99
Kaynaklar	103
EK-A: Etik Komisyonu Onay Bildirimi	114
EK-B: Veri Kullanım İzin Belgesi.....	116
EK-C: Etik komisyon İzni	117
EK-Ç: Etik Beyanı.....	118
EK-D: Yüksek Lisans/Doktora Tez Çalışması Orjinallik Raporu	119
EK-E: Thesis/Dissertation Originality Report	120
Ek-F: Yayınlanma ve Fikri Mülkiyet Hakları Beyanı	121

Tablolar Dizini

Tablo 1 Araştırmada incelenen değişkenler	5
Tablo 2 Bilgisayarda bireyselleştirilmiş sınıflama testi algoritması	11
Tablo 3 Veri setlerinde bulunan madde sayıları	43
Tablo 4 Eksik veri matrisi	46
Tablo 5 1PLM ile 2PLM'nin model veri uyumu için karşılaştırılması	50
Tablo 6 2PLM ile 3PLM'nin model veri uyumunun karşılaştırılması	51
Tablo 7 Yetenek kestirim yöntemi MOK olduğunda iki, üç, dört kategorili sınıflamaya ait test etkililiği ve ölçme kesinliği değerleri	55
Tablo 8 Yetenek kestirim yöntemi AOK olduğunda iki, üç, dört kategorili sınıflamaya ait test etkililiği ve ölçme kesinliği değerleri	68
Tablo 9 Yetenek kestirim yöntemi BSD olduğunda iki, üç, dört kategorili sınıflamaya ait test etkililiği ve ölçme kesinliği değerleri	79
Tablo 10 Madde seçme yöntemlerine göre iki, üç, dört kategorili sınıflamaya ait test etkililiği ve ölçme kesinliği değerleri	88
Tablo 11 Yetenek kestirimlerine göre test etkililiği ve ölçme kesinliği değerleri	90
Tablo 12 Sınıflama kriterlerine göre test etkililiği ve ölçme kesinliği değerleri	91
Tablo 13 Sınıflama kategori sayısına göre test etkililiği ve ölçme kesinliği değerleri	93

Şekiller Dizini

Şekil 1 <i>1PLM için madde karakteristik eğrisi</i>	16
Şekil 2 <i>2PLM için madde karakteristik eğrisi</i>	17
Şekil 3 <i>3PLM için madde karakteristik eğrisi</i>	19
Şekil 4 <i>Veri setine ait yamaç birikinti grafiği</i>	48

Simgeler ve Kısaltmalar Dizini

AOK: Ağırlıklandırılmış Olabilirlik Kestirimi (WLE: Weighted Likelihood Estimation)

AOOT: Ardışık Olasılık Oran Testi (SPRT: Sequential Probability Ratio Test)

BBST: Bilgisayarda Bireyselleştirilmiş Sınıflama Testi (CACT: Computerized Adaptive Classification Test)

BSD: Beklenen Sonsal Dağılım (EAP: Expected a Posteriori)

FB: Farksızlık bölgesi (IR: Indifference region)

GA: Güven Aralığı (CI: Confidence Interval)

MFB: Maksimum Fişher Bilgisi (MFI: Maximum Fisher Information)

MFB-KN: Kesme Noktası Temelli Maksimum Fisher Bilgisi (MFI-CB: Cut-Score Based Maximum Fisher Information)

MFB-KY: Kestirilen Yetenek Temelli Maksimum Fisher Bilgisi (MFI-EB: Ability-based Maximum Fisher Information)

MOK: Maksimum Olabilirlik Kestirimi (MLE: Maximum Likelihood Estimation)

MTK: Madde Tepki Kuramı (Item Response Theory)

OMH: Ortalama Mutlak Hata (MAE: Mean Absolute Error)

OSD: Ortalama Sınıflama Doğruluğu (ACA: Average Classification Accuracy)

OTU: Ortalama Test Uzunluğu (ATL: Average Test Length)

KLB: Kulback Laiber Bilgisi (KLI: Kullback-Leibler Information)

KLB-KN: Kesme Noktası Temelli Kulback Laiber Bilgisi (KLI-CB: Cut-Score Based Kullback-Leibler Information)

KLB-KY: Kestirilen Yetenek Temelli Kulback Laiber Bilgisi (KLI-EB: Ability-based Kullback-Leibler Information)

Bölüm 1

Giriş

Birinci bölümde araştırmanın problem durumu, amacı, önemi, problem cümlesi, alt problemleri, sayıtlarına ve sınırlılıklarına yer verilmiştir.

Problem Durumu

Test uygulamaları, sınava giren kişinin sınavın amacıyla ilişkili olduğu düşünülen yetenek seviyesine ulaştığına dair kanıt sağlamak için kullanılır. Testler sayesinde çoğu zaman kişinin belirli bir özellik için tahmini yetenek düzeyini gösteren bir puan elde edilir. Çoğu zaman bir testten alınan puan, bireyi iki veya daha fazla kategoriye ayırmak için kullanılır. Örneğin, bir akademik başarı değerlendirmesinden elde edilen ölçek puanı, bir öğrencinin bir dersten kredi alıp almayacağını veya bir sonraki sınıf seviyesine geçip geçemeyeceğini belirlemek için kullanılabilir. Birçok durumda sınav katılımcılarını ikiden fazla kategoriden birine (örneğin ortalamanın altında / ortalama / ortalamanın üstünde) sınıflandırmak, basit iki kategoride (örneğin başarılı / başarısız) sınıflandırmaya tercih edilir.

Belli bir özelliğe göre bireyleri sınıflama amacı olduğunda ve daha az sayıda madde ile yüksek doğrulukta sınıflama yapmak için Bilgisayarda Bireyselleştirilmiş Sınıflama Testi (BBST) kullanılabilir. BBST bireylerin yeteneğine uygun ve az sayıda madde ile ölçüm yaptığı için daha yüksek derecede ölçüm hassasiyeti içerir (Davey & Pitoniak, 2006; Wainer ve diğerleri, 2000; Way ve diğerleri, 2006). Kağıt kalem testleri Klasik Test Teorisine göre geliştirildiğinde ortalama yetenek etrafında bir yetenek aralığını ölçmek için tasarlanır. Klasik test teorisine göre geliştirilen testler yetenek seviyeleri ölçülen yetenek aralığından sapan katılımcılara uygulandığında yetersiz ölçüm yapılabilir (Weiss, 2014). Bilgisayarda Bireyselleştirilmiş Testler (BBT) genellikle geniş madde havuzu ile geniş bir yetenek sürekliliğini kapsar ve bireyin yeteneğine en uygun maddeyi belirlemek için tasarlanmıştır. BBT cevaplanan her maddeden sonra bireyin yeteneğini geçici yetenek düzeyinde en fazla bilgi içeren maddelerin seçimi ile kestirdiği için bireylerin yetenek seviyesini gerçek seviyelerinden daha düşük veya daha yüksek kestirebilmektedir. Örneğin şansla bir

maddenin doğru cevaplanması ya da yanlış cevaplanması gibi gerçek yetenek düzeyini ketirmek zor olabilir. BBST'de kestirilen yetenek dışında kesme noktasında da en çok bilgiyi veren maddenin seçilmesi nedeniyle birey sınıflandığında yeteneği de kestirilmiş olur. BBST özellikle sınava giren bireylerin sınıflanması için tasarlanmış değişken uzunluklu bir testtir. Değişken uzunluklu testlerde her birey aynı uzunlukta test almaz. Bunun yerine parametreleri önceden kestirilmiş maddeler madde havuzunda bulunur. Belirli bir başlatma kuralı ile test başlatılır. Bireyler maddelere cevap verdikçe kestirilen geçici yetenek düzeyine uygun maddeler veya kesme noktasında en çok bilgi veren maddeler madde havuzundan seçilir böylece sınav bireyselleştirilmiş olur. Belirli bir kritere ulaşıldıktan sonra ya da sınıflama gerçekleştirildikten sonra test sonlandırılır. Böylece daha az sayıda madde ile bireyler kolayca sınıflandırılabilir. Bu da hem zamandan hem de çabadan tasarruf sağlar. Bilgisayar, belirli bir süre içinde daha fazla sınav katılımcısının test edilmesine olanak tanır. Bu fayda, binlerce veya yüzlerce kişi için düşünüldüğünde önemli mali sonuçlar doğurmaktadır.

Doğası gereği sınıflandırma kararları yüksek risk taşır çünkü yanlış sınıflandırma bireylerin hak ettikleri eğitim veya sağlık gibi fırsatları kaçırmalarına neden olabilir. Sınıflandırma kararlarının bireylerin yaşamları üzerindeki etkisi olabileceğinden test geliştiricilerin görevi sınıflama doğruluğuna ilişkin kanıt sağlamaktır (Lathrop, 2015). Özellikle kesme noktalarındaki sınıflandırma performansı daha büyük önem taşıyabilir çünkü yanlış sınıflandırma bireyleri yanlış yeterlilik seviyelerine yerleştirmek gibi yanlış kararlar alınmasına neden olur. Özellikle çok kategorili sınıflama için sabit test formu yerine BBST kullanmak kesme noktası ve kestirilen yetenek temelinde yüksek bilgi veren maddeleri seçerek daha doğru sınıflamaya ulaşılabilceğini göstermiştir (Lewis & Sheehan, 1990). Çok kategorili sınıflama, iki kategorili sınıflamaya göre daha geniş bir yetenek yelpazesinde bilgi gerektirdiği için sınava girenler tek bir puan yerine iki veya daha fazla kesme puanına göre değerlendirilir.

Bir testin iki ya da üç kesme puanıyla üç ya da dört kategoride sınıflandırması, sınava giren tüm öğrencilerin sınıflandırılması için ihtiyaç duyulan ortalama madde sayısını artırır ve bu da maddelere aşırı maruz kalmadan etkili bir test uygulaması için bir madde havuzunda ihtiyaç duyulan madde sayısını artırır (Spray, 1993). BBST'de sınıflama için gereken madde sayısının az olması ve ortalama sınıflama doğruluğunun yüksek olmasıyla testin etkililiği artarken düşük standart hatalar, yanlılık, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyonun yüksek olması ölçme kesinliğini yükseltir (Thompson, 2009). BBST araştırmasının genel amacı BBST 'nin verimliliğini en üst düzeye çıkararak belirli yöntemleri belirlemek ve olabildiğince az madde kullanarak sınıflama hatalarını düşürerek testin etkililiğini ve ölçme kesinliğini sağlamaktır (Thompson, 2009).

Alanyazın incelendiğinde BBST performansı üzerine yapılan çalışmaların büyük çoğunluğunun simülasyonla oluşturulmuş verilere dayanan Monte-Carlo çalışmaları olduğu görülmektedir (Spray & Reckase, 1996). Ancak oluşturulmuş veri setlerine dayanan simülatif çalışmalar tahmin etme, hızlılık, yorgunluk gibi gerçek sınav katılımcısı davranışlarını temsil etmekte başarısız olabilir (Thompson & Weiss, 2011; Wang ve diğerleri, 1999). Ayrıca alanyazında daha çok iki kategorili sınıflamaya ait çalışmalara rastlanmaktadır (Gündeğer, 2017; Lau, 1996; Reckase,1983; Thompson, 2011). İki kategorili sınıflamalar kadar çok kategorili sınıflamalara da ihtiyaç duyulmaktadır. Çok kategorili sınıflama için Nydick vd. (2012), çok boyutlu veri seti üzerinde üç ve beş kategorili sınıflama çalışması, Eggen ve Straetmans (2000), tarafından pratik kısıtlamalar altında üç kategoride BBST uygulaması çalışılmıştır. Çok kategorili sınıflamalara ilişkin çalışmalar az bulunmaktadır. Çok kategorili sınıflamalarda doğru sınıflama düzeyini yükseltmek için gerekli koşulların neler olduğu net değildir. Özellikle gerçek veri seti ile çok kategorili sınıflama için BBST uygulamasına ait çok sayıda yöntemin çaprazlanarak performansının incelendiği bir araştırma örneğinin bulunmaması literatürde boşluk oluşturmaktadır.

Gerçek veri seti üzerinde BBST' nin çok kategorili sınıflamada uygulanabilirliğinin araştırıldığı çalışmaya ihtiyaç olduğu görülmüştür. Bu çalışmada İngilizce seviye belirleme

sınavından elde edilen gerçek veri seti ile çok kategorili BBST uygulaması için tüm yöntemler çaprazlanarak birlikte kullanıldığında sınıflama doğruluğu, ortalama test uzunluğu, yanlılık, gerçek yetenekler ile kestirilen yetenekler arasındaki korelasyon (r), RMSE ve ortalama mutlak hata (OMH) üzerindeki etkileri 25 tekrarın ortalaması ile hesaplanarak sınıflama koşullarının performansları belirlenmiştir. Sınıflama doğruluğu endeksi ile ölçümü nicelleştirerek sınıflandırmanın kalitesi ölçülmüştür. Bu endeks ile sınav katılımcılarının gerçek sınıflandırmalarının gözlenen sınıflandırmalarla ne ölçüde uyumlu olduğu farklı bir ifade ile sınava girenlerin sınıflandırılma performansı belirlenmiştir (Diao & Sireci, 2018). İki, üç ve dört kategorili sınıflamada gerçek BBST uygulaması için test etkililiği ve ölçme kesinliği en yüksek olan gerçek veri seti ile çok sayıda sınıflama koşullarının çaprazlandığı en uygun desenler belirlenmiştir.

BBST uygulamasında pratik kısıtlamalar adıyla içerik dengeleme ve madde kullanım sıklığı kontrolü gibi ölçmenin geçerliliğini yükselttiği düşünülen yöntemler de bulunmaktadır. BBST uygulamalarında madde kullanım sıklığı kontrol yöntemleri ve içerik dengeleme gibi pratik kısıtlamalar sorununda geçmişteki araştırmalardan kısıtlamaların getirilmesinin gerekli olmadığı özellikle simülasyon çalışmalarında pratik kısıtlamaların getirilmesinin araştırma için zararlı olduğu düşünülmektedir (Thompson, 2007). Bu nedenle ve madde havuzu geniş olduğu için maddeleri hatırlamaktan kaynaklı geçerliliği etkileyecek durum olmayacağı düşünüldüğünden araştırmada pratik kısıtlamalar kullanılmamıştır.

Araştırmanın Amacı ve Önemi

Bu araştırmanın amacı İngilizce seviye belirleme sınavına ait gerçek veri setinden oluşan geniş madde havuzu üzerinde iki, üç ve dört kategorili sınıflama için BBST'nin uygulanabilirliğini araştırmaktır. Bu amaçla Bilkent Üniversitesi Yabancı Diller Yüksekokulundan 919 bireye uygulanan 256 okuma maddelerinden oluşan İngilizce seviye belirleme sınavına ait tek boyutlu gerçek veri seti kullanılmıştır. İki, üç ve dört kategorili Bilgisayarda Bireyselleştirilmiş Sınıflama Testi (BBST) uygulandığında; testin uzunluğunu gösteren (OTU), sınıflamanın doğruluğunu gösteren (OSD) ve kestirimin standart hatasını

gösteren RMSE, ortalama mutlak hata (OMH), gerçek yeteneklerle BBST sonucunda kestirilen yetenekler arası korelasyon (r) ve yanlılığı; sınıflama kriterlerinin Ardışık Olasılık Oran Testi (AOOT), Güven Aralığı (GA), Genelleştirilmiş Olabilirlik Oran (GOO); madde seçme yöntemlerinin Kestirilen Yetenek ve Kesme Noktası Temelli Maksimum Fisher Bilgisi (MFB) ve Kulback Leibler Bilgisi (KLB); yetenek kestirim yöntemlerinin Ağırlıklandırılmış Olabilirlik Kestirimi (AOK), Beklenen Sonsal Dağılım (BSD), Maksimum Olabilirlik Kestirimi (MOK) olduğu durumlarda nasıl değiştiği araştırılmıştır. Tüm yöntemlerin çaprazlanarak oluşturduğu koşullarda performanslarını çok kategorili sınıflama için gerçek veri ile belirlemek ve ölçme kesinliği ile test etkililiğinin yöntemlere göre nasıl değiştiğini incelemek amaçlanmıştır. Araştırmada 108 koşul belirlenmiştir. Oluşturulan 108 durum için en uygun sınıflama koşuluna ait sınıflama desenini 25 tekrarın ortalaması alınarak belirlemek amaçlanmıştır. Bu koşulları oluşturan değişkenler Tablo 1 de gösterilmektedir.

Tablo 1

Araştırmada İncelenen Değişkenler

Veri seti	Sınıflama Kategori Sayısı	Sınıflama Kriterleri	Yetenek Kestirimleri	Madde Seçme Yöntemleri
	İki	AOOT (FB: 0.1)	BSD	MFB-KY
Gerçek veri seti	Üç	GA (%90)	MOK	MFB-KN
	Dört	GOO (FB:0.1)	AOK	KLB-KY KLB-KN

Tablo 1 de üç sınıflama kategori sayısı, üç sınıflama kriteri, üç yetenek kestirim yöntemi, dört madde seçme yönteminin gerçek veri seti ile kullanıldığı görülmektedir. Üç sınıflama kategori sayısı x üç yetenek kestirimi x üç sınıflama kriteri x dört madde seçme yöntemi = 108 tane koşul oluşturulmuştur.

Uluslararası geniş ölçekli sınavlarda PISA, TIMSS gibi bireyleri çok kategorili sınıflandırmanın amaçlandığı görülmektedir. Özellikle yabancı dil sınavları bireyleri dil

becerilerine göre sınıflandırmaktadır. Dil sınavları ve uluslararası sınavlar bu anlamda BBST ile benzer özelliklere sahiptir. Bu nedenle öncelikle dil sınavlarından başlanarak BBST uygulamaları kullanılırsa her bireye uygulanacak madde bireyin daha önce uygulanan maddelere verdiği cevaplar ile kestirilen yetenek düzeyine göre seçilecek ve bireyin yetenek düzeyine uygun maddeler uygulanacağı için sınav bireyselleştirilecektir. Böylece daha az sayıda madde ile daha kısa sürede bireylerin dil seviyeleri hakkında daha doğru kararlar verilebilecektir. BBST için çok kategorili sınıflamada en uygun desenin belirlenmesi PISA, TIMSS gibi uluslararası sınavlarda da gerçek uygulamalara geçiş süreci için önemli olduğu düşünülmektedir. Uluslararası geniş ölçekli sınavların BBST ile uygulanması binlerce sınav katılımcısını kapsadığı için önemli mali sonuçlar doğurmaktadır.

Alanyazında özellikle yurt içinde çoğunlukla BBT için araştırmalar yapılmıştır (Arzu & Doğan, 2021; Boztunç-Öztürk & Doğan, 2015; Eroğlu & Kelecioğlu, 2015; Gelbal & Şahin, 2020). BBST'ye ilişkin yurtiçinde Demir (2019); Gündeğer (2017) olmak üzere az sayıda çalışma bulunmaktadır. Yurt dışında ise yurt içine göre daha fazla sayıda çalışma örneklerine rastlansa da genelde simülatif veri ile yapılan çalışmalardır (Lau, 1996; Kingsbury & Weiss, 1980; Spray & Reckase, 1996). Yurt dışındaki literatür incelendiğinde sınıflama kriterlerinden AOOT ve BUT sınıflama kriterlerinin karşılaştırıldığı çalışmalara (Kingsbury & Weiss, 1980; Nydick, 2013; Thompson, 2011), sınıflamanın iki kategoride yapıldığı çalışmalara (Lau, 1996; Reckase, 1983; Spray & Reckase, 1996) ve madde seçme yöntemlerinin simülatif veri ile performanslarının karşılaştırıldığı çalışmalara (Eggen, 1999; Lin & Spray, 2000) rastlanmaktadır.

Alanyazında gerçek veri seti ve geniş madde havuzu ile çok sayıda yetenek kestirim yöntemleri madde seçme yöntemleri ve sınıflama kriterlerinin çaprazlandığı geniş koşullara ait çok kategorili sınıflamaların ölçme kesinliği ve test etkinliği açısından incelendiği, yöntemlerin performanslarının araştırıldığı gerçek uygulama için oluşturulan koşullara ait en uygun desenlerin belirlendiği geniş çaplı çalışma örneğinin ulusal ve uluslararası düzeyde BBST uygulayıcılarına ve literatüre katkı sağlayacağı ve literatürdeki boşluğu kapatacağı

nedeniyle önemli olduğu düşünülmektedir. Literatürde özellikle GOO sınıflama kriterinin, MOK, BSD yetenek kestirimlerinin tüm madde seçme yöntemleriyle çaprazlandığı performanslarının çok kategorili BBST uygulamasında nasıl olduğu net değildir. Bu çalışmanın çok kategorili sınıflama yapması, gerçek veri seti kullanılması ve BBST uygulaması için önceki çalışmalarda performansları araştırılmayan yöntemlere diğer yöntemlerle çaprazlanarak yer verilmiş olması nedeniyle BBST uygulaması için faydalı olacağı düşünülmektedir. Özellikle uluslararası geniş ölçekli sınavlarda ve yurt içinde iki ve çok kategorili sınıflama gereken tüm sınavlarda BBST uygulamalarının ilerleyen teknoloji ile kullanılmaya başlanılacağı, bireyleri seçme ve yerleştirmede önemli olduğu düşünüldüğünde en doğru BBST desenlerinin belirlenmesi gerekmektedir. Bu anlamda iki, üç ve dört kategoride yapılan BBST çalışmaları ile araştırmanın gerçek veri seti ve geniş madde havuzu üzerinde çalışılmış olması ve birçok yöntemin çaprazlandığı geniş koşulların araştırılmasının alan literatürü için de faydalı olacağı düşünülmektedir.

Araştırma Problemi

Çoklu kesme puanlı BBST uygulamalarında tek boyutlu, iki kategorili maddelerden oluşan dil testine ait gerçek veri seti sonuçlarında test etkililiği ve ölçme kesinliği farklı yetenek kestirim yöntemlerine, farklı sınıflama kategori sayılarına ve sınıflama kriterleri ile madde seçme yöntemlerine göre, kullanılan MTK modeli altında nasıl değişmektedir?

Alt Problemler

1. BBST uygulamasında MOK yetenek kestirim yönteminde, MFB (KY-KN) ve KLB (KY-KN) madde seçme yöntemleri ile iki, üç ve dört kategorili sınıflamada GOO (FB:0.1), GA (%90), AOOT (FB:0.1) sınıflama kriterlerine göre OTU, OSD ve ölçme kesinliği nasıl değişmektedir?

2. BBST uygulamasında AOK yetenek kestirim yönteminde, MFB (KN-KY) ve KLB (KY-KN) madde seçme yöntemleri ile iki, üç ve dört kategorili sınıflamada GOO (FB:0.1), GA(%90), AOOT (FB:0.1) sınıflama kriterlerine göre OTU, OSD ve ölçme kesinliği nasıl değişmektedir?

3. BBST uygulamasında BSD yetenek kestirim yönteminde, MFB (KY-KN) ve KLB (KY-KN) madde seçme yöntemleri ile iki, üç ve dört kategorili sınıflamada GOO (FB:0.1), GA (%90), AOOT (FB:0.1) sınıflama kriterlerine göre OTU, OSD ve ölçme kesinliği nasıl değişmektedir?

4. Test uzunluğu, sınıflama doğruluğu ve ölçme kesinliği sınıflama kriterlerine AOOT (FB:0.1), GOO (FB:0.1), GA(%90) göre nasıl değişmektedir?

5. Test uzunluğu, Sınıflamanın doğruluğu ve ölçmenin kesinliği Yetenek Kestirim yöntemlerine (AOK, MOK, BSD) göre nasıl değişmektedir?

6. Test uzunluğu, sınıflama doğruluğu ve ölçme kesinliği kestirilen yetenek ve kesme noktası temelinde madde seçme yöntemlerine (MFB ve KLB) göre nasıl değişmektedir?

7. Test uzunluğu, sınıflama doğruluğu ve ölçme kesinliği sınıflama kategori sayısına (İki, üç ve dört kategori) göre nasıl değişmektedir?

8. Tüm analiz sonuçlarına göre en uygun BBST' i algoritması nasıldır?

Sınırlılıklar

1.Araştırmada İngilizce seviye belirleme sınavına ait tek boyutlu veri setinde a parametrelerinde yaklaşık 40 maddedeki yakınsama problemi nedeniyle 3PLM'ye göre madde ve yetenek parametreleri kestirilememiştir. Bu nedenle araştırmada yetenek kestirimleri için 2PLM kullanılmıştır.

2.Ortak köklü maddeler nedeni ile madde çiftleri arasındaki düşük düzeyli ilişki gözardı edilmiştir.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

İkinci bölümde BBST'i uygulamasının kuramsal temelini oluşturan psikometrik model olarak Madde Tepki Kuramı (MTK), Madde havuzu, Başlama noktası, Madde seçimi, Yetenek kestirimi ve Sınıflama kriterleri ile ilgili bilgiler verilmiştir. Sonrasında ilgili araştırmalara yer verilmiştir.

Bilgisayarda Bireyselleştirilmiş Sınıflama Testi (BBST)

Başarı ve yetenek testleri sınava katılan bireylerin yeteneklerini bir noktada belirlemek için geliştirilmiş testlerdir. Sınıflama testi ise bireyleri belli bir özelliğe göre sınıflandırma amacı olduğunda kullanılır. Bilgisayar tabanlı testler sonuçları yüksek önem arz eden sınıflandırma kararları vermek için kağıt kalem testlerine oranla tercih edilebilir. Bilgisayarda Bireyselleştirilmiş Sınıflama testi (BBST) özellikle sınava girenlerin sınıflandırılması için tasarlanmış değişken uzunlukta bilgisayarlı bir testtir. BBST testin uzunluğunu her bir birey için uyarlamakla kalmaz aynı zamanda bireyle ilgili bilgiyi, bireyin parametreleri bilinen sorulara verdikleri cevaplar şeklinde kullanır ve bu bilgiye göre uyarlanabilir şekilde maddeleri seçer. Maddenin zorluğu bireyin yeteneği ile ilişkilendirilir, böylece yeteneği yüksek olan birey basit maddelerle zaman kaybetmez ve yeteneği düşük olan birey seviyesinin üzerinde zor maddelerle karşılaşmaz. Değişken uzunluklu testler, rastgele madde seçimi ve sıralı madde seçimi algoritmalarıyla oluşturulabilir. Sıralı seçim, bir kesme puanı noktasında bilgiyi en üst düzeye çıkarmak için maddelerin seçimi gibi maddeleri akıllıca seçen rastgele olmayan bir algoritmadır. Değişken uzunluklu BBST oluşturmak için 3 ana özelliğin tanımlanması gerekmektedir.

1. Seçili bir psikometrik modelle güçlendirilmiş madde havuzu
2. Madde seçim algoritması
3. Sınıflama kriterleri

Değişken uzunluklu testlerde sınav amacına ulaşırsa çok az sayıda madde kullanılsa bile sonlandırılır. Bilgisayarda Bireyselleştirilmiş Testler yeteneğin kestirimi için en uygun değişken uzunluklu testlerdir. BBST ise sınıflama yapmak amacıyla kullanılan en uygun değişken uzunluklu testlerdir.

BBST kağıt kalem testlerinde olduğu gibi tek seferde verilmek yerine “tur” sistemiyle işleyen bir test olarak kavramsallaştırılabilir. Genel algı, bir maddenin veya madde grubunun (test takımı) her turun başında seçilmesi ve testi cevaplayan kişinin maddelere yanıt vermesi ve bilgisayarın, bireyin sınıflandırılıp sınıflandırılmayacağını değerlendirmek için yanıtları kullanmasıdır. Sınıflama kriteri bu değerlendirme için nicel temeli sağlar. Sınava giren kişi sınıflandırıldığında, test sonlandırılır. Sınıflama kriteri karar veremiyorsa süreç başka bir turla kendini tekrar eder (Spray & Reckase, 1994). Bireye cevaplayacağı bir madde sunulur. Bilgisayar hemen maddeyi puanlar ve sınıflama kriterinin bireyin sınıflandırılmasıyla ilgili bir sonuca ulaştırıp ulaştırmayacağını kontrol eder. Bireyin sınıflandırılmasına ilişkin kararın bir maddeden sonra gerçekleşmesi olası değildir, bu nedenle bilgisayar bir sonraki tura geçer ve cevaplanması için başka bir madde seçer. Birey bu maddeye yanıt verir ve sınıflama kriteri yeniden değerlendirilir. Birey sınıflandırılmıyor ise sınıflandırılması için üçüncü bir madde yöneltilir. Bu işlem sınıflama kriteri bireyi sınıflandıran kadar, madde havuzunda madde kalmayana kadar veya maksimum test uzunluğuna ulaşılan kadar devam eder. Sınıflama kriteri ve madde seçim süreci verimli çalıştığı için, sınıflama kriterleri bazen sadece birkaç maddeden sonra sınıflandırma yapabilirler. BBST’de sınıflama kriteri hem sonlandırma hem de sınıflama görevini yerine getirir (Weiss & Kinsbury, 1984). BBST’de bireyin yeteneği kestirildikten sonra kestirilen geçici yetenek düzeyinde en çok bilgi (KY) veren madde ya da kesme noktasında en çok bilgi veren madde (KN) kullanılarak sınav bireye uyarlanmış olur.

BBSTaraştırmalarının genel amacı, BBST’nin verimliliğini en üst düzeye çıkaran belirli yöntemleri belirlemektir. Bir BBST iki kategorili sınıflandırmalı bir durumda kesme puanının çok üstünde veya altında olan bireylerin sınıflandırılmasında yalnızca birkaç

maddeden sonra bir sınıflandırma yapabilirken yetenek seviyesi kesme puanına yakın olan bireyler soru havuzundaki her maddeye cevap verebilir ve sınıflama kriteri yine de bir sınıflandırma yapamayabilir. Bu nedenle daha etkili sınıflama koşullarının geliştirilmesi, yüksek düzeyde kesin ve doğru sınıflandırma seviyelerini korurken birey başına düşen madde sayısı ortalamasını azaltmada yardımcı olacaktır. Kişi başına sınav süresinde bir azalma, çok sayıda teste giren bireyler için zaman finansal ve kaynaklar yansımalar içerebilir. Bir BBST'i tasarlamak için, bir test geliştirme uzmanı, yukarıda belirtilen üç özelliğin her birini ele alacak bir yöntem belirtmelidir. İlk olarak, maddeleri kalibre etmek için bir psikometrik model seçimini gerektiren kalibre edilmiş bir madde havuzu geliştirilmelidir. Bir BBST için psikometrik bir model gereklidir çünkü madde seçim yöntemleri ve sınıflama kriterleri, madde ve yetenek parametrelerini belirlemek ve hesaplamaları gerçekleştirmek için bir psikometrik model kullanır. Başlatma kuralı, yetenek belirleme yöntemleri, test boyunca maddeler ve madde grupları dinamik olarak seçildiğinden, bu seçimi gerçekleştirmek için madde seçme yöntemi, hem testin ne zaman durdurulacağını hem de bireyin nasıl sınıflandırılacağını belirlemek için bir sınıflama kriteri gereklidir.

Thompson'a (2007) göre BBST psikometrik model, madde havuzu, başlama noktası, madde seçme yöntemi, sınıflama kriteri olarak beş ana yapıdan oluşmaktadır.

Tablo 2

Bilgisayarda Bireyselleştirilmiş Sınıflama Testi Algoritması

BBST i Bileşenleri	Uygulanabilen Seçenekler
Psikometrik model	KTK ve MTK
Madde havuzu	Sivri ve Basık
Başlama Noktası	Q= 0 ve Ön bilgi
Madde Seçimi	Kestirim temelli, Kesme noktası temelli, Global
Yetenek Kestirimi	MOK, AOK, BSD, MSD, Owen' ın Bayesçi Yaklaşımı
Sınıflama Kriteri	AOOT, GOO, BUT, GA, BKK

Thompson (2007) ile Weiss ve Kingsbury'e (1984) göre BBST algoritması Tablo 2'de özetlenmiştir. Tablo 2'ye göre BBST'de psikometrik madel için Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) kullanılmaktadır. Madde havuzu sivri ve basık, teste başlama noktası yetenek düzeyi sıfır veya ön bilgiler kullanılarak yapılmaktadır. Madde seçiminde kestirilen yetenek temelli, kesme noktası temelli ve global (ortak) madde seçme yöntemleri kullanılmaktadır. Yetenek kestirim yöntemi olarak Maksimum Olabilirlik Kestirimi (MOK), Ağırlıklı Olabilirlik Kestirimi (AOK), Beklenen Sonsal Dağılım (BSD), Owen'ın Bayesçi Yetenek Kestirimi gibi yöntemler, sınıflama kriterleri olarak Ardışık Olasılık Oran Testi (AOOT), Genelleştirilmiş Olabilirlik Oran (GOO), Bireyselleştirilmiş Uzmanlık Testi (BUT), Güven Aralığı yöntemi (GA), Bayesçi Karar Kuramı (BKK) gibi yöntemler kullanılabilir. Daha kesin ve etkili BBST uygulaması için daha doğru algoritmaların geliştirilmesi gerekir çünkü BBST bileşenlerinin uygun algoritmalarla oluşturulması, testin etkililiğini ve güvenilirliğini belirler (Thompson, 2009). Bu nedenle bu araştırmada en uygun BBST algoritması belirlenecektir.

Psikometrik Model

Bilgisayarda Bireyselleştirilmiş Sınıflama Testinin en temel bileşeni psikometrik modelin seçilmesidir. KTK ve MTK kullanılan psikometrik modellerdir. KTK sabit madde sayısı ile geleneksel olarak ortalama yetenek etrafında bir yetenek aralığını ölçmek için tasarlanan sınavlarda kullanılır. MTK da madde havuzundaki maddeler, öğrenci yetenekleri ve kesme puanı aynı ölçekte yer aldığı için daha avantajlıdır (Thompson, 2007). Ayrıca MTK'nın yerel bağımsızlık özelliğine göre madde parametreleri örneklemeden, yetenek parametreleri maddelerden bağımsız kestiriliyor olması nedeni ile BBST için genelde MTK modeli tercih edilmektedir. Bu çalışmada İngilizce seviye belirleme sınavına ait tek boyutlu gerçek veri setinin analizleri için MTK modeli kullanılmıştır. Bu nedenle MTK detaylı olarak anlatılmıştır.

Madde Tepki Kuramı (MTK). MTK bireylerin gözlenen yeteneğinden ölçülmek istenen özelliğe ait baskın yetenekler arasında ilişki kurarak bireyin yeteneği tahmin edilir. Madde tepki kuramı bilgisayarda uyarlanabilir test programlarını uygularken de kullanılan model tabanlı bir ölçüm sistemidir. Bireyin maddeye verdiği cevap ile belirlenen özellik arasındaki ilişkiyi gösterir (Demars, 2010).

Geleneksel kağıt kalem testlerinde KTK temelli testin tamamı için analizler yapılır, MTK da ise madde seviyesinde analizler yapıldığı için BBT ve BBST için uygun bir ölçüm modelidir. MTK'ya göre madde parametrelerinin farklı evrenlerde testi alan bireyler için aynı olması ya da madde istatistiklerinin gruptan bağımsız kestirilebilmesi (madde parametrelerinin değişmezlik özelliği), yetenek parametrelerinin değişmezlik özelliği olan öğrenci yeteneklerinin maddelerden bağımsız kestirilmesi ve öğrencilerin kestirilen yeteneklerine ait kestirimin standart hatası belirlenebilmektedir (Hambleton & Swaminathan,1985) . MTK' da değişmezlik özelliği evrendeki parametrelere uygulanır.

MTK da yetenek düzeyi ile doğru yanıt olasılığı arasındaki ilişki, madde karakteristik eğrisi ile gösterilen doğrusal olmayan, monoton artan fonksiyondur (Lord, 1952). Madde karakteristik eğrisi maddeye verilen tepki ile bireyin yeteneği arasındaki ilişkiyi gösteren fonksiyon grafiğidir. Madde karakteristik eğrisinde yetenek düzeyi yükseldikçe öğrencinin doğru cevap verme olasılığı artar (Hambleton & Swaminathan,1985).

Tek boyutlu, çok boyutlu MTK modelleriyle, iki ve çok kategorili maddelerden oluşan testlerde uygulanabilen MTK modelleri bulunmaktadır. Kullanılacak MTK modeli belirlenirken model veri uyumunu gösteren değerler incelenmelidir. Yeteneğin doğru kestirilebilmesi için modelin veriye uyumlu olması gerekir. BBST'de madde seçim algoritması nedeniyle test bireye uyarlanır. MTK'nın bireye uyarlanmış uygulamalarda değişmezlik özelliği nedeniyle kullanılmasının etkili olacağı düşünülmektedir. Bu araştırmada iki kategorili tek boyutlu maddeler kullanılmıştır bu sebeple tek boyutlu MTK modellerine yer verilmiştir.

MTK Modellerinin Varsayımları. MTK da modeller vardır. Veri seti modeller ile uyumlu olmalıdır. Tüm modellerin varsayımları vardır. MTK'nın temel varsayımları tek boyutluluk, yerel bağımsızlık ve testin hız testi olmamasıdır (Hambleton & Swaminathan, 1985). Tek boyutluluk test maddelerinin başat bir özelliği ölçmesidir. Sınava giren kişilerin test maddelerine verdiği yanıtların tek bir gizil özellik veya yeteneğin gözlemlerini temsil ettiğini ima eder (Embretson & Reise, 2000). Her öğrenci için bir tane özelliğe ait yetenek değerinin bulunduğu anlamına gelir ve o tek özellik için maddeye tepki verilir, bunun dışında bir boyut sadece o maddeye özgü tesadüfi hata ya da öngörülme boyut olarak ele alınır, diğer maddeleri etkileyen bir durum değildir. Bu varsayımın yok sayılması parametrelerin veya standart hatanın doğru kestirilmemesine sebep olabilir (Demars, 2010). Test maddeleri birden fazla özelliği ölçüyorsa çok boyutlu MTK modellerinden yararlanılmalıdır.

Yerel bağımsızlık varsayımı ise belli bir yetenek düzeyinde bir maddeye verilen tepkilerin diğer maddelere verilen tepkilerden istatistiksel olarak bağımsız olmasıdır. Testteki bir maddenin diğer maddeye ipucu vermemesidir. Aynı yetenek düzeyindeki bireylerin maddelere verdikleri cevapların birbirinden bağımsız olmasıdır. Yerel bağımsızlık varsayımının ihlali, başka bir boyutun maddelere verilen yanıtları etkilediğini ve dolayısıyla tek boyutluluk varsayımını ihlal ettiğini düşündürülebilir (Demars, 2010). Aynı yetenek düzeyinde maddeler arası kovaryans sıfır ise test yerel bağımsızdır ve tek boyutludur (Hambleton & Swaminathan, 1985). MTK da yerel bağımsızlığa madde düzeyinde bakılır. Yerel bağımsızlık boyutluluktan ayrı değerlendirilmelidir (Demars, 2010).

Bireylerin maddelere cevap verirken hız ya da zamandan etkilenmesi testin yerel bağımsızlık ve tek boyutluluk özelliğini etkilediği için maddeleri cevaplayabilecekleri kadar sürenin verilmesi uygulanan testin hız testi olmamasını sağlar. Bu durum testi yarım kalmış öğrenci sayısına bakılarak anlaşılır (Hambleton & Swaminathan, 1985). Bir testin hız testi olması tek boyutlu olmasını engellemektedir çünkü hız test performansını etkileyen bir özellik olarak test puanlarına yansımaktadır. Testin hız testi olup olmadığını belirlemek için

aynı testin zaman sınırlaması olmadan yapılan uygulamasından elde edilen puanların karşılaştırılması yapılabilir.

MTK Modelleri. Çoktan seçmeli maddelerde her bir madde için doğru/yanlış puanlama sistemi kullanıldığından iki kategorili MTK modelleri kullanılabilir. Likert tipi ölçeklerde tüm maddeler için çeşitli düzeyler yer almakta ve maddeler çoklu puanlanmaktadır (Hambleton & Swaminathan, 1985). MTK' da her model kendine özgü madde karakteristik eğrisi ile tanımlanır. MTK da modeller doğru yanıt olma olasılığı ile yetenek düzeyi arasındaki ilişkiye dair matematiksel eşitlik sunmaktadır. MTK da yetenek arttıkça başarı, performans ya da ölçülen özelliğin miktarı artmalıdır. Her bir modelde sayısal değerleri belirli bir madde karakteristik eğrisini tanımlayan bir veya daha fazla parametre yer alır. En temel ayırım maddeyi tanımlamakta kullanılan parametre sayısıdır. Yaygın kullanılan modeller bir parametrelili lojistik model (1 PLM), iki parametrelili lojistik model (2 PLM), üç parametrelili lojistik model (3 PLM) ve dört parametrelili (4 PLM) lojistik modeldir. Modellerin her birinde çeşitli yetenek düzeyinde maddeyi doğru yanıt olma olasılığının tespiti için matematiksel eşitlikler kullanılır sonra bu eşitliklere karşılık gelen madde karakteristik eğrisi çizilir. MTK da tüm modellerin varsayımları vardır. Bu varsayımlar iki grupta ele alınabilir. Birincisi MTK ya ait ve tüm modeller için geçerli olan varsayımlar olan tek boyutluluk, yerel bağımsızlık ve testin hız testi olmamasıdır. İkinci grup ise modellere özgü varsayımlardır. Örneğin 1 PLM de madde ayırt edicilik güçleri eşit şans parametresi sıfır kabul edilmektedir. Bu çalışmada iki kategorili puanlanan İngilizce seviye belirleme sınavına ait gerçek veri seti kullanıldığı için ikili puanlanan tek boyutlu maddeler için kullanılan 1 PLM, 2 PLM, 3 PLM MTK modellerine yer verilmiştir.

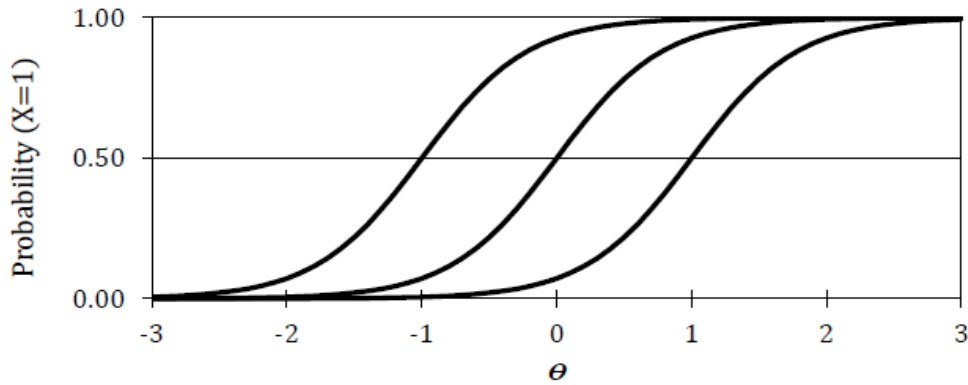
1 Parametrelili Lojistik Model. Model tek bir madde parametresinin tahminlerini sağlar (Rasch,1960).Tüm maddeler için ayırt edicilik (a parametresi) güçlerinin eşit olduğunu kabul eder. Madde güçlük parametresi b_i ile temsil edilir. 1 PLM için kestirilen tek madde parametresidir.1 PLM için bir maddeye doğru yanıt verme olasılığı şu şekilde tanımlanır (Hambleton & Swaminathan, 1985).

$$P(X_i = 1|\theta_j) = \frac{e [D\tilde{a}(\theta - b_i)]}{1 + e [D\tilde{a}(\theta - b_i)]} \quad (1)$$

Formülde teta (θ) sınava giren kişinin yetenek seviyesini temsil eder, bir j kişinin doğru cevap verme olasılığını gösterir, b_i i maddesinin zorluk seviyesini veya konumunu temsil eder. %50 olasılıkla maddeyi doğru yanıtlayabilecek yetenek düzeyini gösterir. Teorik olarak b_i parametresi $-\infty$ ile $+\infty$ arasında değişir uygulamada değeri $[-3,+3]$ arasında değişmektedir. Formüldeki a parametresi madde ayırt ediciliği olup tüm maddeler için sabit ve birbirine eşittir bu nedenle tüm maddelerin zorluk seviyelerine eşit olarak ayırım yaptığı varsayılarak madde karakteristik eğrileri çakışmaz. Amaç bireyin sadece yeteneğini kestirmek olduğunda az sayıda parametre içeren 1 PLM kullanılabilir.

Şekil 1

1PLM İçin Madde Karakteristik Eğrisi



Madde karakteristik eğrisi sınava giren kişinin yeteneği, madde güçlük parametresi ve maddeye verilen yanıtın olasılığı arasındaki ilişkiyi göstermek için kullanılır. Bireyin yeteneği x ekseninde yer alan θ ölçeği ile temsil edilir. Madde güçlüğü x ekseninde yer alan yetenek ölçeği ile aynı ölçekte yer alır bu sınava giren kişinin yetenek düzeyi ile madde güçlüğü arasında doğru yanıt olasılığına ilişkin doğrudan karşılaştırmayı kolaylaştırır. Bir madde için

b_i parametresi yetenek ölçeğinde doğru cevaplama olasılığı 0.50 ye karşılık gelen noktadır. Yetenek ölçeğiyle bağlantılı olarak madde karakteristik eğrisinin konumunu gösterir. +3 e yakın maddeler zor -3 e yakın maddeler kolay maddelerdir.

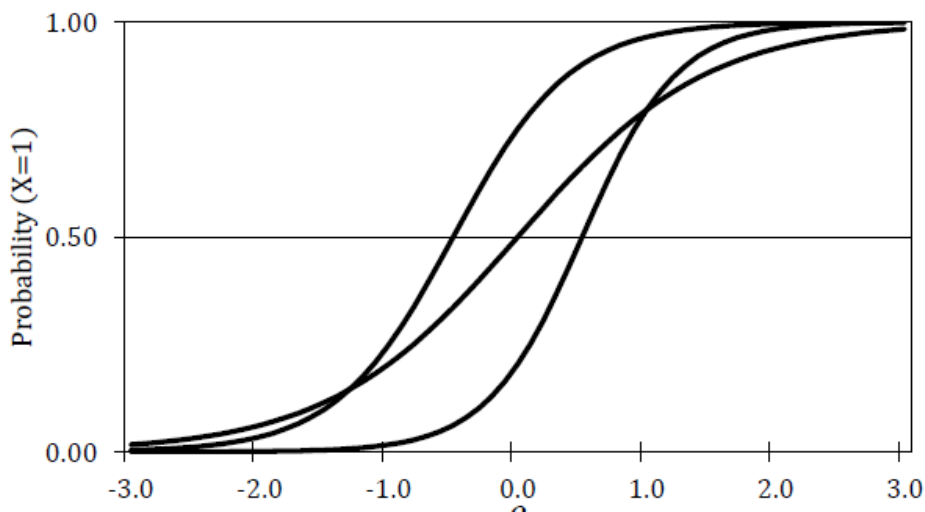
2 Parametrelili lojistik model. 2 PLM de madde güçlüğü (b_i) ve madde ayırt ediciliği (a_i) olarak adlandırılan iki parametre bulunur. Theta (θ) bireyin yetenek seviyesini temsil eder. Ayırt edicilik parametresi (a) yüksek maddeleri doğru cevaplayan öğrenciler daha yüksek yeteneğe sahiptir (Hambleton ve Swaminathan, 1985). Belirli bir (θ) ile sınava giren j bireyinin bir i maddesine doğru yanıt verme olasılığı şu şekilde tanımlanır.

$$P(X_i = 1 | \theta_j) = \frac{e [D a_i (\theta - b_i)]}{1 + e [D a_i (\theta - b_i)]} \quad (2)$$

Burada b_i madde güçlüğü, veya maddenin ölçülen özellik için konumunu temsil eder, a_i i maddesinin ayırt etme gücünü temsil eder. D ölçekleme sabitidir. Modelde a parametresi ; $(-\infty, \infty)$ arasındadır uygulamada $[0, 2]$ dir.

Şekil 2

2 PLM İçin Madde Karakteristik Eğrisi



Grafikte eğimler farklı olduğu için eğriler birbirini kesebilir. Ayırt edicilik parametresi (a) bükülme noktasındaki eğimle orantılıdır. Madde karakteristik eğrisinin eğimi arttıkça maddenin ayırt etme gücü de artmaktadır. Daha yüksek a parametresi maddenin güçlük seviyesine yakın farklı yetenek seviyeleri arasında daha iyi ayırım yapılabildiğini gösterir. Madde güçlük parametresi (b) bükülme noktasına karşılık gelen yetenek seviyesidir. Bükülme noktasında maddeye doğru yanıt verme olasılığı 0.50 dir.

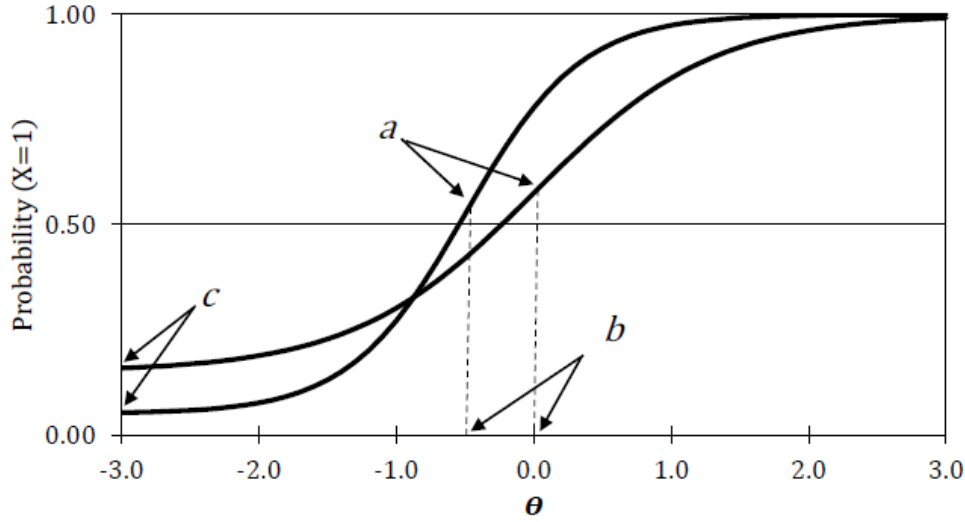
3 Parametrelili Lojistik Model. 3 PLMde madde güçlüğü (b), madde ayırt ediciliği (a), madde şans parametresi (c) üç parametre tahmin edilir (Birnbaum, 1968). 3 PLM de teta öğrencinin yetenek seviyesini temsil eder. Belirli bir θ ile sınava giren bir j nin bir maddeye doğru yanıt verme olasılığı aşağıda tanımlanmıştır.

$$P (X_i= 1 | \theta_j) = C_i + (1 - C_i) \frac{e [D_{ai} (\theta - b_i)]}{1 + e [D_{ai} (\theta - b_i)]} \quad (3)$$

Burada b_i , i maddesinin gizil özellik boyunca zorluk seviyesini veya konumunu temsil eder, a_i i maddesinin ayırt etme gücünü, c_i sınava giren j bireyinin doğru yanıt seçeneğini tahmin etme olasılığını temsil eden şansla doğru yanıt verme olasılığıdır. D ölçekleme sabitidir. 3 PLM de diğer modellerden farklı olarak c parametresi şansla doğru yanıtlama olasılığını gösterir ve madde karakteristik eğrisinin alt asimptotu olarak tanımlanır. Alt asimptot sıfırdan büyük olabileceğinden grafiğin bükülme noktası 0.50 ye eşit olan doğru yanıt verme olasılığına karşılık gelmeyecek daha büyük olacaktır. C parametresinin teorik olarak aralığı [0,1] dir. 3 PLM de b parametresi c parametresinden dolayı 0.50 olasılıkla doğru yanıtlanabilecek yetenek düzeyini değil bu yetenek düzeyinin üstüne denk gelen $1+c/2$ olasılıkla doğru yanıtlanabilecek yetenek düzeyini gösterir. 3 PLM c parametresi nedeni ile çoktan seçmeli sınavlarda kullanılır.

Şekil 3

3 PLM İçin Madde Karakteristik Eğrisi



BBST uygulamasında 3 PLM ile sınıflama analizleri yapıldığında daha yüksek sınıflama doğruluğu ve daha az madde ile sınavın sonlandığı yapılan araştırmalardan anlaşılmaktadır (Reckase, 1983; Lau, 1996; Jiao & Lau, 2003). MTK modellerinde 3 PLM kullanıldığında parametre kestirimleri için yakınsama problemi de yaşanabilmektedir (Eckes, 2014).

Model Seçimi

MTK da uygun modelin seçimi için model varsayımlarının test edilmesi gerekir. Bir verinin MTK nin varsayımlarını sağladıktan sonra hangi modelle analiz edileceğine ilişkin çeşitli ölçütler bulunmaktadır.

Model veri uyumu, puanlamada maddelerin nasıl ağırlıklandırıldıkları, ölçmede istenilen ölçek özellikleri, parametre kestiriminin amacıdır (Hambleton & Swaminathan, 1985). Parametre ve yetenek kestirimlerinin hatasız olabilmesi adına kullanılan verinin tercih edilen modelin varsayımlarını karşılar nitelikte, model ile uyumlu olması gereklidir. Modellerin en genel hali 3 PLM dir. Modelin kendine özgü varsayımları yoktur. Şans parametresi (c_i) sıfır olduğunda 2 PLM elde edilir. 2 PLM 'e ayırt edicilik parametresi (a_i) değerlerinin eşit olduğu varsayımı eklenirse 1 PLM elde edilir.

Model veri uyumunun iyiliği için AIC ve BIC değerlerinin düşük – loglikelihood değerlerinin ise mutlak değerlerinin küçük olması beklenir. -Loglikelihood değeri, verinin modelden uzaklaşma derecesini göstermektedir. Değer ne kadar düşük olursa veri modele o kadar iyi uyum sağlamıştır (Embretson & Reise, 2000). Değer ne kadar büyükse veri modelden o kadar uzaklaşır. Model-veri uyumunu değerlendirmede AIC, BIC, -loglikelihood uyum istatistikleri ANOVA ile incelenerek iki model karşılaştırılabilir.

MTK da madde parametrelerinin gruptan ve birey parametrelerinin maddelerden bağımsızlığı sayesinde bilgisayarda bireyselleştirilmiş testlerde ve bilgisayarda bireyselleştirilmiş sınıflama testlerinde bireylerin yeteneği maddelerden bağımsız olarak kestirildiği için bireyler farklı maddeleri cevaplasalar bile karşılaştırılabilmektedir. (Hambleton & Swaminathan, 1985).

Madde Havuzu

Etkili bir BBST uygulaması için büyük bir madde havuzuna ihtiyaç duyulur. Madde havuzunda bulunması istenen çok sayıda maddeyi aynı gruba tek seferde uygulayıp elde etmek zordur. Aynı gruba farklı zamanlarda ya da aynı evrenden farklı gruplara uygulama yapılır (Kolen ve Brennan, 2004). Bu durumda yapılan uygulamalarda ortak “(anchor)” maddeler kullanılarak testlerin ölçekleme yapılarak kalibre edilmesi uygun olacaktır (Nydick & Weiss, 2009).

Etkili bir madde havuzu kalibrasyon sürecinde seçilen MTK modeline, madde seçme yöntemlerine ve madde havuzunun yapısına bağlıdır. Madde havuzunda maddeler kağıt kalem formatında uygulanır ve uygun MTK modeli ile kalibrasyonu yapılır (Yoes, 1995). Kesme noktası ve kestirilen yetenek temelli madde seçme algoritmalarına bağlı olarak madde parametreleri için istenen ranj aralığı belirlenir. Kesme noktası temelli madde seçme algoritması kullanılıyorsa kesme noktasında yüksek bilgi veren ve bu bölgede güçlük değerine sahip maddelere ihtiyaç duyulur. Kestirilen yetenek temelli madde seçme algoritması kullanılıyorsa bireyin kestirilen yetenek düzeyinde yüksek bilgi veren ve bu düzeydeki güçlükte madde seçim ranjı geniş olan maddeler olmalıdır. Bilgisayarda

Bireyselleştirilmiş Test Uygulamalarında amaç not vermek ise yetenek ölçeğinin her noktasında maddeye ihtiyaç duyulmaktadır (Thompson & Weiss, 2011).

İki kategorili sınıflamada tek bir kesme puanı olduğu için madde havuzu kesme puanı etrafında yüksek bilgi veren maddelerden oluşabilir. Çok kategorili sınıflamada birden fazla kesme puanı olduğu için sınav katılımcısının puanı iki kesme puanı arasındaysa sorun yaşanabilir. Bu nedenle kestirilen yetenek temelli madde seçme yöntemlerinin de kullanılması gerekir (Haring, 2014). Stocking (1994) BBST uygulamasında madde havuzunun büyüklüğü için kağıt kalem formatının en az 6 katı madde olması gerektiğini belirtmektedir. Bu araştırmada kestirilen yetenek temelli ve kesme noktası temelli madde seçme yöntemleri kullanılmıştır. Madde havuzunda kağıt kalem formatında uygulanan 9 setten oluşan, her sette yaklaşık 35-39 arasında toplam 256 madde bulunmaktadır. Her setteki madde sayısının yaklaşık 6 katından fazla madde havuzunda madde bulunmaktadır.

Başlama Noktası

BBST uygulaması için bireyler hakkında ön bilgiler mevcutsa ön bilgi kullanılarak ve önceden belirlenen yetenek düzeyleri başlama noktası olarak belirlenebilir (Weiss & Kingsbury, 1984; Yang ve diğerleri, 2006). Başlama noktasının bireyin yetenek düzeyi ile farklı olmasının kaygıya sebep olduğu düşünülmektedir (Wainer & Kiely, 1987). Önsel bilgi olmadığında yetenek ranjının ortasındaki bir madde ile ($\theta = 0$) başlanması kullanılan bir yöntemdir (Thompson, 2007). Bu araştırmada koşulların tamamı için başlama noktası $\theta = 0$ olarak belirlenmiştir.

Madde Seçimi

BBST uygulamasında madde seçim prosedürü sınıflandırmayı sağlayan temel bileşendir. BBST uygulamasının amacı sınava giren kişinin en doğru şekilde sınıflandırılmasını sağlamaktır. Ölçümün kesinliği sınava giren kişinin yeteneğini belirleyecek en bilgilendirici maddenin seçilmesiyle sağlanır. Madde seçim yöntemlerindeki farklılıklar, aynı sınav katılımcıları için farklı madde seçimlerine ve dolayısıyla farklı

sınıflandırma kararlarına yol açabilir. Bu nedenle, madde seçim yönteminin seçimi, bilgisayarda bireyselleştirilmiş sınıflama testin en önemli parçalarından biridir (Nurakhmetov, 2019).

BBST de en temel madde seçimi rastgele madde seçimi ile geliştirilmiştir (Ferguson, 1969). Rastgele madde seçimi ile ne seçilen maddeler ne de öğrenciler hakkında bilgi kullanmamaktadır. Rastgele madde seçiminin verimsiz olduğunu Kingsbury ve Weiss (1983) yaptıkları çalışmalarla ortaya koymuşlardır.

Akıllı madde seçimi belli bir yöntemle madde seçimi olarak daha doğru sınıflama yapmaktadır. Kesme noktasında bilgi veren ve kestirilen yetenekte bilgi veren maddeler olarak iki yöntem bulunmaktadır (Thompson, 2007). Kestirim temelli madde seçme yöntemi ile kesme puanından bağımsız sınava giren kişinin testte o noktaya kadar kestirilen yetenek düzeyinde en yüksek bilgiyi veren madde seçilmektedir. Sınava giren bireyin testin o noktasına kadar verdiği yanıt vektörü ile hesaplanır. Kesme puanı temelli madde seçme yönteminde maddeleri kesme puanına göre değerlendirir yetenek tahminini içermez, kesme noktasında en yüksek bilgiyi veren madde seçilir. Kesme puanı temelli üç yöntem vardır. Maksimum Fisher Bilgisi (MFB), Kullback Leibler Bilgisi (KLB) ve Log-odds ratio yöntemleridir (Lin & Spray, 2000). MFB tek bir noktada bilginin maksimize edilmesini sağlar, KLB'da kesme puanı etrafındaki küçük bir bölgedeki bilgiyi değerlendirir ($\theta_0 - \theta_1$). Log-Odds Oranı yönteminde kesme puanının her iki tarafında olasılık değerleri arasında en büyük farklılıkları üreten madde seçilir (Embretson & Reise, 2000). Lin ve Sprey'e (2000) göre bu üç yöntem de aynı şeyi değerlendirdiği için karşılaştırıldığında üçü de neredeyse eşdeğerdir. Bir madde seçim kriteri açısından en yüksek bilgi veren madde diğeri için de en yüksek bilgi verecektir.

Kestirim temelli madde seçiminde de MFB ve KLB kullanılabilir. (Reckase, 1983; Spray & Reckase,1994; Eggen,1999). Burada kesme puanında değil kestirilen yetenek düzeyinde ve yetenek düzeyi çevresinde en çok bilgi veren maddenin seçimi benzerdir. Kestirilen yetenek düzeyine göre birey için madde seçilir. Bu da teste bireysellik

katar. Bunların dışında global (ortak) madde seçim yöntemi de bulunmaktadır. Yetenek düzeyinde madde bilgisini değerlendiren geniş yetenek ranjında bilgiye ihtiyaç duyulduğunda bireyin yeteneği ile ilgili test başlangıcında en az bilgi sahibi olduğunda ve birden fazla kesme puanı olduğunda kullanılabilir. İstenen bilginin tek kesme puanı olduğu durumlarda etkili değildir (Weissman, 2006).

Bu çalışmada kesme puanının bir, iki ve üç olduğu sınıflama kategori sayısının iki, üç ve dört olduğu tek boyutlu dil testine ait gerçek veri seti ile BBST uygulamalarında MFB kestirilen yetenek (MFB-KY) ve MFB kesme noktası (MFB-KN), KLB kestirilen yetenek (KLB-KY), KLB kesme noktası (KLB-KN) temelli madde seçme yöntemleri kullanılmıştır.

Yetenek Kestirimi

MTK temelli puanlama öğrencilerin maddelere verdikleri cevaplara ve kestirilmiş madde parametrelerine dayanır. Bir yetenek kestirim prosedürü yetenek kestirimini elde etmek için sınav katılımcısının maddelere verdiği yanıt örüntüsünü madde parametreleriyle birlikte kullanır. Bilinen madde parametreleri kullanılarak sınav katılımcısının bilinmeyen yeteneği tahmin edilir.

BBST de ise sınıflamanın etkililiği ve maddeleri kalibre etmek için psikometrik model seçimini gerektiren kalibre edilmiş madde havuzu, madde seçimi ile sınıflama kriterleri ve parametreleri belirlemek için psikometrik model, madde seçimi için ve buna bağlı sınıflama kararlarının uygunluğu için yetenek kestirim yöntemi ve testin ne zaman durdurulacağı bireyin nasıl sınıflanacağını belirlemek için bir sınıflama kriteri gereklidir (Thompson, 2009). BBST’de madde seçiminin doğru olması yetenek kestirim yönteminin doğru belirlenmesine bağlıdır. Az sayıda madde ile doğru sınıflama yapılmasının temeli yetenek kestirim yöntemine bağlıdır. BBST’ uygulamasının amacı da az madde ve yüksek sınıflama doğruluğu ile sınıflama yapmaktır (Thompson, 2009). BBST için yetenek kestirim yönteminin performansı bu nedenle çok önemlidir.

Alanyazında MTK temelli Maksimum Olabilirlik Kestirimi (MOK; Birnbaum, 1968), Marjinal Maksimum Olabilirlik Kestirimi (MMOK; Bock ve Aitkin, 1981), Ağırlıklandırılmış Olabilirlik Kestirimi (AOK; Warm, 1989) ve Bayeşçi yetenek kestirim yöntemleri olan Ardışık Bayeşçi Kestirim Yöntemi (Owen, 1975), Maksimum Sonsal Dağılım yöntemi (MSD; Samejima, 1969), Beklenen Sonsal Dağılım (BSD; Bock ve Aitkin, 1981), olmak üzere farklı yetenek kestirim yöntemleri bulunmaktadır. Yetenek kestirim yöntemlerinin tamamı bir miktar da olsa yanlı kestirim yapmaktadır (Warm, 1989). BBST çalışmalarında yetenek kestirim yöntemlerinin performanslarına ait bilgi az bulunmaktadır (Wang & Vispoel, 1998).

Yetenek kestirim yöntemlerinin bazı olumsuz yanları bulunmaktadır. Örneğin Maksimum olabilirlik yöntemi maddelerin tümü doğru tümü yanlış olduğunda kestirimde bulunmaz. Verinin modele uygun olduğu düşünülerek kestirim yapılır (Warm (1989)). Yansız olması tetanın beklenen değerinin gerçek değerine eşit olması etkili bir kestirici olması, hatalarının normal dağılım göstermesi olumlu özellikleridir (Embretson & Reise, 2000). Literatüre göre değişken uzunluktaki bireyselleştirilmiş testlerde yetenek kestirimi için AOK kullanıldığında MOK'a göre daha az sayıda madde ile test sonlandırıldığı görülmektedir. Bu durum da madde kullanım sıklığını düşürür test süresini azaltır testin kullanılabilirliğini artırır (Warm, 1989). Tüm maddelerin doğru veya yanlış cevaplanması durumunda MOK'un kestirimde bulunamaması özelliğinden dolayı Bayeşçi kestirim yöntemlerinin MOK'a alternatif kestirim yapabildiği literatüre göre bilinmektedir. Bu çalışmada bayeşçi kestirim yöntemlerinden BSD ve MOK'a göre daha kullanışlı olduğu literatürde belirtilen AOK yöntemi ve tüm maddelerin yanlış veya tüm maddelerin doğru cevaplandığı veri madde havuzunda bulunmadığı için, MOK'un sınıflama çalışmalarında literatürde performansı çok araştırılmadığı için MOK yetenek kestiriminin de performansı araştırılmıştır. Aşağıda BSD ve AOK ve MOK açıklanmıştır.

Beklenen Sonsal Dağılım Yetenek Kestirim Yöntemi (BSD). BSD iteratif bir yöntem değildir. BBST için istenen durum olan daha hızlı yetenek kestirimini yapabilmektedir. Tüm maddeler doğru ya da tüm maddeler yanlış işaretlene bile yetenek

kestirimi yapabilir (Embretson & Rise, 2000). BSD önsel dağılımın modu yerine ortalamayı bularak türetilen bir bayesian kestirimidir. Önsel dağılım tetanın evrende nasıl dağıldığı hakkındaki bilgidir. Örneğin birey normal dağılım gösteren bir evrenin üyesi ise cevap örüntüsü için tetanın olabilirlik kestiriminin -5 değerini alması çok olası değildir (Demars 2010). Madde sayısı az olduğunda BSD'in kestirimi yanlı olur. Bayes istatistiklerinde önsel dağılım gözlenen veriye dayanan olabilirlik fonksiyonu ile çarpılır sonuç sonsal olabilirlik olarak adlandırılır (Demars, 2010). Madde sayısı az olduğunda yetenek kestirimlerinin değeri ortalama yönünde yanlılık gösterir. Ortalamaya doğru yanlılık uç noktalara yakın değerleri, yetenek tahminlerinin ortalamasına yakın olan değerlerden daha fazla etkiler (Lord, 1986). Bu sebeple bayesian yetenek kestirimlerinin maddelerin uygulanma sıralarına bağlı etkilerinin olduğunu, aynı madde setini alan fakat maddelerin sunulma sırası değişen sınav katılımcıları farklı yetenek tahminlerine sahip olabilir (Parshall ve diğerleri, 2002). Olumsuz yanı kestirimin yanlı olmasıdır (Wainer & Thissen, 1987; Embretson & Reise, 2000). Yetenek kestirim yönteminin hızlı olması, etkili kestirim yapması ve çalışmada madde havuzunun büyüklüğünün az olmaması, gerçek veri üzerinden kestirimlerin yapılacak olması, önsel dağılımların bilinmesi, MOK'a alternatif olması nedeniyle çalışmada MOK ve BSD iki yetenek kestirim yönteminin de performansını BBST çalışmasında görebilmek adına araştırılmıştır. Bu sebeplerden dolayı BSD'in çalışma için uygun bir kestirim yöntemi olduğu düşünülmektedir.

Ağırlıklandırılmış Olabilirlik Kestirim Yöntemi (AOK). Maksimum Olabilirlik Kestirimi, Maksimum sonsal Kestirim, Beklenen Sonsal Dağılım ve Ağırlıklandırılmış Olabilirlik Kestirim yetenek kestirim yöntemlerinde örtük özelliğin normal dağılım gösterdiği kabul edilir. MOK'de fonksiyonun maximum değerini bulmak için en çok kullanılan yöntem Newton-Raphson yöntemidir. Bu yöntemde yetenek ranjı üzerinde olabilirlik fonksiyonunun modu bulunmaktadır. Madde ve kişi sayısı arttığında çok zaman alıcıdır (Warm,1989). AOK yöntemi Warm (1989) tarafından geliştirilmiştir. Yetenek kestirim yöntemlerinin yanlı olması sınıflamayı sistematik olarak etkilemektedir (Wang & Wang, 2001). Yetenek

kestiriminde olabilirlik fonksiyonunun ortalamasını dikkate alarak yanlılığı azaltan AOK yöntemidir (Warm, 1989). Sabit uzunluklu testlerde AOK yönteminin daha yansız olduğu ve daha az sayıda madde ile testi sonlandırdığı bilinmektedir. Değişken uzunluklu testlerde performansının nasıl olduğu tam olarak bilinmemektedir (Wang, Hanson & Lau, 1999; Yi, Wang & Ban, 2000; Gündeğer, 2017). Araştırmada AOK yetenek kestirim yönteminin değişken uzunluklu test olan BBST de nasıl performans gösterdiği araştırılmak istenmiştir.

Maksimum Olabilirlik Yetenek Kestirim Yöntemi (MOK). MOK prosedürü, olabilirlik fonksiyonunu maksimize eden yetenek (θ) değerini bulmak için sınav katılımcısının belirli bir kalibre edilmiş madde setine verdiği yanıt dizisini kullanır. Belirli bir madde setine verilen herhangi bir yanıt dizisi için, yetenek sürekliliği boyunca her (θ) değeri için log birimleri cinsinden bir olabilirlik değeri hesaplanabilir. Maksimum olabilirlik yetenek tahmini, her (θ) değerine koşullu olabilirliklerin toplanması ve ardından bir Newton-Raphson yineleme prosedürü aracılığıyla olabilirlik fonksiyonunun modunun belirlenmesiyle elde edilir (Lord, 1980; Embretson & Reise, 2000). MOK, BSD ve MSD yetenek kestirim yöntemlerine göre daha az yanlı yetenek kestirimi yapar (Lord, 1986). MOK yetenek kestiriminin en önemli sorunu maddelerin tümü yanlış ya da tümü doğru cevaplandırıldığında bir yetenek kestirimi sağlayamamasıdır. Yetenek tahmini için iteratif yöntem kullanılır. Bu sebeple daha yavaş kestirim yapar. Literatürde MOK'un özellikle çok kategorili BBST uygulamalarında performansına ait çalışmaya rastlanmamıştır. Gerçek veriye dayanan yanıt vektöründe tümü yanlış tümünün doğru olduğu veri seti bulunmamaktadır. Daha az yanlı olması ve BSD ve AOK yetenek kestirimlerine göre performansını araştırmak amacı ile araştırmada kullanılmıştır.

Sınıflama Kriterleri

Bir sınav katılımcısını sınıflandırmak için kullanılan matematiksel algoritma sonlandırma olarak adlandırılır. BBST 'inde sınıflama kriterleri aynı zamanda sonlandırma kriterleridir. Ardışık Olasılık Oran Testi (Wald, 1947; Eggen, 1999), Güven Aralığı (Kingsbury & Weiss, 1983; Eggen & Straetmans, 2000) ve Bayesçi Karar Kuramı

(Vos,1999; Rudner, 2002) olmak üzere üç sınıflama kriteri vardır. Thompson (2009) tarafından yapılan çalışmada GOO sınıflama kriterinin AOOT'nin genişletilmiş hali olarak kullanımı önerilmiştir. Kullanılan kriterin doğruluğu psikometrik model, madde seçme yöntemi ve yetenek kestirim yöntemine bağlıdır. Bu üç sınıflama kriterinin de az sayıda madde ile benzer doğrulukta sınıflama yaptıkları bilinmektedir (Kingsbury & Weiss, 1983; Rudner, 2002). Literatüre göre KY temelinde madde seçimi kullanıldığında GA sınıflama kriterinin daha etkili olduğu, KN temelinde madde seçimi kullanıldığında AOOT ve GOO sınıflama kriterinin daha başarılı olduğu bilinmektedir (Eggen & Straetmans, 2000; Spray & Reckase,1996; Thompson, 2009; Demir, 2019). BBST uygulamalarında alanyazında en çok AOOT yöntemi kullanılmaktadır (Thompson, 2009). Bu çalışmada AOOT, GOO ve GA yöntemlerinin performanslarını karşılaştırabilmek ve madde seçme ve yetenek kestirim yöntemleriyle oluşturulacak çaprazlamada en etkili koşulu belirlemek için üç kriter de kullanılmıştır. Aşağıda üç sınıflama kriterine ait bilgiler verilmektedir.

Ardışık Olasılık Oran Testi Sınıflama Kriteri (AOOT). AOOT prosedürü, Abraham Wald (1947) tarafından İkinci Dünya Savaşı sırasında kalite kontrol test prosedürlerinde malzemeleri korumanın bir yolu olarak geliştirilmiştir. Kingsbury and Weiss'e (1983) göre geleneksel sınıflandırma test yöntemlerinin ve AOOT yöntemlerinin maddelerin çoğunun kesme puanına yakın max bilgiye sahip olduğu durumda sınıflama doğruluk oranları %87 ve %86 gibi çok benzerdir. AOOT için optimal düzeyden daha az bir madde havuzu kullanıldığında sınıflandırma doğruluğunun önemli ölçüde düştüğünü ve madde havuzunun önemini vurgulamıştır. AOOT'de olabirlik fonksiyonu kullanılmaktadır. Cevap dağılımının olabirliği Hipotezlerin birinin doğru olmasına karar vermek için kullanılır. Olabirliği yüksek olan hipotez kabul edilir. Hipotezlerin olabirlikleri aynı ise farklı bir madde öğrenciye uygulanır (Reckase,1983; Gündeğer, 2017). Hipotezler şu şekilde yapılandırılmıştır.

$$H_0: \theta \leq \theta_k - \delta = \theta_1 \quad (4)$$

$$H_1 : \theta \geq \theta_k + \delta = \theta_2 \quad (5)$$

θ sınava giren kişinin yetenek seviyesini, θk geçme puanını ya da kesme puanını, δ farksızlık bölgesi sabitini başka bir ifade ile bireyin yetenek düzeyinin kesme puanının altındaki ve üstündeki sabit değere eşitliğini test ederek bu iki sabit değer arasındaki aralık farksızlık bölgesidir. Örneğin alt sınır θ_1 , üst sınır θ_2 ise ikisinin arasındaki uzaklık kesme puanına yakın sınıflama kararları için tolere edilebilen belirsizlik düzeyi ya da hatayı belirten farksızlık bölgesidir. θ_1 sınıflandırma için maksimum alt sınır karar değerini ve θ_2 sınıflandırma için minimum üst sınır karar değerini temsil eder ve dolayısı ile $\theta_1 < \theta k < \theta_2$ (parshall ve diğerleri, 2002). Sıfır hipotezi incelenen kişinin yetenek seviyesinin farksızlık bölgesinin alt sınır noktasına eşit olduğunu, belirtir ve kesme puanının altında bir sınıflandırma ile sonuçlanır. Alternatif hipotez H_1 incelenen kişinin yeteneğinin farksızlık bölgesinin üst sınır noktasına eşit olduğunu belirtir ve bu da kesme puanının üzerinde bir sınıflandırma ile sonuçlanır. Önce θk kesme puanı farksızlık bölgesinin alt ve üst sınırı olarak tanımlanmalıdır. θ_1 ve θ_2 her ikisi de kesme puanı değerine bağlıdır. Farksızlık bölgesi ne kadar büyük olursa test özellikle kesme puanından çok uzak yeteneklere sahip bireyler için o kadar kısa olacaktır. Kesme puanına yakın bireyler sınıflama hatalarında artışların yanı sıra biraz daha uzun testlere sahip olacaktır. Daha küçük farksızlık bölgeleri sınıflandırma doğruluğunu daha iyi koruyabilir ancak daha uzun testler gerekir. θ_1 ve θ_2 değerlerinin log olabilirliklerinin oranlanmasıyla hipotezler test edilir.

$$LRk (\theta_2, \theta_1; \mathbf{x}) = \frac{LRk (\theta_2; \mathbf{x})}{LRk (\theta_1; \mathbf{x})} \quad (6)$$

X bir bireyin i sayıda maddeye verdiği yanıtlardır. ($x = x_1, x_2, x_3, \dots, x_i$). Oranın yüksek olması bireyin yeteneğinin kesme puanının üstünde olduğunu, oranın düşük olması yeteneğin

kesme puanının altında olduğunu gösterir. Hipotezlerin doğruluğu hakkında karar verilirken kabul edilebilir hata oranları

$$P(H_0 \text{ kabul} \mid H_0 \text{ doğru}) \geq 1-\alpha \quad (7)$$

$$P(H_0 \text{ kabul} \mid H_1 \text{ doğru}) \leq \beta \quad (8)$$

Eşitsizlikleri ile tanımlanır. α 1. tip hata oranını başka bir ifade ile yanlış pozitif sınıflandırma hata oranı gösterir. Gerçekte yetenek düzeyi kesme puanından düşük olan kişinin geçer olarak sınıflandırılması örnek olarak verilebilir. β 2. Tip hata oranını gösterir başka bir ifade ile yanlış negatif sınıflandırma hata oranıdır. Gerçek yetenek düzeyi kesme puanından yüksek olduğu halde başarısız olarak sınıflandırılan sınav katılımcısı örnek olarak verilebilir. Her iki hata oranı türü de 0 ve 1 arasında değişir ancak araştırmalarda 0.05 veya 0.10 olarak genelde kullanılır (Parshall ve diğerleri, 2002). Hata oranları sınıflama kararlarındaki alt ve üst sınırları hesaplamak için kullanılır. Sınıflandırmada üst sınır $A = (1-\beta) / \alpha$, alt sınır $B = \beta / (1-\alpha)$ dır.

Karar verilirken ölçütler

$$B / (1-\alpha) < LRk(\theta_2, \theta_1; \mathbf{x}) < (1-\beta) / \alpha \text{ olduğunda teste devam edilir.} \quad (9)$$

$$LRk(\theta_2, \theta_1; \mathbf{x}) \leq \beta / (1-\alpha) \text{ olduğunda } H_0 \text{ kabul edilir.} \quad (10)$$

$$LRk(\theta_2, \theta_1; \mathbf{x}) \geq (1-\beta)/\alpha \text{ olduğunda } H_0 \text{ reddedilir} \quad (11)$$

şeklinde belirlenmiştir (Eggen,1999).

Çok kategorili sınıflamada da iki kategoriye benzer prosedür uygulanır. Eş zamanlı olarak birden fazla kategori değerlendirilir (Eggen, 2000).Örneğin sınıflandırmak için iki kesme puanı kullanıldığında sınav katılımcıları üç kategoriye ayrılır.

Hipotezler ise

$$H_{01}: \theta = \theta_{k1} - \delta_{01} = \theta_{01} \text{ (seviye 1)} \quad (12)$$

$$H_{11}: \theta = \theta_{k1} + \delta_{11} = \theta_{11} \text{ (seviye 1'in üstünde)} \quad (13)$$

$$H_{02}: \theta = \theta_{k2} - \delta_{02} = \theta_{02} \text{ (seviye 3'ün altında)} \quad (14)$$

$$H_{12}: \theta = \theta_{k2} + \delta_{12} = \theta_{12} \text{ (seviye 3)} \quad (15)$$

Burada θ_{k1} iki kesme puanından düşük olanını temsil eder. δ_{01} ve δ_{11} farksızlık sabitleri θ_{k2} ise yüksek olan kesme puanını temsil eder. δ_{02} ve δ_{12} ikinci kesme noktası için farksızlık bölgesi sabitleri.

θ_{01} , θ_{11} değerleri kendi arasında θ_{02} ve θ_{12} değerlerinin kendi arasında log olabilirlikleri oranlanır. Oranın yüksek ya da düşük olması kesme puanının üstünde ya da altında olarak karar verilir. Tek kesme puanında olduğu gibi sınıflandırmada üst sınır $A = (1-\beta) / \alpha$ ve alt sınır $B = \beta / (1-\alpha)$ karar vermede ölçüt olarak kullanılabilir.

Bu çalışmada AOOT sınıflama kriteri için farksızlık bölgesini belirlerken literatüre göre farksızlık bölgesi sabiti ne kadar küçük olursa sınıflama doğruluğunun o kadar yüksek olduğu düşünülerek Thompson (2011) çalışması da dikkate alınarak 0.10 olarak belirlenmiştir. Gerçek veri seti ile iki, üç ve dört kategorili sınıflama yapılacağı için kesme noktası literatüre göre rastgele de belirlenebilir ya da Eggen ve Straetmans 'ın (2000) çalışmasında olduğu gibi BBST öncesinde belirlenen yetenek parametreleri iki kategorili sınıflama için seviye bir ve seviye iki, üç kategorili sınıflama için seviye üç ve dört kategorili sınıflama için de seviye dört de olacak şekilde kodlanıp iki kategorili sınıflama için 1. Seviyedeki en büyük yetenek düzeyinin %70 i alınarak -0.0977 olarak belirlenmiştir. Üç kategorili sınıflama için belirlenecek iki kesme puanı için küçük kesme puanı 1. Seviyedeki en büyük yetenek düzeyinin %70 i belirlenerek büyük kesme puanı da 2. Seviyedeki büyük yeteneğin %70 i belirlenerek -0.34841 ile 0.3099 olarak hesaplanmıştır. Dört kategorili sınıflama içinde belirlenen seviyelerdeki seviye 1, seviye 2, seviye 3 e ait en yüksek yeteneklerin %70 i alınarak -0.5087, -0.0031, 0.4922 olarak kesme noktaları belirlenmiştir.

Güven Aralığı Sınıflama Kriteri (GA). Bu sınıflama kriterinde amaç bireylere ait yetenek düzeyleri için belirlenen güven aralığını kesme puanı ile karşılaştırarak bireyleri uygun kategorilere ayırmaktır.

Çok kategorili sınıflama için Weiss ve Kingsbury (1984) GA sınıflama kriterinin iki kategorili sınıflamadan kolayca uyarlanabilse de çok kategoride sınıflamak için θ ölçeğinde güven aralıklarının kullanılmasını önermişlerdir. Her maddeden sonra sınav katılımcısının yetenek seviyesi kestirilir. Sonra $1-\alpha$ güven aralığı maksimum olabirlik kestirimi etrafında ölçmenin koşullu standart hatası (CSEM) kullanılarak oluşturulur. Bu güven aralığı $\theta_j - z\alpha (CSEM) \leq \theta_j \leq \theta_j + z\alpha (CSEM)$ şeklinde gösterilir. Burada $z\alpha$, $1-\alpha$ ya karşılık gelen sapma değeridir. Her maddeden sonra güven aralığı kesme puanının üstünyse birey başarılı altındaysa başarısız, kesme puanına eşit ise yeni madde ile sınava devam edilir. Literatüre göre güven aralığı değeri arttikça testin sınıflama yapması için gereken madde sayısı ve sınıflamanın doğruluğu artar (Eggen & Straetmans, 2000). Bu nedenle bu araştırmada GA değeri Eggen ve Straetmans'ın (2000) çalışmaları da dikkate alınarak %90 olarak belirlenmiştir.

Genelleştirilmiş Olabirlik Oran Sınıflama Kriteri (GOO). AOOT sınıflama kriteri kullanıldığında madde havuzunda ideal olarak kesme puanına yakın veya kesme puanında parametrelere sahip maddeler ağırlıkta olacaktır. Gerçek bir madde havuzunda maddeler kesme puanı etrafında dağılmıştır ve her parametrede madde bulunacaktır. Bu durumda kesme puanından daha uzak güçlük derecelerine sahip maddeler daha az verimli olur. Bu nedenle GOO sınıflama kriteri güçlük parametresinin kesme puanıyla yakından eşleşmediği maddelerin verimsizliğini gidermek için önerilmiştir (Haug, 2004; Bartroff, Finkelman ve Lai, 2008; Thompson, 2007, 2009). GOO işlevsel olarak parametrelerin belirlenmesi, olabirlik oranının hesaplanması ve sınıflandırma süreçleri olarak AOOT prosedürüne benzerdir. Farkları θ_1 ve θ_{02} AOOT için sabitken GOO da değişmesine izin verilir. AOOT de madde güçlüğü belirtilen farksızlık bölgesinin dışında kalabilir. GOO sınıflama kriteri olabirlik fonksiyonunun maksimumunun farksızlık bölgesi içinde yer almadığında

hesaplamaya dahil edilmesini sağlar. Örneğin kesme puanı 0 olsaydı ve simetrik bir farksızlık bölgesi ile $\theta_1 = -0.2$ ve $\theta_2 = 0.2$ maddenin güçlüğü 0.5 ise GOO sınıflama kriteri prosedürü olabilirlik fonksiyonunun maksimum değeri önceden belirlenen 0.2 yerine 0.5 olur. Başka bir ifadeyle olabilirlik fonksiyonunun maksimumu sınırların altında veya üstünde olduğunda farksızlık bölgesi sınırların değerlerini o tek madde için olabilirlik fonksiyonunun maksimumunu içerir. Thompson (2007, 2009) GOO sınıflama kriteri prosedürünün hiçbir zaman daha düşük performansın olmayacağını ortalama test uzunluğu açısından AOOT sınıflama kriterinden bazı durumlarda performansın daha iyi olduğunu ortaya koymuştur. Bu araştırmada GOO sınıflama kriterinin performansının BBST çalışmalarında fazla araştırılmış olmaması ve AOOT'nin modifiye edilmiş daha genel hali olması yani AOOT' nin bahsedilen dezavantajını ortadan kaldırması nedeniyle ve farksızlık bölgesi tolere edilebilir hatalar için AOOT ile benzer değer olan 0.10 değeri ile performansı araştırılmıştır.

İlgili Araştırmalar

Demir (2019) doktora tezinde iki kategorili puanlanan 500 maddeden oluşan madde havuzunu ve 5000 bireye ait madde parametrelerini R yazılımı kullanarak simülasyonla oluşturmuştur. İki, üç, dört kategoride yapılan BBST için AOK yetenek kestirimi ile OTU, OSD ve RMSE, OMH, gerçek yeteneklerle kestirilen yetenekler arası korelasyon ve yanlılığın AOOT ve GA sınıflama kriterleri ile MFB-KY ve MFB-KN madde seçme yöntemlerine göre içerik dengeleme ve madde kullanım sıklığı kontrolü gibi pratik kısıtlamalara göre nasıl değiştiğini araştırmıştır. Araştırmasında 48 koşul oluşturulmuştur. Tüm koşullar için 30 tekrarın ortalaması aile hesaplamalar yapılmıştır. Güven Aralığı sınıflama kriterinin ortalama test uzunluğu için Ardışık Olasılık Oran Testi sınıflama kriterinin ise ortalama sınıflama doğruluğu için başarılı olduğu sonucuna ulaşmıştır. MFB-KY madde seçme yönteminin performansının MFB-KN madde seçme yönteminden daha yüksek olduğu anlaşılmıştır. İçerik dengesi için Modifiye Edilmiş Multinomial Model yönteminin başarılı olduğu, Kısıtlanmış Bireyselleştirilmiş Bilgisayarlı Test yönteminin test etkililiği için daha başarılı olduğu sonucuna ulaşmıştır. Madde Uygunluk yönteminin, Sympson- Hetter'e

göre madde kullanım sıklık yöntemi olarak daha iyi performans gösterdiği araştırmada belirtilmiştir.

Gündeğer (2017) doktora tezi çalışmasında BBST uygulaması için iki kategorili sınıflama yapılmıştır. AOOT, GA, GOO üç sınıflama kriterleri ile BSD, AOK olarak iki yetenek kestirimi ve KLB-MFB (KY-KN) temelli dört madde seçme yöntemi için oluşturulan 96 koşulda OTU, OSD, yanlılık, kestirimin standart hataları, kestirilen yetenekler ile gerçek yetenekler arası korelasyon değerlerini incelemiştir. Araştırmada simülasyonla 3000 birey 3PLM ile 500 madde R yazılımında oluşturulmuştur. Ayrıca 994 bireye uygulanan 80 maddelik gerçek veri seti kullanılmıştır. Hem üretilen hem gerçek veri seti üzerinde iki kategorili sınıflama için 25 tekrarın ortalaması ile sonuçlara ulaşılmıştır. OTU ve OSD için GOO ile GA sınıflama kriterlerinin, kestirimin standart hata değerleri ve yanlılık için iki kategorili sınıflamada AOOT sınıflama kriterinin daha iyi performans gösterdikleri belirlenmiştir. Farksızlık bölgesi genişledikçe OTU'nun azaldığı, BSD yetenek kestiriminin kestirimin standart hatası için, gerçek yetenekler ile kestirilen yetenekler arası korelasyon için ise AOK ve BSD yetenek kestirimlerinin iyi performans gösterdikleri sonucuna ulaşılmıştır. Madde seçme yöntemlerinden kestirilen yetenek temelli Maksimum Fisher Bilgisinin daha başarılı olduğu araştırmanın sonuçlarındandır.

Demir (2022) araştırmasında madde havuzu ve madde seçim yöntemlerinin BBST performansı üzerindeki etkisini araştırmıştır. Bu amaçla 1000 bireye ait cevap örüntüsü ve 15, 300, 450, 600 maddeden oluşan farklı dağılımlara sahip 8 madde havuzu oluşturmuştur. Araştırma sonunda büyük madde havuzuna sahip madde havuzlarında ortalama test uzunluğunun daha yüksek olduğunu madde seçim yöntemlerinin etkilerinin OTU ile OSD açısından benzer olduğu sonuçlarına ulaşılmıştır. Sympon Hetter yönteminin test etkililiği açısından avantajlı olduğu, Modifiye edilmiş multinominal modelin ise içerik dengeleme için daha etkili olduğu sonuçlarına ulaşılmıştır.

Kaya (2022) araştırmasında İngilizce yeterlilik sınavına ait maddelerin BBST ve kağıt kalem testi olarak uygulanmasından elde edilen puanları sınıflandırma performansları

olarak karşılaştırmıştır. Karşılaştırma için farklı sonlandırma kurallarına göre gerçek veri ile simülasyon çalışması yapmıştır. Yetenek kestirimlerinin BBST ve kağıt kalem testlerinde benzer olduğu görülmüştür. BBST ile daha az madde sayısı ile sınıflama yapıldığı ve sabit madde sayısı kuralı ile sonuçlandırılan testlerde 20, 25, 30 madde sayılarının düşük hata değerleri oluşturduğu görülmüştür. Araştırmada sınıflandırma doğruluğu ve sınıflandırma tutarlılığı belirlenmiştir. Farklı kesme puanları kullanılmıştır. Tek kesme puanıyla yani sınıflamanın iki kategoride yapıldığı durumlarda daha yüksek sınıflandırma performansı görülmüştür. Çok kategorili sınıflamanın sınıflandırma performansının iki kategorili sınıflamaya göre biraz daha düşük olduğu sonuçlarına ulaşılmıştır.

Kingsburry ve Weiss (1980) BBST araştırmasında kağıt kalem testi ile BBST çalışmasını OTU, OSD ile kestirimin standart hataları açısından karşılaştırmıştır. Araştırmasında 100 maddelik dört madde havuzu ile 500 öğrenciye ait normal dağılımdan yetenek parametreleri simülasyonla oluşturmuştur. Madde havuzları oluşturulurken tek biçimli dağılım, b parametresinin 0.5 aralıklarla değişkenlik gösterdiği ve a ve b parametrelerinin 0.5 aralıklarla değiştiği, a, b, c parametrelerinin de 0.1, 0.2, 0.3 şeklinde değişkenlik gösterdiği dört madde havuzu oluşturulmuştur. Sınıflama kriteri olarak AOOT, BUT sınıflama kriterleri kullanılmıştır. Araştırmanın sonunda AOOT sınıflama kriterinin en az madde ile sınıflama yaptığı geleneksel kağıt kalem testinin en yüksek sayıda madde ile sınıflama yaptığı görülmüştür. Geleneksel kağıt kalem testlerinin tüm koşullar için en yüksek sayıda madde ile sınıflama yaptığı görülmüştür.

Reckase (1983) araştırmasında 1 PLM ve 3 PLM olmak üzere iki MTK modelinde AOOT sınıflama kriterinin performansını üç farksızlık bölgesi ile ortalama test uzunluğu ortalama sınıflama doğruluğu açısından iki kategorili sınıflama için araştırmıştır. Araştırmada yetenek kestirimi olarak maksimum olabilirlik kestirimi, madde seçme yöntemi için kestirilen yetenek temelli MFB ve kesme noktası olarak 0 kullanılmıştır. Araştırma sonunda farksızlık bölgesi azaldıkça test uzunluğunun arttığı ortalama sınıflama doğruluğunun fazla etkilenmediği 3 PLM'nin 1 PLM den daha başarılı olduğu anlaşılmıştır.

Spray ve Reckase (1994) arařtırmalarında madde seme yntemlerini karřılařtırmıřtır. Bu amala kesme noktasında, gerek ve kestirilen yetenek dzeylerinde en yksek bilgi veren madde seme yntemleri ortalama test uzunluęu aısından karřılařtırılmıřtır. Arařtırmada 200 maddelik gerek veri seti,  farklı kesme noktası, sınıflama iin en fazla 50 madde kullanılması, 1,0 farksızlık blgesi olarak belirlemiřtir. Birey yeteneklerini madde havuzundan 0.25 aralıklarla oluřturulmuřtur. alıřmanın sonunda kesme noktası temelli madde seme yntemlerinin kestirilen yetenek farklı bir ifadeyle gerek ve kestirilen yetenek dzeylerinde en yksek bilgiyi veren madde seme yntemine gre performansının daha iyi olduęu, daha az madde ile sınıflama yapıldıęı sonucuna ulařılmıřtır.

Spray ve Reckase (1996) arařtırmalarında iki kategorili sınıflama iin AOOT, BUT sınıflama kriterlerini kestirimin standart hata deęerleri ve madde sayıları bakımından karřılařtırmıřlardır. Simlasyonla 1000 bireye ait yetenek parametreleri ve 200 maddeye ait madde havuzu oluřturulmuřtur. Kesme noktası temelli MFB madde seme yntemi olarak kullanılmıřtır. Sınıflama iin en fazla 50 madde kullanılacak řekilde kořullar ayarlanmıřtır. alıřmanın sonunda AOOT sınıflama kriterinin daha az sayıda madde ile sınıflama yaptıęı sınıflama hatalarının benzer olduęu sonucuna ulařılmıřtır.

Lau (1996) Veri setinin tek boyutlu olmadıęı durumlarda AOOT sınıflama kriterinin iki kategorili sınıflamada performansını arařtırmıřtır. Simlasyonla iki boyutlu 1 PLM ve 3 PLM ile madde havuzu oluřturulmuřtur. En fazla 50 madde ile sınıflamanın yapılacaęı řekilde kořullar oluřturulmuřtur. Ortalama test uzunluęu ortalama sınıflama doruęu bakımından karřılařtırmalar yapılmıřtır. Arařtırma sonunda tek boyutluluk ihlal edildięinde ardıřık olasılık oran testi sınıflama kriterinin 3 PLM altında 1 PLM e gre daha etkili sonular oluřturduęu test uzunluęunun azaldıęı gzlenmiřtir. Kullanılan MTK modelinin test uzunluęu zerinde etkisinin olduęu arařtırmadan anlařılmıřtır.

Spray vd. (1997) ok boyutlu madde havuzu olması durumunda AOOT sınıflama kriterinin etkisini belirlemek iin iki boyutlu madde havuzunda simlasyon alıřması

yapılmıştır. Madde sabitlemesi yapılmadan ve en fazla 60, 120 madde ile sınıflamanın yapıldığı koşullar ile madde kullanım sıklığı kontrolü altında OTU ve OSD gibi koşulların performansı belirlenmiştir. Araştırmanın sonunda test uzunluğu arttıkça sınıflama doğruluğu artmış, yetenek düzeyi zayıf olan durumlarda madde kullanım sıklığının daha fazla etkiye sebep olduğu, ardışık olasılık oran testi sınıflama kriterinin çok boyutlu madde havuzunda yüksek sınıflama doğruluğu ile sınıflama yaptığı görülmüştür.

Lau ve Wang (1998) AOOT'nin iki kategorili ve çok kategorili gerçek veri seti üzerinde performansı araştırılmıştır. Araştırmada rastgele madde seçme yöntemi, kesme noktası temelli madde seçme yöntemi ve kesme noktası temelli ile rastgele madde seçiminin birleşimi olarak üç madde seçme yöntemi kullanılmıştır. Araştırma sonunda iki kategorili çok kategorili madde havuzlarında kesme noktası temelli madde seçme yönteminin performansının ortalama test uzunluğu açısından en iyi olduğu ve ardışık olasılık oran testinin çok kategorili madde havuzlarında da çok iyi performans gösterdiği sonucuna ulaşılmıştır.

Eggen (1999) 250 maddelik gerçek veri seti ile BBST uygulamasında iki, üç kategoride yapılan sınıflama için kestirilen yetenek ve kesme noktası temelli MFB ve KLB madde seçme yöntemlerinin performansını ortalama test uzunluğu ve ortalama sınıflama doğruluğu için 5000 tekrar ile araştırmıştır. Araştırmanın sonunda KLB ve MFB madde seçme yöntemlerinin OTU ve OSD için benzer performans gösterdiği görülmüştür. KN temelli madde seçme yöntemlerinin KY temelli madde seçme yöntemlerine göre daha başarılı olduğu anlaşılmıştır.

Lau ve Wang (1999) araştırmalarında AOOT'nin performansını çok kategorili madde havuzunda MFB be KLB madde seçme yöntemleri, madde kullanım sıklığı kontrol yöntemleri ile gerçek veri seti üzerinde en düşük 3 en yüksek 30 madde ile sınıflama yapılacak şekilde koşullar oluşturularak ortalama test uzunluğu, ortalama sınıflama doğruluğu olarak sonuçlar belirlenmiştir. Araştırmanın sonunda madde seçme yöntemleri arasında ortalama test uzunluğu ve kestirimin standart hataları yönünden çok fark

bulunmadığı, madde kullanım sıklık kontrol yöntemi olmadığı durumda düşük madde sayısı ile yüksek doğrulukta sınıflama yapıldığı, farksızlık bölgesi değeri arttıkça hatanın yükseldiği madde sayısının azaldığı gibi sonuçlar elde edilmiştir.

Eggen ve Straetmans (2000) yaptıkları çalışmada sınıflama kriterlerinin performansını gerçek veri seti ile üç kategorili sınıflama için incelemiştir. 250 maddelik madde havuzu 1198 bireye uygulanmış ve 2 PLM ile kalibre edilmiştir. Gerçek veri ile yapılan çalışmada ardışık olasılık oran ve güven aralığı olmak üzere iki sınıflama kriteri, tesadüfi madde seçme yöntemi ile MFB (KY-KN) madde seçme yöntemleri olmak üzere üç madde seçme yöntemi ile içerik dengeleme ve madde kullanım sıklığı kontrol yöntemleri de araştırmada koşul olarak kullanılmıştır. Araştırmada yetenek kestirimi olarak AOK kullanılmıştır. Gerçek veri seti ile yapılan simülasyon çalışmasında maksimum 25 madde belirlenmiştir. Araştırmanın sonunda AOOT sınıflama kriterinin GA sınıflama kriterinden test etkililiği bakımından yüksek performans gösterdiği kestirilen yetenek temelli MFB madde seçme yönteminin daha az sayıda madde ile, kesme noktası temelli madde seçme yönteminin daha düşük standart hatalarla sınıflama yaptığı görülmüştür. Çalışmada içerik dengeleme yönteminin madde kullanım sıklığı kontrol yönteminden hata bakımından yüksek performans gösterdiği anlaşılmıştır.

Lin ve Spray (2000) araştırmalarında AOOT sınıflama kriterinin performansını MFB, KLB ve Log Odds oranı üç madde seçme yöntemine göre araştırmışlardır. Araştırmalarında gerçek maddelerden oluşan madde havuzu ve madde kullanım sıklığı kontrol yöntemleriyle birlikte OTU ve OSD açısından incelemiştir. Normal dağılımdan yetenek düzeyleri 10.000 birey için türetilmiş ve tüm koşullar için benzer sonuçlar elde edilmiştir.

Thompson (2007) BBST için çok kategorili maddeleri kullanarak 25 tekrar ile ardışık olasılık oran testi, bireyselleştirilmiş uzmanlık testi olarak iki sınıflama kriteri, MFB (KY-KN) olarak iki madde seçme yöntemi ile iki psikometrik model, tek biçimli ve normal dağılımdan oluşturulan iki madde havuzu kullanılmıştır. Araştırmanın sonunda AOOT ve BUT sınıflama kriterlerinin MFB-KN madde seçme yöntemiyle çaprazlandığında en düşük test uzunluğu

ile sınıflamanın yapıldığı görülmüştür. Araştırmanın sonunda OTU ve OSD için standart hata değerlerinin düşük olduğu görülmüştür. Kesme noktası sayısının başka bir ifade ile sınıflama kategori sayısının ve MTK modelinin OTU ve OSD 'nu etkilediği sonucuna ulaşılmıştır.

Thompson ve Ro (2007) sınıflama kriterlerini OTU ve OSD açısından performanslarını karşılaştırmışlardır. Simülasyonla 750 maddeye ait madde parametreleri ile madde havuzu ve 10000 kişiye ait yetenek parametreleri oluşturulmuştur. Araştırmada AOOT ve GOO ile GA sınıflama kriterleri kullanılmıştır. Madde seçme yöntemleri olarak MFB kesme noktası ve kestirilen yetenek temelli madde seçme yöntemleri belirlenmiştir. Araştırmada farksızlık bölgesi küçüldükçe sınıflama doğruluğu ve test uzunluğunun arttığı görülmüştür. MFB-KN temelli madde seçme yönteminin KY madde seçme yöntemine göre daha düşük test uzunluğu ile sınıflama yaptığı görülmüştür. Araştırmanın tüm koşulları için sınıflama doğruluğu 0.93 den yüksek hesaplanmıştır.

Thompson (2009) tarafından yapılan araştırmada simülasyonla 350 ve 750 madde ile madde havuzu sivri ve basık dağılımda oluşturulmuştur. MFB (KN-KY) madde seçme yöntemleri ve AOOT,GA sınıflama kriterlerinin performansları ortalama test uzunluğu ortalama sınıflama doğruluğu bakımından araştırılmıştır. Araştırmada 10000 birey için yetenek düzeyleri normal dağılımla oluşturulmuştur. 750 maddelik sivri dağılımda madde havuzu kullanıldığında en düşük madde sayısı ile en yüksek sınıflama doğruluğu ile sınıflama yapıldığı görülmüştür. MFB-KY madde seçme yönteminin GA sınıflama kriteri ile ortalama test uzunluğu olarak, basık madde havuzlarında ise ortalama sınıflama doğruluğu olarak iyi performans gösterdiği görülmüştür. MFB-KN madde seçme yönteminin basık madde havuzlarında GA sınıflama kriteri ile çaprazlandığında performansının yüksek olduğu görülmüştür.

Thompson (2011) BBST uygulamasında GOO, AOOT, GA sınıflama kriterleri ile simülasyonla ürettiği 50, 100, 200 maddelik madde havuzlarını kullanarak 0.20 ve 030 farksızlık bölgelerini kullanıp karşılaştırmıştır. Thompson (2011) başka bir simülasyon

çalışmasıyla 200 maddelik tek bir madde havuzunda GOO ve AOOT sınıflama kriterlerini 0.10 farksızlık bölgesi ile karşılaştırmıştır. Araştırmalarda simülasyonla 1000 birey için yetenek parametresi oluşturmuştur. Sonuç olarak GOO sınıflama kriterinin diğer iki sınıflama kriteri kadar performans gösterdiği ayrıca AOOT sınıflama kriterinden daha yüksek performans gösterdiği sonucuna ulaşılmıştır. Farksızlık bölgesi yüksek olduğunda testin yüksek sınıflama doğruluklarıyla sonlandığı görülmüştür.

Nydick vd. (2012) BBST uygulaması için AOOT, GOO ile GA ve Stokastik Azaltmalı Ardışık Olasılık Oran Testi (SA- AOOT) sınıflama kriterlerini ortalama test uzunluğu, ortalama sınıflama doğruluğu açısından üç ve beş kategorili sınıflamalar için performanslarını karşılaştırmıştır. Araştırmada 600 maddelik sekiz boyutlu gerçek veri seti 3PLM ile madde parametreleri belirlenmiştir. Çalışmanın en az 8 en çok 21 madde ile kullanılacak şekilde koşullar oluşturulmuştur. Araştırmanın sonunda GOO sınıflama kriteri ile SA-AOOT sınıflama kriterlerinin benzer performans gösterdiği fakat GOO sınıflama kriterinin uygulayıcıların daha kolay uygulayacağı yöntem olması karmaşık olmaması nedeniyle önerildiği görülmüştür.

Huebner (2012) araştırmasında madde kullanım sıklık oranı, madde kullanım sıklık kontrol yöntemleri, sınıflama doğruluğu ve test uzunluğu açısından fazla kullanılan maddelerin kullanım sıklığı ve oranı olarak ardışık olasılık oran testi sınıflama kriteri ile karşılaştırmıştır. 1000 bireye ait yetenek kestirimleri simülasyonla oluşturulmuştur. Araştırmanın sonunda madde kullanım sıklık yöntemlerinin sınıflama doğruluğu ve test uzunluğu için benzer performans gösterdikleri sonucuna ulaşılmıştır. Testin zorluk düzeyi yükseldikçe sınıflamak için test uzunluğu ve sınıflamanın doğruluğunun arttığı anlaşılmıştır.

Van Groen ve diğerleri (2016) çok boyutlu BBST ile tek boyutlu BBST için OTU ve OSD açısından performanslarını ortaya koymuştur. AOOT sınıflama kriteri ile birlikte madde seçim yöntemi için KY ve KN temelli madde seçme yöntemleri kullanılmıştır. Çok boyutlu BBST de sınıflama yönteminin verimliliğini ve etkinliğini göstermek amaçlanmıştır. Çalışma sonunda çok boyutlu maddeler için AOOT sınıflama kriterinin farksızlık bölgesi genişledikçe

OTU azalmıştır. Araştırmada çok boyutlu MTK ile uygulanan AOOT sınıflama kriterinin daha uzun testlerle sınıflama yaptığı ve daha doğru sınıflandırma sonuçlarının olduğu belirlenmiştir.

Huebner ve diğerleri (2021) BBST için iki kategorili sınıflama için araştırmanın çok olduğunu çok kategorili sınıflama için simülasyonla AOOT ve GOO sınıflama kriterlerinin performansını incelemişlerdir. Birden fazla kesme noktası olduğunda AOOT testini çoklu hipotez durumuna genişleten çalışmaların az olduğu belirtilmiştir. Araştırmada Modifiye edilmiş GOO sınıflama kriterinin çok kategorili sınıflamalarda sınıflama doğruluğundan ödün vermeden daha kısa ortalama test uzunluğu ile sınıflama yaptığı daha iyi performans gösterdiği sonucuna simülasyonla ulaşılmıştır.

Liu ve diğerleri (2022) Çokboyutlu Bilgisayarda Bireyselleştirilmiş Sınıflama Testi (ÇBBST) üzerine araştırma yapmışlardır. Çalışma tek boyutlu BBST araştırmalarına göre daha verimli bulunmuş olsa da aralarındaki farkın ihmal edilebilir düzeyde olduğu görülmüştür. Çalışmada sınıflandırma kriteri olarak AOOT sınıflama kriteri kullanılmıştır. Her bir boyut üzerinde bir sınıflandırma kararı vermek için yalnızca o boyuta yüklenen maddeler kullanılarak olasılık oranı hesaplanmıştır. Araştırmada iki boyutlu ve iki kategorili simülasyon ile ÇBBST kullanılmıştır. Gizil özelliğin koşullu dağılımı ile Ardışık Olasılık Oran Testi (AOOT-K) sınıflama kriterinin ölçüm verimliliği AOOT sınıflama kriteri ile karşılaştırılmıştır. AOOT-K sınıflama kriteri ile İlk olarak, diğer boyutların gizil özellik dağılımları, o anda sınıflandırılmakta olan boyutun üst ve alt sınırlarına bağlı olarak tahmin edilir. Koşullu gizil özellik dağılımı önceden hesaplanır ve daha sonra bu bilgi AOOT-K' de dikkate alınır. Boyutlar arasında 0.0, 0.5, 0.8 olarak sıfır orta ve yüksek korelasyonlar belirlenmiştir. AOOT-K sınıflama kriteri ile AOOT sınıflama kriteri bağımsız değişkenlerdir. OTU, OSD bağımlı değişkenlerdir. Boyutlar arasındaki korelasyon arttıkça AOOT-K nin AOOT den daha iyi performans göstermesi beklenmiştir. Sonuç olarak testin boyutları arasında en azından orta derecede ilişki olduğunda AOOT ye kıyasla test uzunluğunun ölçüm verimliliğinin önemli ölçüde arttığı görülmüştür.

İlgili Araştırmalar Özet

BBST ile ilgili araştırmaların çoğunlukla yurt dışında çalışılmış olduğu görülmektedir. Araştırmalar genelde iki kategorili sınıflama üzerine yapılan çalışmalardır (Gündeğer, 2017; Kingsbury & Weiss,1980; Lau, 1996; Reckase,1983; Spray & Reckase,1996; Thompson, 2011;). Araştırmaların çoğunluğu madde seçme yöntemlerinin karşılaştırılması (Eggen, 1999; Lin & Spray, 2000; Spray & Reckase,1994) yetenek kestirim yöntemlerinin karşılaştırılması ve sınıflama kriterlerinin AOOT ve BUT olarak karşılaştırılması ve simülatif veri ile yapılan çalışmalara (Eggen, 1999; Lin & Spray, 2000) rastlanmaktadır. Çok kategorili sınıflama için simülasyon verisi ile pratik kısıtlamaların etkisi, çok boyutlu veri seti üzerinde üç ve beş kategorili sınıflama örnekleri Nydick vd. (2012), Eggen ve Straetmans, (2000); Demir, (2019) de bulunmaktadır. Genelde sınıflama kriterlerinden AOOT sınıflama kriterinin performansı, yetenek kestirim yöntemlerinden en temel yöntem olan AOK'un performansı gibi araştırmalar yapılmıştır. Bu araştırmada geniş madde havuzu ve gerçek veri kullanılmıştır. Çok kategorili sınıflama için sınıflama kriterleri madde seçme yöntemleri, yetenek kestirim yöntemlerinden oluşan fazla sayıda koşul çaprazlanarak araştırma yapılmıştır. Özellikle MOK yetenek kestirimi, GOO sınıflama kriterlerinin KY ve KN temelli tüm madde seçme yöntemlerinin diğer yöntemlerle çaprazlandığı desenlerin performansı çok kategorili sınıflama için belirlenmesinin önemli olduğu düşünülmektedir.

Bölüm 3

Yöntem

Üçüncü bölümde araştırmanın yöntemi, verilerin elde edilmesi, verilerin MTK koşullarına uygunluğu ve verilerin analizine yer verilmiştir.

Araştırma Yöntemi

Araştırmada 220K161 numaralı “ İngilizce hazırlık sınıflarında Bilgisayarda Bireyselleştirilmiş Test Kullanımı ” başlıklı Tübitak projesi kapsamında Bilkent Üniversitesinde uygulanan İngilizce seviye belirleme sınavına ait okuma maddelerini içeren gerçek veri ile simülasyon olarak Bilgisayarda Bireyselleştirmiş Sınıflama Testi (BBST) koşullarının performansının iki, üç, dört kategorili sınıflamada sınıflama doğruluğu, test uzunluğu, yanlılık, korelasyon, RMSE ile ortalama mutlak hata açısından incelenmesi amaçlanmıştır. Araştırma betimsel araştırma ve gerçek veri seti ile yapılan simülasyon çalışmasıdır. Betimsel araştırmalar bir durumu tanımlayan açıklayan özetleyen araştırmalardır (Büyüköztürk ve diğerleri, 2008). Kaptan ‘a (1995) göre betimsel araştırmalar mevcut durumları ortaya koyar ve daha önceki olaylarla da ilişkisini dikkate alarak açıklamayı hedefler. Simülasyon çalışmaları ise gerçek uygulama olsaydı sonuçlar ne olurdu? Sorusuna cevap aramaktadır. İstatistiksel yöntemlerin performansının araştırıldığı çalışmalardır (Dooley, 2002).

Verilerin Elde Edilmesi

Bu araştırmada 220K161 numaralı Tübitak projesi kapsamında okuma ve dinleme maddelerinden oluşan İngilizce seviye belirleme sınavı oluşturulmuştur. Toplam 336 okuma, 174 dinleme maddelerinden oluşan 9 set bulunmaktadır. Bilkent Üniversitesi Yabancı Dil Yüksekokulu hazırlık sınıflarında okuyan 1020 öğrenciye İngilizce seviye belirleme sınavına ait dokuz set farklı zamanlarda uygulanmıştır. Veri temizleme işlemi yapıldıktan sonra toplam 919 kişiye uygulanmış verilerin okuma maddeleri üzerinde araştırma yapılmıştır. Kağıt kalem formatında uygulanmış İngilizce seviye belirleme sınav

sonuçları kullanılarak gerçek uygulamada tek boyutlu iki, üç ve dört kategoride sınıflama için BBST algoritmasının belirlenmesi amacıyla literatürde bulunan BBST yöntem ve tekniklerinin gerçek veri üzerinde simülasyonla araştırması gerçekleştirilmiştir. Gerçek veri ile gerçekleştirilen simülasyonda bir manipülasyon uygulanmamıştır.

Araştırmanın amacına bağlı olarak verilerin analizi üç aşamada yapılmıştır. Birinci aşamada BBST için madde havuzunun oluşturulmasına yönelik analizler yapılmıştır. İkinci aşamada madde havuzunun madde tepki kuramı analizleri yapılmıştır. Üçüncü aşamada gerçek veri setinden oluşturulan madde havuzundan 108 koşul için gerçek zamanlı tek boyutlu iki, üç ve dört kategorili sınıflama için en uygun BBST koşulları belirlenmiştir.

Madde Havuzunun Oluşturulması

Madde havuzunun oluşturulması için Bilkent Üniversitesinde farklı zamanlarda uygulanan dokuz setten oluşan İngilizce seviye belirleme sınavı verileri kullanılmıştır.

Tablo 3

Veri Setlerinde Bulunan Madde Sayıları

Setler	Okuma Madde sayısı	Dinleme Madde sayısı
Set 1	37	19
Set 2	36	21
Set 3	38	19
Set 4	37	21
Set 5	37	21
Set 6	38	18
Set 7	39	18
Set 8	37	19
Set 9	37	18
Toplam	336	174

Madde havuzunun için dokuz sette bulunan madde sayıları Tablo 3'te gösterilmektedir. Toplam 336 okuma 174 dinleme maddesi bulunmaktadır. Araştırmada okuma maddelerinin 10 tanesi ortak maddelerdir. Toplamda 256 okuma maddesi bulunmaktadır. Geniş bir madde havuzu oluşturmak amaçlandığı için zaman ve motivasyon nedeniyle madde havuzunda yer alan 256 maddenin tamamı aynı gruba ya da aynı zamanda uygulanamamıştır. MTK'nın değişmezlik özelliği varsayımına göre madde sayısı çok olduğunda aynı evrenden farklı gruplara bu maddeler farklı zamanlarda da uygulanabilir (Kolen & Brennan, 2004). MTK'nın değişmezlik özelliği kullanılarak aynı evrenden farklı gruplara bu maddeler uygulanmıştır. Farklı gruplara yapılan uygulamalarda testlerin ölçeklenebilmesi için ortak maddeler kullanılmalıdır (Nydick & Weiss, 2009).

Öğrencilerin tamamı her setteki maddeleri yanıtlamadıkları için yanıt birey matrisi eksik matristir. Tam matrisi elde etmek için eksik yanıt matrisindeki gerçek veri setinden madde ve yetenek parametreleri " anchor " maddeler de kullanılarak tüm setleri aynı ölçek düzeyine getirmek için R yazılımında mirt (Chalmers, 2012) paketi kullanılmıştır. Dikey ölçekleme yöntemlerinden eş zamanlı kalibrasyon ile setlerdeki maddeler on ortak madde ile birbirine bağlanmış eş zamanlı olarak madde parametreleri coef kodu ile yetenek parametreleri de aynı pakette fscores kodu ile bir parametrelili lojistik model, iki parametrelili lojistik model ve üç parametrelili lojistik modele göre kalibrasyon ile üretilmiştir. Lau (1996) literatürde 1 PLM ile araştırılan BBST'nin test etkililiğinin (OTU, OSD) diğer modellere göre daha düşük olduğunu belirtmiştir. Reckase (1983) de çalışmasında 1PLM ile araştırılan BBST'de kullanılan madde sayısının daha fazla olduğunu belirtmiştir. Araştırmada 3 parametrelili lojistik model ile kalibrasyon yapıldığında 256 maddenin yaklaşık 40 maddesinin a parametrelerinin kestiriminde yakınsama problemi yaşanmıştır. MTK modellerinde 3PL kullanıldığında parametre kestirimleri için yakınsama problemi yaşanabilmektedir (Eckes, 2014). Çalışmada madde havuzunu daha etkin kullanabilmek amacıyla yakınsama problemi yaşanan 40 madde madde havuzundan çıkarılmamıştır. 2PLM ile kalibrasyon yapılmıştır. Kalibrasyon, parametreleri dönüşüm ile aynı ölçeğe yerleştirme sürecidir

(Kolen, 2004). Farklı gruplara uygulanan testlerden elde edilen verileri aynı ölçeğe yerleştirmek için farklı kalibrasyon yöntemleri kullanılmaktadır. Eş zamanlı ve ayrı kalibrasyon yöntemleri en çok kullanılan yöntemlerdir. Eş zamanlı kalibrasyonda farklı sınıf düzeylerindeki verilerin aynı anda kalibre edilmesiyle dikey ölçek oluşturulmaktadır (Meng, 2007). Bu yöntem ile eş zamanlı olarak parametreler ölçeklendiğinden parametre dönüşümlerine ihtiyaç duyulmadan parametreler aynı ölçek üzerinde bulunur (Kim ve diğerleri, 2009). Araştırmada MTK ile ölçekleme yöntemi kullanılarak eş zamanlı (birlikte) kalibrasyon ile madde ve yetenek parametreleri aynı ölçekte elde edilmiştir. 2PLM ile kalibrasyon yapılan 919 bireye ait yetenek parametreleri 2.72 ile -2.29 arasındadır. Ayırt edicilik (a) parametresi değerleri 0.02 ile 2.9 arasında, madde güçlük (b) parametresi 2.9, -2.6 değerleri arasında belirlenmiştir.

Araştırmada İngilizce seviye belirleme sınavına katılan bireyler alt alta sıralandığında her biri ilk on ortak madde ile birlikte bir madde setini yanıtlamıştır. Bu nedenle yanıt matrisinde diğer sekiz madde setini yanıtlamadıkları için eksik yanıt matrisi oluşmuştur. Gerçek veriye dayalı simülasyon uygulaması için her bireyin kalan sekiz madde setine verdikleri yanıtlar ölçekleme sonucu elde edilen madde setlerindeki maddelerin parametreleri ve bireylerin yanıtladıkları madde setinden elde edilen yetenek kestirimleri göz önüne alınarak yanıtlamadıkları madde setleri için mirt (Chalmers, 2012) paketi kullanılarak " simdata " fonksiyonu ile yanıt oluşturulmuştur. Tablo 4'te bireylerin yabancı dil testindeki setlere ait eksik veri matrisi bulunmaktadır. Tabloda doksan hücre bulunmaktadır. İlk dokuz hücredeki on maddeyi tüm sınav katılımcıları yanıtlamıştır. Kalan 81 hücrede "X" simgesi bulunan hücreler satırdaki bireylerin sütundaki madde setini yanıtlamış olduğu anlamına gelir. Diğer 72 hücre için yanıt üretme işlemi gerçekleştirilmiştir. Elde edilen eksiksiz yanıt matrisi kullanılarak madde tepki kuramı analizleri ve gerçek veriye dayalı simülasyonlar gerçekleştirilmiştir.

Tablo 4*Eksik Veri Matrisi*

	ilk 10 madde	27 madde	26 madde	28 madde	27 madde	27 madde	28 madde	29 madde	27 madde	29 madde	Toplam 256
Set1 100 kişi	x	x									37 madde
Set2 104 kişi	x		x								36 madde
Set3 100 kişi	x			x							38 madde
Set4 100 kişi	x				x						37 madde
Set5 100 kişi	x					x					37 madde
Set6 105 kişi	x						x				38 madde
Set7 103 kişi	x							x			39 madde
Set 8 102 kişi	x								x		37 madde
Set 9 105 kişi	x									x	37 madde
Top 919 kişi											Toplam 336 made

Veri seti MTK varsayımları açısından incelenmiştir. Tek boyutluluk varsayımı R yazılımında faktör analizi ile araştırılmıştır. Yerel bağımsızlık varsayımı için bireylere ait cevap örüntüsü üzerinden maddeler arasındaki ilişkiler ve Yen'in Q3 istatistiği kullanılarak maddelerden elde edilen artıklar arasındaki korelasyon matrisi incelenmiştir. Model veri uyumu için -2Loglikelihood (-2LL), (AIC) ve (BIC) uyum istatistiklerinden yararlanılmıştır.

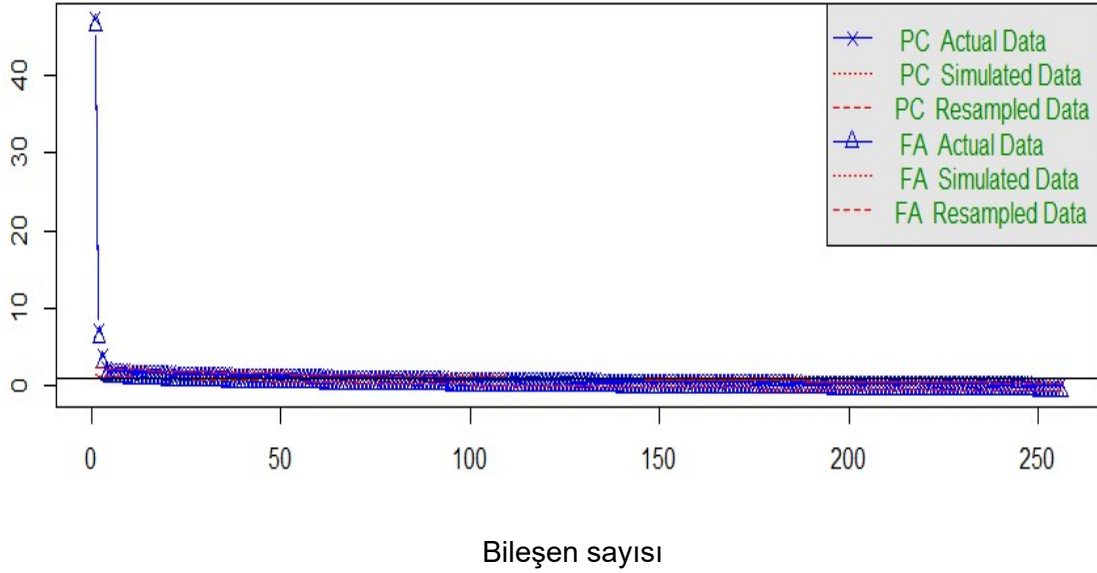
Veri Setinin Tek Boyutluluk Varsayımının İncelenmesi. MTK'nın tek boyutluluk özelliği bireyin test göstergesi ile ilgili başat bir gizil özelliğinin olduğunu belirtir (de Ayala, 2009). Bu varsayımın karşılanması için test göstergesini etkileyen baskın bir faktörün

olması tek boyutluluğun göstergesi olarak kabul edilmektedir (Hambleton & Swaminathan, 1985). Bu nedenle tek boyutluluk varsayımını test etmek için R yazılımında açımlayıcı faktör analizi yapılmıştır. Veriler iki kategorili (1-0) olduğu için açımlayıcı faktör analizi tetrakorik korelasyon matrisi oluşturularak gerçekleştirilmiştir. Faktör analizi öncesinde verinin faktör analizine uygun olup olmadığı Kaiser- Meyer- Olkin (KMO) ve Bartlett küresellik testi ile incelenmiştir. Faktör analizi öncesinde faktör analizi yapılabilmesi için KMO değerinin 0.60'dan yüksek, Bartlett testinin ise istatistiksel olarak anlamlı çıkması gerekmektedir (Tabachnick & Fidel, 2001). R yazılımında "psych" paketinde KMO fonksiyonu ile KMO değeri ve " bartlett.test " fonksiyonu ile bartlett küresellik testi yapılmıştır. Aynı pakette " fa.parallel " fonksiyonu kullanılarak paralel analiz ayrıca " scree " komutuyla faktör sayısı ve veriye ait yamaç birikinti grafiği elde edilmiştir. Hesaplanan korelasyon matrisinin bartlett testi sonuçları (Ki-Kare = 5086, sd = 255, p=0,00) ile anlamlı çıktığı görülmektedir (p<0.05). Hesaplanan KMO değeri 0.96 ile örneklem büyüklüğünün faktör analizi için oldukça yeterli olduğunu göstermektedir.

Şekil 4'te R ortamında " scree "komutuyla çizilen yamaç-birikinti grafiğine göre birinci faktörden sonra keskin düşüş olmuştur. İkinci faktörden itibaren eğim önemli ölçüde azalmıştır. İkinci faktörle birlikte diğer faktör varyansa büyük oranda katkı sağlamamıştır. Veri setinin tek boyutluluk varsayımını sağladığı görülmektedir. Açımlayıcı faktör analizi ile tek faktör için elde edilen uyum indeksleri incelendiğinde Ki-kare değeri serbestlik derecesi ile oranlanarak değerlendirmeye alınmıştır. " X^2/df (83877.46/31877)" oranı 2.63 olarak hesaplanmıştır. Büyük örneklerde X^2 /df oranınının 3'ün altında olması mükemmel uyuma; 5'in altında olması orta düzeyde uyuma karşılık gelmektedir (Kline, 2005; Sümer, 2000). Bu çerçevede elde edilen değer mükemmel uyuma işaret etmektedir. RMSEA değeri incelendiğinde 0,019 olarak bulunmuştur. RMSEA'nın 0,05'ten küçük olması mükemmel uyum, 0,005- 0,08 arasında olması iyi uyumu göstermektedir (Jöreskog & Sörbom, 1993). 0,08-0,10 arası zayıf uyumun göstergesidir (Tabachnick & Fidel, 2001). Bu nedenle elde edilen değerden yola çıkarak mükemmel uyum olduğu söylenebilir.

Şekil 4

Veri Setine Ait Yamaç Birikinti Grafiği



CFI değeri incelendiğinde değer, 0.99 olarak hesaplanmıştır. CFI değerinin 0.95'in üzerinde olması mükemmel uyuma, 0.90'ın üzerinde olması ise iyi uyuma karşılık gelmektedir (Sümer, 2000). Bu nedenle CFI değerinin mükemmel uyuma karşılık geldiği görülmektedir. Son olarak SRMR değeri 0.04 olarak bulunmuştur. SRMR değerinin 0.05'in altında olması mükemmel uyumun, 0.08'in altında olması ise iyi uyumun göstergesidir (Brown, 2006). Diğer uyum indeksleri gibi bu değer de iyi uyuma karşılık geldiği söylenebilir. İki kategorili veri setinin MTK'nın lojistik modellerine ait tek boyutluluk varsayımını karşıladığı görülmüştür.

Veri Setinin Yerel Bağımsızlık Varsayımının İncelenmesi. MTK'nın önemli varsayımlarından birisi de yerel bağımsızlıktır. Bireylerin yetenekleri sabit tutulduğunda farklı maddelere verdikleri yanıtların birbirinden istatistiksel olarak bağımsız olmasıdır (Lord, 1980;akt. Hambleton & Swaminathan, 1985). Aynı örtük özelliği ölçmesi amacıyla geliştirilmiş bir teste ait iki madde arasındaki ilişkinin sadece ölçülmek istenen özellik üzerinden bir ilişkiye sahip olmasıdır. Bu özellik düzeyi sabit tutulduğunda iki madde arasındaki ilişkinin ortadan kalkması beklenir. Yerel bağımsızlık, tek boyutluluk

varsayımının sağlanmış olmasını gerektirir (Hambleton & Swaminathan, 1985). Aynı zamanda tek boyutluluk varsayımının sağlanmış olması yerel bağımsızlık için de bir kanıt olarak sayılabilir. Yerel bağımsızlığı test etmek için madde artıkları arasındaki ilişkileri gösteren Q3 istatistiği incelenmiştir. Q3 korelasyonel bir istatistiktir. Q3 hesaplanırken sınava giren tüm öğrencilerin yetenekleri tüm maddelerdeki beklenen performansını belirlemek için kestirilir. Öğrencilerin her madde için gözlenen ve beklenen performansı arasındaki fark artık değer olarak hesaplanır. Seçilen iki maddenin Q3 istatistiği, fark puanlarının bütün yanıtlayıcılardaki korelasyonudur. Q3 istatistiğinin değeri 0,20'den büyük olması ilgili madde çifti için yerel bağımsızlık varsayımının sağlanamadığına işaret etmektedir (DeMars, 2010). Değeri -1,+1 arasında değişir ve Q3 ün yüksek mutlak değeri yerel bağımsızlığın önemli ölçüde ihlal edildiğini gösterir (Peak & Cole, 2019). Q3 istatistiği için De Ayala (2009) mutlak değerinin 0.2236'nın üzerinde olmasının yerel bağımlı madde çiftlerinin göstergesi olduğunu söylemiştir. Yen 0.20'den yüksek korelasyonlara yerel bağımlı yaklaşmayı tavsiye etmiştir. Maddeler için R programında mirt (Chalmers, 2012) paketinde Q3 istatistiği incelenmiştir. Potansiyel olarak bağımlı madde çiftlerini saptamak için 256*256 matris (65.536 hücre) incelenmiştir. 0.2236'nın üzerinde Q3 istatistiklerine sahip madde çiftleri yerel bağımlı maddeler olarak belirlenmiştir. 65.536 hücre için madde çiftleri arasındaki korelasyonun çoğunluğunun anlamsız; manidar olanların (%4) düşük düzeyde ilişkili olduğu görülmüştür. Madde havuzunu oluşturan setlerde ortak on madde bulunmaktadır. Ortak köklü maddeler nedeniyle %4 oranında düşük düzeyli ilişkinin olduğu düşünülmektedir. Yerel bağımsızlık varsayımının büyük oranda sağlandığı görülmüştür. Benzer yetenek düzeyindeki bireyler için maddeler arasındaki kovaryans sıfıra eşit ise test tek boyutludur (Akt: Hambleton ve Swaminathan, 1985). Farklı bir ifade ile veri setinin anlamlı bir açıklıkla tek boyutluluk özelliği göstermesi setlerde yer alan maddelerin yerel bağımsız olduğunun diğer bir göstergesidir şeklinde yorumlanabilir.

Testin Hız Testi Olmaması. Bir testte cevaplanamayan maddeler, paralel formlar ile testi bitirememiş öğrenci sayısı testin hız testi olması hakkında bilgi verir (Hambleton &

Swaminathan). Testin hız testi olup olmadığı bireylerin maddelere cevap verirken hız ya da zamandan etkilenmesi testin yerel bağımsızlık ve tek boyutluluk özelliğini etkilediği için maddeleri cevaplayabilecekleri kadar sürenin verilmesi hız testi olmamasını sağlar. Bu durum testin sonuna ulaşamamış birey sayısı dikkate alınarak kontrol edilebilir (Hambleton & Swaminathan, 1985). Araştırmada testi cevaplayamayan birey, cevaplanmamış maddeler bulunmamaktadır. Bu sonuca dayanarak testin hız testi olmadığı yorumu yapılabilir. Tek boyutluluk yerel bağımsızlık gibi MTK'nın temel varsayımlarını olumsuz etkileyen bir durum bulunmamaktadır. Araştırmada kullanılan İngilizce seviye belirleme sınavına ait gerçek verilerin MTK'nın temel varsayımlarını karşıladığı söylenebilir. Bu aşamadan sonra veri setine ait model veri uyumu incelenerek gerçek veri setine en uygun MTK modeli belirlenmiştir.

Veri Setinin Model Veri Uyumu Varsayımının İncelenmesi. Model veri uyumunu saptamak için R ortamında mirt (Chalmers, 2012) paketinde fit fonksiyonu ile 1 PLM, 2 PLM ve 3 PLM'ye göre veri seti analiz edilmiştir. Parametre ve yetenek kestirimlerinin hatasız olabilmesi adına kullanılan verinin tercih edilen modelin varsayımlarını karşılar nitelikte, yani model ile uyumlu olması gereklidir.

Tablo 5

1 PLM İle 2 PLM' İçin Model Veri Uyumu Karşılaştırılması

	AIC	SABIC	HQ	BIC	Loglike	X2	Df
1PLM	256856.9	257280.2	257329.9	258096	-128171.4	-8617.754	-255
2PLM	248549.1	249592.6	249691.5	251218.6	-123862.5		

Model-veri uyumunu değerlendirmede AIC, BIC,-loglikelihood uyum istatistikleri incelenerek iki model karşılaştırılmıştır. Model veri uyumunun iyiliği için AIC ve BIC değerlerinin düşük - loglikelihood değerlerinin ise mutlak değerlerinin küçük olması beklenir.-Loglikelihood değeri, verinin modelden uzaklaşma derecesini göstermektedir.

Düşük değer model veri uyumunun daha iyi olduğunu göstermektedir (Embretson & Reise, 2000).

Tablo 5'e göre -loglikelihood değerleri incelendiğinde en düşük değer (-123862.5) 2 parametrelili modelde elde edildiği görülmektedir. AIC değerleri incelendiğinde en düşük değer 2 parametrelili model için elde edilmiştir. BIC değerlerine bakıldığında en düşük değer 2 parametrelili model kullanıldığında elde edildiği görülmektedir. İki parametrelili modelin tek parametrelili modele göre daha düşük değerlere sahip olduğu veriye daha iyi uyum sağladığı görülmüştür.

Tablo 6

2 PLM İle 3 PLM İçin Model Veri Uyumu Karşılaştırılması

	AIC	SABIC	HQ	BIC	LogLik	X2	Df
2PLM	248749.1	249592.6	249691.5	251218.6	-123862		
3PLM	248895.0	249960.2	250108.7	252399.3	-123579.5	9183.813	511

Model uyum karşılaştırması 2 PLM ile 3 PLM için de yapılmıştır. Tablo 6'da bu karşılaştırmanın sonuçları görülmektedir. AIC, SABIC ve BIC model endekslerinin 2 PLM de daha düşük oldukları görülmektedir. Loglikelihood değerinin 3 PLM de mutlak değer olarak bir miktar daha düşük olduğu görülse de 2 PLM'nin uyum endeksleri olarak veriye daha uygun olduğu 3 PLM e göre veriden çok uzaklaşmadığı anlaşılmaktadır. Literatürde kalibrasyonla BBST sürecinin farklı modellerle yapıldığı çalışmalar da mevcuttur. Reckase (1983), BBST çalışmasında c parametresinin BBST üzerindeki etkisini anlamak için madde kalibrasyonu ile madde ve yetenek parametrelerini 3 PLM elde etmiş olmasına rağmen Bireyselleştirilmiş test sürecini ve olabilirlik oranını 1 PLM'i temel alarak hesaplamıştır. Lau (1996) ve Reckase (1983) e göre 1 PLM ile yapılan BBST'nin test etkililiğinin (OTU ve OSD) diğer modellere göre daha düşük olduğunu belirtmeleri ve 3 PLM ile kalibrasyon yapıldığında 256 maddenin yaklaşık 40 maddesinin a parametrelerinin kestiriminde

yakınsama problemi yaşanması madde havuzunu daha etkin kullanabilmek için 40 maddeyi çıkarmadan 2 PLM model ile analizlere devam edilmiştir. Sonuç olarak araştırmmanın Bilgisayarda Bireyselleştirmiş Sınıflama testi koşullarının belirlenebilmesi için gerçek veri setinden oluşturulan simülasyon koşullarında 2 PLM dikkate alınmıştır.

BBST Simülasyon Koşulları

Çalışmada gerçek zamanlı çok kategorili BBST algoritmasının belirlenmesi için öğrencilerin İngilizce seviye belirleme sınavına ait yanıtları ve bu maddelere ait madde parametreleri gerçek veriye dayalı simülasyonla kullanılmıştır. Her alt problem için gerçek veri ile BBST simülasyon sürecine geçilmiştir. BBST simülasyonu için Nydick (2014) tarafından yazılan CatIrt paketi ve Revelle (2020) tarafından yazılan Psych paketi kullanılmıştır. Tüm simülasyon koşulları için başlama noktası $\theta=0$ olarak belirlenmiştir. Belirlenen 108 simülasyon koşulunun tamamı için 25 tekrarın ortalaması alınarak değerler hesaplanmıştır. Yetenek kestirimi, madde seçme yöntemleri, sınıflama kriterleri, sınıflama kategori sayıları çalışmanın amacına uygun şekilde çaprazlanarak manipüle edilmiştir. Çalışmada sınıflama kriterleri AOOT (FB= 0,1), GA (%90), GOO (FB= 0,1) olarak, madde seçme yöntemleri ise MFB (KY-KN), KLB (KY-KN) olarak yetenek kestirimleri ise MOK, BSD, AOK olarak belirlenmiştir. Literatüre göre madde havuzunda testin bireye uyarlanmasıyla yanıtlayacağı madde sayısının 6-12 katı arasında madde olması gerektiği belirtilmiştir (Patsula & Steffan, 1997; Luecht, 1998). Madde havuzunda bulunan 256 maddede için gerçek uygulamaya yakın olması amacıyla minimum madde sayısı 10 maksimum madde sayısı 40 olarak belirlenmiştir. İki, üç, dört kategorili gerçek veriye dayalı BBST simülasyonunda her bir koşul için değerler 25 tekrarın ortalaması ile hesaplanmıştır.

Verilerin Analizi

Araştırmada gerçek veri seti kullanılarak farklı madde seçme yöntemleri, farklı yetenek kestirim yöntemleri ve farklı sınıflama kriterleri ile oluşturulan 108 simülasyon koşulu ile 25 tekrarın ortalaması alınarak iki, üç, dört kategorili sınıflama yapılmıştır. Araştırmanın bağımlı değişkenleri ortalama test uzunluğu (OTU), ortalama sınıflama

doğruluğu (OSD), gerçek yetenekler ile kestirilen yetenekler arası korelasyon (r), yanlılık ve RMSE (root mean squared error), ortalama mutlak hata (OMH) değerleridir. Bilkent Üniversitesi Yabancı Dil Yüksekokulunda uygulanan İngilizce seviye belirleme sınavına ait öğrencilerin gerçek yetenekleri (θ_i) ile BBST sonrası kestirilen son yetenekleri ($\hat{\theta}_i$) arasındaki tutarlılığa dair ilişki R ortamında “cor” komutuyla Pearson Momentler Çarpımı Korelasyon katsayısı olarak (r) her koşul için tekrarların ortalaması ile hesaplanmıştır. Simülasyon sonunda ortalama test uzunluğu \$cat_length koduyla hesaplanmıştır. Ortalama sınıflama doğruluğu için \$cat_categ koduyla hesaplanan sınıflar ile bireylerin yetenekleri ve kesme noktalarına göre türetilen gerçek sınıfları arasındaki uyum Cohen'in Kappa istatistiği ile hesaplanmıştır.

Yanlılık, son yetenek düzeylerinin ($\hat{\theta}_i$) gerçek yetenek düzeylerinden (θ_i) farkları toplamının birey sayısına (n) oranına eşittir (Miller & Miller, 2004).

$$\text{Yanlılık} = \frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)}{n} \quad (1)$$

RMSE, kestirilen son yetenek düzeylerinin ($\hat{\theta}_i$) gerçek yetenek düzeylerinden (θ_i) farklarının kareleri toplamının birey sayısına (n) oranının kareköküne eşittir

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2}{n}} \quad (2)$$

OMH, kestirilen son yetenek düzeylerinin ($\hat{\theta}_i$) gerçek yetenek düzeylerinden (θ_i) farklarının mutlak değerleri toplamının birey sayısına (n) oranına eşittir.

$$\text{OMH} = \frac{\sum_{i=1}^n |\hat{\theta}_i - \theta_i|}{n} \quad (3)$$

Bölüm 4

Bulgular, Yorumlar Ve Tartışma

Dördüncü bölümde araştırmanın alt problemlerine ait bulgulara ve yorumlara yer verilmiştir.

Birinci Alt Probleme Ait Bulgular Ve Yorumlar

Araştırmanın birinci alt problemde iki, üç, dört kategorili BBST uygulamasında yetenek kestirimi MOK olduğunda MFB (KN-KY) ve KLB (KN-KY) madde seçme yöntemleri ile GA (%90), GOO (FB=0.1), AOOT(FB=0.1) sınıflama kriterlerinin çaprazlandığı koşullara göre OTU, OSD, yanlılık, RMSE, OMH ve gerçek yetenekler ile kestirilen yetenekler arasındaki korelasyonun nasıl değiştiği araştırılmıştır.

Tablo 7'ye göre (MOK) yetenek kestirim yöntemi ile yeteneğin kestirildiği iki kategorili sınıflama için MFB (KY-KN) madde seçme yöntemleri ile GOO (FB:0.1) sınıflama kriterleri çaprazlandığında ortalama 17-18 madde ve 0.83-0.86 sınıflama doğruluğu, GA (%90) sınıflama kriteri ile MFB (KY-KN) madde seçme yöntemleri birlikte kullanıldığında 16-19 madde ve 0.83-0.86 sınıflama doğruluğu, AOOT (FB:0.1) sınıflama kriterleri MFB (KY-KN) madde seçme yöntemleri ile birlikte sınıflama yapıldığında ortalama 26-27 madde ve 0.84-0.87 sınıflama doğruluğu hesaplanmıştır. Buna göre MOK yetenek kestirimi ile yeteneğin kestirildiği MFB-KY madde seçme yöntemi ile GA sınıflama kriterinin çaprazlandığı iki kategorili koşullara ait sınıflamanın en az madde ile yapıldığı görülmektedir. MFB-KN madde seçme yöntemi kullanıldığında ise GOO sınıflama kriterinin testi sonlandırmak için en az madde ile sınıflama yaptığı söylenebilir. OSD açısından GA ile GOO sınıflama kriterlerinin benzer performans gösterdikleri, AOOT sınıflama kriteri yönteminin performansının yüksek olduğu anlaşılmaktadır. AOOT sınıflama kriterine göre farksızlık bölgesi ne kadar büyük olursa test kesme puanından çok uzaktaki yeteneklere sahip bireyler için o kadar kısa olacaktır. Küçük farksızlık bölgesi yüksek sınıflama doğruluğunda daha uzun testlerle sınıflama yapılacağı özelliğinden dolayı AOOT sınıflama kriterinin daha fazla madde ile daha yüksek doğrulukta sınıflama yaptığı söylenebilir.

Tablo 7

Yetenek Kestirim Yöntemi MOK Olduğunda İki, Üç, Dört Kategorili Sınıflamaya Ait Test Etkilliliği Ve Ölçme Kesinliği Değerleri

Koşullar		Bağımlı Değişkenler						
Madde Seçme Yöntemleri	Sınıflama Kriterleri	SKS	OTU	OSD	r	Yanlılık	RMSE	OMH
MFB-KY	GOO(FB:0.1)	İki	17.80	0.834	0.894	-0.125	0.584	0.393
		Üç	24.28	0.789	0.905	-0.120	0.566	0.366
		Dört	27.46	0.724	0.91	-0.119	0.557	0.358
	GA(%90)	İki	16.10	0.83	0.892	-0.12	0.581	0.397
		Üç	22.26	0.781	0.902	-0.122	0.566	0.373
		Dört	26.48	0.725	0.911	-0.117	0.551	0.355
	AOOT(FB:0.1)	İki	26.02	0.841	0.914	-0.165	0.525	0.333
		Üç	31.00	0.788	0.922	-0.166	0.503	0.313
		Dört	33.11	0.729	0.924	-0.171	0.492	0.301
MFB-KN	GOO(FB:0.1)	İki	18	0.864	0.83	0.002	0.883	0.585
		Üç	24.80	0.807	0.901	-0.147	0.612	0.393
		Dört	28.19	0.727	0.884	0.023	0.607	0.372
	GA(%90)	İki	19	0.857	0.893	-0.08	0.448	0.321
		Üç	22.92	0.802	0.897	-0.146	0.623	0.407
		Dört	28.67	0.725	0.937	-0.007	0.35	0.254
	AOOT(FB:0.1)	İki	27	0.871	0.872	-0.104	0.655	0.408
		Üç	31.24	0.808	0.915	-0.2	0.551	0.33
		Dört	33.92	0.724	0.924	-0.076	0.411	0.296
KLB-KY	GOO(FB:0.1)	İki	18	0.83	0.892	-0.127	0.603	0.408
		Üç	24.36	0.779	0.903	-0.124	0.580	0.380
		Dört	27.06	0.718	0.908	-0.123	0.573	0.37
	GA(%90)	İki	16	0.824	0.889	-0.124	0.598	0.411
		Üç	22.35	0.771	0.899	-0.126	0.587	0.390
		Dört	26.45	0.714	0.905	-0.152	0.577	0.375
	AOOT(FB:0.1)	İki	26	0.825	0.911	-0.171	0.53	0.336
		Üç	30.98	0.780	0.915	-0.124	0.518	0.321
		Dört	32.99	0.718	0.919	-0.18	0.511	0.314
KLB-KN	GOO(FB:0.1)	İki	18	0.864	0.83	0.004	0.873	0.578
		Üç	24.70	0.797	0.906	-0.124	0.596	0.391
		Dört	28.11	0.728	0.876	0.024	0.642	0.391
	GA(%90)	İki	19	0.859	0.887	-0.076	0.462	0.326
		Üç	22.93	0.79	0.904	-0.125	0.594	0.397
		Dört	28.75	0.725	0.928	-0.079	0.375	0.264
	AOOT(FB:0.1)	İki	27	0.868	0.865	-0.098	0.671	0.671
		Üç	31.28	0.8	0.924	-0.186	0.515	0.323
		Dört	33.87	0.728	0.925	-0.094	0.406	0.266

SKS: Sınıflama kategori sayısı, OTU: Ortalama test uzunluğu, OSD: Ortalama sınıflama doğruluğu, r: Gerçek ve kestirilen yetenekler arası korelasyon, OMH: Ortalama mutlak hata

Araştırmanın bu bulgusu iki kategorili sınıflama için Nydick (2012), Thompson (2009) araştırma sonuçlarıyla da benzerlik göstermektedir. Bu durumda Tablo 7'ye göre MFB-KY madde seçme yöntemi GA sınıflama kriteri ile çaprazlandığı koşullarda test etkililiğinin (OTU, OSD) daha yüksek olduğu söylenebilir. MFB-KN madde seçme yöntemi GOO sınıflama kriteri ile test etkililiğinin daha yüksek olduğu anlaşılmaktadır. GOO sınıflama kriteri güçlük parametresi kesme puanıyla eşleşmeyen daha uzaktaki maddeleri de verimli hale getirdiği için madde havuzu daha etkin kullanıldığından AOOT sınıflama kriterine göre daha az madde ile sınıflama yaptığı söylenebilir. GA sınıflama kriteri de her maddeden sonra belirlenen yetenek düzeylerini kullanarak belirlenen güven aralığını kesme puanı ile karşılaştırdığı için KY temelli madde seçme yöntemleri ile performansının daha yüksek olduğu görülmüştür.

Tablo 7'ye göre iki kategorili sınıflamada bireylerin gerçek yetenekleri ile kestirilen yetenek düzeyleri arasındaki en yüksek korelasyonun (r) KY temelli madde seçme yöntemi ile AOOT sınıflama kriteri kullanıldığında 0.914, KN temelli madde seçme yöntemi ile GA sınıflama kriteri kullanıldığında 0.893 ile en yüksek korelasyonlar elde edildiği görülmektedir. Bu bulguya göre GA ve AOOT sınıflama kriterleri GOO sınıflama kriterine göre simülasyon sonunda kestirilen yeteneklerle gerçek yetenekler arası korelasyon olarak daha iyi performans gösterdikleri anlaşılmaktadır. Tablo 7 incelendiğinde iki kategorili sınıflamada MFB (KY-KN) temelli madde seçme yöntemi AOOT sınıflama kriteri ile yanlılık açısından -0.165,-0.104 ile en küçük yanlılık değeri ile sınıflama yaptığı görülmektedir. Buna rağmen KN temelli madde seçme yöntemi için GOO sınıflama kriteri 0.002 ile KY temelli madde seçme yöntemi için GA sınıflama kriterininin -0.12 ile en yüksek yanlılık değerine sahip oldukları bu araştırmanın bulgularından anlaşılmaktadır. Yanlılık değerlerinin madde seçme yöntemi ve kullanılan sınıflama kriterine bağlı olduğu anlaşılmaktadır. Tablo 7'de kestirimin standart hatasını gösteren RMSE ve OHM değerlerine göre iki kategorili sınıflamada en yüksek RMSE 0.584-0.883 ile MFB (KY- KN) madde seçme yöntemleri ile birlikte GOO sınıflama kriteri çaprazlandığı koşullarda görülmektedir. İki kategorili sınıflama için en düşük

RMSE değeri 0.525 ile AOOT ve MFB-KY madde seçme yönteminin birlikte kullanıldığı koşulda görülmektedir. MFB-KN temelli madde seçme yöntemi kullanıldığında iki kategorili sınıflamada GA sınıflama kriteri için en düşük RMSE değeri 0.448 hesaplanmıştır. Benzer şekilde en yüksek OMH değeri 0.397 ile iki kategorili sınıflamada MFB-KY madde seçme yöntemi ile GA sınıflama kriteri birlikte kullanıldığı koşullarda gözlenmiştir. MFB-KN madde seçme yöntemi için GOO sınıflama kriteri ile çaprazlandığı koşullarda en yüksek OMH değeri 0.585 gözlenmiştir. En düşük OMH değeri 0.333 ile MFB-KY ve AOOT sınıflama kriteri çaprazlandığında, MFB-KN için GA sınıflama kriteri ile 0.321 olarak en düşük OMH değeri gözlenmiştir. Kısacası iki kategorili sınıflamada GOO sınıflama kriterinin AOOT ve GA sınıflama kriterlerine göre en yüksek RMSE ve OMH değeri ile sınıflama yaptığı anlaşılmaktadır. AOOT sınıflama kriteri ise en düşük RMSE ve OMH değerini KY temelli madde seçme yöntemi ile birlikte kullanıldığında göstermiştir. Benzer şekilde GA sınıflama kriteri KN temelli madde seçme yöntemi ile en düşük RMSE ve OMH değerleri ile sınıflama yaptığı görülmektedir. MFB-KY madde seçme yöntemi ile oluşturulan koşulların MFB-KN madde seçme yöntemi ile oluşturulan koşullara göre daha düşük RMSE ve OMH değerleri hesaplanmıştır.

Özetle iki kategorili sınıflamada test etkililiği için OTU ve OSD olarak GA sınıflama kriteri ile MFB-KY madde seçme yönteminin çaprazlanarak oluşturulduğu desenin daha etkili performans gösterdiği görülmektedir. GOO sınıflama kriteri ise MFB-KN madde seçme yöntemi ile daha etkili performans göstermiştir. Ölçme kesinliği için yani hata değerleri yanlışlık korelasyon açısından GA sınıflama kriterinin KN temelli madde seçme yöntemi ile, AOOT sınıflama kriterinin ise KY temelli madde seçme yöntemiyle sınıflama yapıldığında performansının etkili olduğu söylenebilir. GOO sınıflama kriterinin hata değerleri yüksek olduğu için ölçme kesinliği olarak düşük performans gösterdiği gözlenmektedir.

Kulback Laiber madde seçme yöntemleri için yetenek kestiriminin MOK olduğu iki kategorili sınıflamada KLB(KY-KN) temelli madde seçme yöntemlerinin gerçek veri ile BBST simülasyonunda Tablo 7'ye göre GOO sınıflama kriteri ile ortalama 18-18 madde ve 0.83-

0.864 sınıflama doğruluğuyla sınıflama yapıldığı görülmektedir. Güven aralığı (GA) sınıflama kriteri ile ortalama 16-19 madde, 0.824-0.859 sınıflama doğruluğuyla sınıflama yapıldığı anlaşılmaktadır. AOOT sınıflama kriteri KLB (KY-KN) madde seçme yöntemiyle birlikte kullanıldığında ortalama 26-27 madde ve 0.825-0.868 sınıflama doğruluğu ile sınıflama yaptığı görülmektedir. Buna göre KLB-KY madde seçme yöntemi BBST uygulamasında testi sonlandırmak en az madde ile bireyi sınıflamak için GA sınıflama kriteriyle birlikte kullanıldığı koşulların en uygun olduğu anlaşılmaktadır. KLB- KN temelli madde seçme yöntemi için GOO sınıflama kriterinin daha az madde ile ortalama sınıflama doğruluğu daha yüksek sınıflama yaptığı görülmektedir. Bununla beraber KLB-KN temelli madde seçme yöntemi KLB-KY temelli madde seçme yöntemine göre test etkililiği (OSD, OTU) açısından daha iyi performans gösterdiği söylenebilir. Bireyselleştirilmiş Bilgisayarlı Sınıflama testlerinde amaç daha az madde ile daha doğru sınıflama yapmaktır. AOOT sınıflama kriteri KN ve KY temelli iki madde seçme yöntemiyle de diğer sınıflama kriterlerine göre daha yüksek sayıda madde ile sınıflama yaptığı için yeteneğin MOK olduğu KLB-KN ve KLB-KY madde seçme yöntemleriyle performansı daha düşüktür.

Tablo 7'de bireylerin gerçek yetenek düzeyleri ile kestirilen yetenek düzeyleri arasındaki en yüksek korelasyonun yetenek kestirimi MOK olduğunda KLB-KY madde seçme yönteminde iki kategorili sınıflamada AOOT sınıflama kriteri için 0.911, GA için 0.889, GOO sınıflama kriteri için 0.892 olarak hesaplanmıştır. Bu bulguya göre korelasyon değeri olarak AOOT sınıflama kriterinin en yüksek performansı gösterdiği söylenebilir. KLB-KN madde seçme yöntemi kullanıldığında de AOOT sınıflama kriteri 0.865, GA sınıflama kriteri 0.887, GOO sınıflama kriteri ile kullanıldığında ise 0.83 korelasyon değerleri hesaplanmıştır. Bu bulguya göre GA sınıflama kriterinin iki kategorili sınıflamada KLB-KN madde seçme yöntemiyle birlikte kestirilen yeteneklerle gerçek yetenekler arasındaki korelasyon için en yüksek performansı gösterdiği anlaşılmaktadır. Tüm sınıflama kriterleri için KLB-KY madde seçme yönteminin KLB-KN madde seçme yöntemine kıyasla daha yüksek korelasyonların belirlendiği görülmektedir. Tablo 7 incelendiğinde KLB (KY-KN)

madde seçme yöntemleriyle hesaplanan yanlılığın iki kategorili sınıflamada AOOT sınıflama kriteri ile -0.171,-0.098, GA sınıflama kriteriyle birlikte -0.124,-0.076, GOO sınıflama kriteriyle birlikte -0.127- 0.004 olduğu tablodan anlaşılmaktadır. Bu bulgulara göre her iki madde seçme yöntemi ile birlikte AOOT sınıflama kriterinin daha düşük yanlılığa sahip olduğu anlaşılmaktadır. Tüm sınıflama kriterleri için KLB-KY madde seçme yönteminin kullanıldığı durumlarda yanlılığın daha düşük olduğu araştırmanın önemli bulguları arasındadır. Tablo 7’de kestirimin standart hata değerlerini gösteren RMSE ve OMH değerleri için en yüksek RMSE değeri 0.603-0.83 olarak KLB (KY-KN) madde seçme yöntemleri ile GOO sınıflama kriterinin kullanıldığı koşullarda en düşük ise 0.53 olarak AOOT sınıflama kriterinin KY temelli madde seçme yöntemleriyle birlikte, GA sınıflama kriterinin KN temelli madde seçme yöntemi ile kullanıldığı koşullarda 0.462 olarak hesaplanmaktadır. Benzer şekilde en yüksek OMH 0.671 değeri ile AOOT sınıflama kriterinin KLB-KN madde seçme yöntemiyle, en düşük OMH değeri 0.326 olarak GA sınıflama kriteri ile birlikte kullanıldığında görülmektedir. KY temelli madde seçme yöntemi ile GA sınıflama kriteri birlikte 0.411 ile en yüksek, AOOT ile birlikte 0.336 ile en düşük OMH değerleri oluşturmuştur. Sonuç olarak bu bulgulara göre iki kategorili sınıflamada KLB-KN madde seçme yönteminin KLB-KY madde seçme yöntemine göre daha yüksek RMSE ve OMH değerlerine sahip olduğu anlaşılmıştır. GA sınıflama kriteri ile KN temelli madde seçme yönteminin birlikte kullanıldığı koşullarda en düşük RMSE ve OMH değerleri hesaplanmıştır.

Tablo 7’de iki kategorili sınıflama için kestirilen yetenek temelli KLB ve MFB her iki madde seçme yönteminde GA sınıflama kriterinin daha az madde ile sınıflama yaptığı, OSD olarak yakın performanslar görünse de MFB ile daha yüksek performansa sahip olduğu anlaşılmaktadır. GOO sınıflama kriteri ise MFB ve KLB her iki madde seçme yöntemi ile KN temelinde daha yüksek test etkililiğine sahip olduğu farklı bir ifade ile yüksek OSD ve düşük OTU ile sınıflama yaptığı anlaşılmaktadır. Kestirilen yetenek ve gerçek yetenekler arası korelasyon için AOOT sınıflama kriterinin MFB-KY madde seçme yöntemi ile oluşturduğu

koşulun yüksek korelasyon ile sınıflama yaptığı görülmektedir. Yanlılık olarak AOOT sınıflama kriterinin her iki madde seçme yöntemi için de KY temelli yöntemle çaprazlandığında en düşük olduğu, hata değerleri bakımından GA sınıflama kriterinin MFB-KN ile en düşük hata ile sınıflama yaptığı görülmektedir.

Üç kategorili sınıflama için yetenek kestirimi MOK olduğunda MFB (KY-KN) madde seçme yöntemlerinde test etkililiği için başka bir ifade ile OSD yüksek OTU düşük olduğu sınıflama için sınıflama kriterlerinin performansı incelendiğinde KY-KN temelli madde seçme yöntemleri ile GOO sınıflama kriteri birlikte kullanıldığında ortalama 24-24 madde ile 0.789-0.807 sınıflama doğruluğu ile GA sınıflama kriteri ile ortalama 22-22 madde ile 0.781-0.802 sınıflama doğruluğu ile AOOT sınıflama kriteri ile ortalama 31-31 madde 0.788-0.802 sınıflama doğruluğu ile sınıflama yapılmıştır. Bu bulgular doğrultusunda test etkililiği için GA sınıflama kriterinin MFB (KY-KN) her iki madde yöntemi ile birlikte en düşük madde ve en yüksek OSD ile üç kategorili sınıflama için en etkili sınıflama yöntemi olduğu görülmektedir. Tablo 7'ye göre üç kategorili sınıflamada KN temelli madde seçme yöntemi KY temelli madde seçme yöntemine göre sınıflama doğruluğu olarak daha iyi performans göstermiştir. Tablo 7'de üç kategorili sınıflamada simülasyon sonucu kestirilen yetenekler ile gerçek yetenekler arası korelasyon (r) değerleri KY ve KN madde seçme yöntemleriyle GOO sınıflama kriteri birlikte kullanıldığında 0.905-0.901, GA sınıflama kriteri ile 0.902-0.897, AOOT sınıflama kriteriyle 0.922-0.915 korelasyon değerleri görülmüştür. Bu bulgulara göre AOOT sınıflama kriteri her iki madde seçme yöntemi ile kullanıldığında GA ve GOO sınıflama kriterlerine göre en yüksek korelasyon değerleri hesaplandığı söylenebilir. GA sınıflama kriterinin ise en düşük korelasyon değerleri gösterdiği görülmektedir. Tablo 7 incelendiğinde GOO sınıflama kriteri KY ve KN madde seçme yöntemleri ile birlikte kullanıldığında hesaplanan yanlılık -0.120,-0.147, GA sınıflama kriteri kullanıldığında -0.122, -0.146, AOOT sınıflama kriteri kullanıldığında -0.166, -0.2 değerleri olduğu tablodan anlaşılmaktadır. Bu bulguya göre AOOT diğer sınıflama kriterlerinin her iki madde seçme yöntemi ile en düşük yanlılık değerlerine sahip olduğu anlaşılmaktadır. Üç

kategorili sınıflama için KN temelli madde seçme yöntemi kullanıldığında tüm kriterler için KY temelli madde seçme yöntemine göre daha düşük yanlılık değerlerinin hesaplandığı görülmektedir. Tablo 7'de üç kategorili sınıflamada yanlılıkla birlikte kestirimin standart hatasını da dikkate alan RMSE ve OMH değerleri için GOO sınıflama kriteri KY ve KN temelli madde seçme yöntemleri ile 0.566-0.612 RMSE, 0.366-0.393 OMH değerleri, GA sınıflama kriteri kullanıldığında 0.566-0.623 RMSE değerleri, 0.373-0.407 OMH değerleri, AOOT sınıflama kriterleri kullanıldığında 0.503-0.551 RMSE değerleri ve 0.301-0.296 OMH değerleri tabloda hesaplanan değerlerdendir. Bu bulgulara göre ölçme kesinliğinin yüksek olduğu yani yanlılık ve kestirimin standart hata değerlerinin düşük olduğu sınıflama AOOT sınıflama kriterinin her iki madde seçme yöntemiyle kullanıldığında en düşük hata ve yanlılık değerlerinin hesaplandığı söylenebilir. GA sınıflama kriterinin ise en yüksek hata değerleri ile üç kategorili sınıflamada sınıflama yaptığı anlaşılmaktadır.

Özetle Tablo 7'de hesaplanan değerler dikkate alınarak gerçek veri ile BBST simülasyonu için üç kategorili sınıflamada yetenek kestirim yöntemi MOK iken MFB (KN-KY) temelli her iki madde seçme yönteminde test etkililiği yani OSD ve OTU bakımından GA sınıflama kriterinin daha başarılı performans gösterdiği söylenebilir. GOO sınıflama kriterinin de KN temelli madde seçme yöntemi ile birlikte kullanıldığında GA sınıflama kriterine test etkililiği bakımından benzer performans gösterdiği araştırmanın bulguları arasındadır. Ölçme kesinliği olarak yanlılık RMSE, OMH değerleri, gerçek yetenekler ile kestirilen yetenekler arası korelasyon bakımından üç kategorili sınıflamada AOOT sınıflama kriterinin her iki madde seçme yöntemi ile kullanıldığı koşullarda daha yüksek korelasyon değerleri ve daha düşük yanlılık, RMSE ve OMH değerleri hesaplanmıştır. Bu bulgulara göre AOOT sınıflama kriterinin bireylerin son yetenek düzeylerini gerçek yeteneklerine daha yakın ve daha hassas kestirdiği yorumu yapılabilir.

Üç kategorili sınıflamada yeteneği MOK yöntemi ile kestirdiğimizde KLB (KY-KN) madde seçme yöntemleri GOO sınıflama kriteri ile birlikte kullanıldığında ortalama 24-24 madde ve 0.779-0.797 OSD ile GA sınıflama kriteri ise KLB (KY-KN) madde seçme

yöntemleri birlikte kullanıldığında ortalama 22-22 madde 0.771-0.79 OSD ile AOOT sınıflama kriteri KLB (KY-KN) temelli madde seçme yöntemleri ile 30-31 madde 0.780-0.8 OSD ile sınıflama yapıldığı anlaşılmaktadır. Bu bulgular doğrultusunda test etkililiği için en az madde ile en yüksek sınıflama doğruluğunu KLB madde seçme yöntemi ile GA sınıflama kriteri KY-KN temelli her iki madde seçme yöntemi ile birlikte kullanıldığında göstermiştir. AOOT sınıflama kriterinin test etkililiği açısından en düşük performansı gösterdiği Tablo 7'ye göre söylenebilir. Bireylerin gerçek yetenekleri ile kestirilen yetenekleri arasındaki korelasyonların (r) üç kategorili sınıflamada KLB (KY-KN) madde seçme yöntemleriyle GOO sınıflama kriteri birlikte kullanıldığında 0.903-0.906 değerleri, GA sınıflama kriteri ile 0.889-0.904, AOOT sınıflama kriteri ile 0.915-0.924 değerleri olduğu anlaşılmaktadır. Buna göre AOOT sınıflama kriterinin kullanıldığı durumlarda KLB için her iki madde seçme yöntemiyle en yüksek korelasyonların gözlemlendiği söylenebilir. GA sınıflama kriteri kullanıldığında ise en düşük korelasyonların olduğu görülmektedir. Üç sınıflama kriterleri için de KLB-KN madde seçme yönteminin dahil olduğu koşullarda KLB-KY madde seçme yöntemine göre daha yüksek korelasyonların belirlendiği görülmektedir. Tablo 7 incelendiğinde GOO ile KLB (KY-KN) yöntemleri çaprazlandığı koşullarda hesaplanan yanlılığın -0.124,-0.124 olduğu, GA sınıflama kriteri kullanıldığında -0.126,-0.125, AOOT sınıflama kriteri kullanıldığında -0.124,-0.186 olduğu Tablo 7'den anlaşılmaktadır. Buna göre yanlılık olarak üç kategorili sınıflamada KLB madde seçme yöntemi ile en düşük yanlılığı AOOT sınıflama kriteri her iki madde seçme yöntemi ile birlikte kullanıldığında hesaplanmıştır. En yüksek yanlılık GA sınıflama kriteri ile birlikte görülmektedir.

Tablo 7'de kestirimin standart hatasını gösteren RMSE ve OMH değerleri için KLB (KY-KN) madde seçme yöntemleriyle GOO sınıflama kriteri birlikte kullanıldığında 0.580 - 0.596 RMSE ve 0.380-0.391 OMH değerleri, GA sınıflama kriteri kullanıldığında 0.587-0.594 RMSE ve 0.390-0.397 OMH, AOOT sınıflama kriteriyle birlikte 0.518-0.515 RMSE, 0.321-0.323 OMH değerleri görülmektedir. Bu durumda ölçme kesinliği için en düşük hata ve yanlılık değerleri KLB (KY-KN) her iki madde seçme yöntemi için de AOOT sınıflama kriteri

ile üç kategorili sınıflama yapıldığında görüldüğü, en yüksek hata değerlerinin ise GA sınıflama kriteri kullanıldığında oluştuğu görülmektedir.

Özetle Tablo 7’de üç kategorili sınıflamada KLB (KY-KN) madde seçme yöntemleri ile test etkililiği için daha az maddeyle daha yüksek ortalama sınıflama doğruluğunun GA sınıflama kriteri kullanıldığında oluştuğu anlaşılmaktadır. AOOT sınıflama kriteri test etkililiği olarak en düşük performansı göstermiştir. Ölçme kesinliği için yanlılık ve hata değerlerinin düşük, bireylerin son kestirilen yetenekleri ile gerçek yetenekleri arasındaki korelasyonun yüksek olduğu sınıflama kriterinin AOOT sınıflama kriteri olduğu anlaşılmaktadır. Genel anlamda KN temelli madde seçme yönteminin KY temelli madde seçme yöntemine göre ölçme kesinliği bakımından daha iyi performans gösterdiği araştırmanın bulguları arasındadır. Tüm bu bulgulara göre üç kategorili sınıflamada GA sınıflama kriteri MFB-KN temelli madde seçme yöntemiyle birlikte en az madde ve en yüksek sınıflama doğruluğu ile sınıflama yaptığı için kullanılabileceği anlaşılmaktadır. Ölçme kesinliği için AOOT sınıflama kriterinin MFB-KY madde seçme yöntemi ile birlikte çaprazlanarak daha düşük hata ve yanlılık değerleri ve yüksek korelasyonla kullanılabileceği söylenebilir.

Dört kategorili sınıflama için MOK yetenek kestirimi ile MFB (KY-KN) madde seçme yöntemleri GOO sınıflama kriteri ile birlikte kullanıldığında ortalama 27-28 madde ve 0.724 -0.727 OSD ile sınıflama yapıldığı görülmektedir. GA sınıflama kriteri ile MFB (KY-KN) madde seçme yöntemleri birlikte ortalama 26-28 madde, 0.725-0.727 OSD ile sınıflama yapıldığı, AOOT sınıflama kriteri kullanılırsa MFB (KY-KN) madde seçme yöntemleriyle 33-33 madde ve 0.729-0.724 OSD ile sınıflama yapıldığı görülmektedir. Bu bulgulara göre MFB için GA sınıflama kriteri kullanıldığında her iki madde seçme yöntemi ile daha az madde ile daha yüksek sınıflama doğruluğu ile sınıflama yapıldığı söylenebilir. AOOT sınıflama kriteri kullanıldığında ise en düşük test etkililiği görülmektedir. KN temelli madde seçme yönteminin performansının başka bir ifade ile bireyleri sınıflamak için KN’ı temel alan madde seçme yönteminin KY temelli madde seçme yöntemine göre daha yüksek olduğu bulgulardan anlaşılmaktadır. Tablo 7’de bireylerin gerçek yetenekleri ile simülasyonla

kestirilen yetenekleri arasındaki korelasyonun yüksek olması beklenir. GOO sınıflama kriteri ile MFB (KY-KN) madde seçme yöntemleri karşılaştırıldığında 0.91-0.884 korelasyon değerleri, GA sınıflama kriteriyle sınıflama yapıldığında 0.911-0.937, AOOT sınıflama kriteriyle sınıflama yapıldığında 0.924-0.924 olduğu görülmektedir. Bu bulguya göre bireylerin kestirilen yetenekleri gerçek yeteneklerine en yakın değerleri MFB-KN temelli madde seçme yöntemiyle GA sınıflama kriteri, KY temelli madde seçme yöntemi ile AOOT sınıflama kriterinin birlikte analize dahil edildiği koşullarda sağlandığı söylenebilir. Tablo 7'de hesaplanan yanlılık değerleri incelendiğinde GOO sınıflama kriteri dört kategorili sınıflamada MFB (KY-KN) madde seçme yöntemleriyle -0.119, 0.023, GA sınıflama kriteri -0.117,-0.007, AOOT sınıflama kriteri ise -0.171,-0.076 yanlılık değerleri gösterdiği Tablo 7'den anlaşılmaktadır. Buna göre en düşük yanlılık değeri AOOT sınıflama kriteri ile KY temelli madde seçme yöntemi birlikte kullanıldığında elde edilmektedir. En yüksek yanlılık değeri ise GOO sınıflama kriteri KN temelli madde seçme yöntemiyle göstermektedir. Yanlılık değeri için dört kategorili sınıflamada KY temelli madde seçme yönteminin KN temelli madde seçme yöntemine göre daha iyi performans gösterdiği söylenebilir. Tablo 7'de yanlılıkla birlikte kestirimin hata değerlerini dikkate alan RMSE, OMH değerleri için GOO sınıflama kriteri MFB (KY-KN) madde seçme yöntemleriyle karşılaştırıldığında 0.557-0.607 RMSE ve 0.358-0.372 OMH değerleri, GA sınıflama kriteri MFB (KY-KN) madde seçme yöntemleri ile birlikte kullanıldığında 0.551-0.35 RMSE ve 0.355-0.254 OMH değerleri, AOOT sınıflama kriteri ise MFB (KY-KN) madde seçme yöntemleri ile karşılaştırıldığında 0.492-0.411 RMSE ve 0.301-0.296 OMH değerleri hesaplanmıştır. Tablo 7'deki bu bulgulara göre ölçme kesinliği için en düşük hata değeri KN temelli madde seçme yöntemi için GA sınıflama kriteri ile yapılan sınıflamada KY temelli madde seçme yöntemi için AOOT sınıflama kriteri ile yapılan sınıflamada görülmektedir. En yüksek hata değerleri GOO sınıflama kriteri kullanıldığında görüldüğü söylenebilir. Genel anlamda KN temelli madde seçme yöntemi KY temelli madde seçme yöntemine göre ölçme kesinliği yani hatanın düşük olması açısından daha etkili performans gösterdiği söylenebilir.

KLB (KY-KN) madde seçme yöntemleriyle GOO sınıflama kriteri birlikte kullanıldığında ortalama 27-28 madde, 0.718-0.728 OSD, GA sınıflama kriteri ile ortalama 26-28 madde 0.714-0.79 OSD, AOOT sınıflama kriteri ile 32-33 madde ve 0.718-0.728 OSD ile sınıflama yapıldığı anlaşılmaktadır. GA sınıflama kriterinin test etkililiği için KLB madde seçme yöntemi ile daha iyi performans gösterdiği söylenebilir. KLB madde seçme yönteminin MFB madde seçme yöntemi ile benzer performans gösterse de MFB-KY temelli madde seçme yöntemi ile test etkililiği olarak en iyi performansın gösterildiği anlaşılmaktadır. BBST simülasyonu ile kestirilen yeteneklerle gerçek yetenekler arası korelasyon değerlerinin KLB (KY-KN) madde seçme yöntemleri ile GOO sınıflama kriteri karşılaştırıldığında 0.908-0.876, GA sınıflama kriteri ile birlikte 0.905-0.928, AOOT sınıflama kriteri ile birlikte 0.919-0.925 olduğu tablodan anlaşılmaktadır. Buna göre AOOT sınıflama kriteri KY temelli madde seçme yöntemi ile GA sınıflama kriteri KN temelli madde seçme yöntemi ile karşılaştırıldığında en yüksek korelasyon değerlerinin hesaplandığı görülmektedir. Yanlılık değerleri için KLB madde seçme yöntemi KY ve KN madde seçme yöntemleriyle GOO sınıflama kriteri karşılaştırıldığında -0.123, -0.024, GA sınıflama kriteri ile -0.152, -0.079, AOOT sınıflama kriteri ile -0.18, -0.094 yanlılık değerleri ile sınıflama yapıldığı görülmektedir. AOOT sınıflama kriterinin iki madde seçme yöntemiyle de en düşük yanlılıkla sınıflama yaptığı anlaşılmaktadır. Ayrıca kestirimin standart hata değerlerini gösteren RMSE ve OMH değerleri için KLB (KY-KN) madde seçme yöntemleriyle GOO sınıflama kriteri karşılaştırıldığında 0.573-0.642 RMSE ve 0.37-0.391 OMH değerleri, GA sınıflama kriteri ile karşılaştırıldığında 0.577-0.375 RMSE ve 0.375-0.264 OMH, AOOT sınıflama kriteri ile 0.511-0.406 RMSE ve 0.314-0.266 OMH değerlerinin hesaplandığı tablodan anlaşılmaktadır. Bu bulgulara göre KY temelli madde seçme yöntemi ile AOOT sınıflama kriteri, KN temelli madde seçme yöntemi ile GA sınıflama kriteri karşılaştırıldığında en düşük RMSE ve OMH değerlerinin hesaplandığı söylenebilir. Özetle ölçme kesinliği için AOOT sınıflama kriterinin hata değerleri yanlılık korelasyon değerleri olarak KY temelli madde seçme yöntemi ile en iyi performansı gösterdiği söylenebilir. Ayrıca GOO sınıflama kriterinin ise ölçme kesinliği için en düşük performansı gösterdiği araştırmanın bulguları arasındadır.

Tablo 7’de tüm sınıflama kategorileri birlikte değerlendirildiğinde yeteneğin maksimum olabilirlik kestirimi ile belirlendiği sınıflamada, sınıflama kategori sayısı arttıkça ortalama test uzunluğunun arttığı ortalama sınıflama doğruluğunun azaldığı test etkililiğinin düştüğü söylenebilir. Üç, dört kategorili sınıflamalarda madde havuzunda madde sayısı azaldığı için sınıflamak için gereken madde sayısının arttığı yorumu yapılabilir. Gerçek yeteneklerle BBST simülasyonu sonucunda kestirilen yetenekler arası korelasyonların sınıflama kategori sayısı arttıkça arttığı, RMSE ve OMH gibi hata değerlerinin azaldığı daha hassas ölçme yapıldığı tablodan anlaşılmaktadır. İki kategorili sınıflama için GA sınıflama kriterinin her iki madde seçme yöntemiyle de en az maddeyle sınıflama yaptığı, OSD için GA sınıflama kriterinin MFB madde seçme yöntemi ile yüksek performans gösterdiği söylenebilir. GA sınıflama kriteri her maddeden sonra belirlenen yetenek düzeylerini kullanarak hesaplanan güven aralığını kesme puanı ile karşılaştırarak bireyleri sınıflamaktadır. GOO sınıflama kriteri MFB-KN madde seçme yöntemi ile etkili performans göstermesiyle birlikte sınıflamanın standart hata değerleri ölçme kesinliğinin düşük olduğu söylenebilir. AOOT sınıflama kriterinin ise KY temelli madde seçme yöntemleriyle ölçme kesinliği olarak performansının etkili olduğu söylenebilir. Bu durumun AOOT sınıflama kriterinin farksızlık bölgesinin küçük olması yüksek sınıflama doğruluğunda sınıflama yapabildiği özelliğinden kaynaklandığı yorumu yapılabilir. Üç kategorili sınıflamada GA sınıflama kriteri MFB-KN madde seçme yöntemi ile test etkililiği için, AOOT sınıflama kriteri MFB-KY madde seçme yöntemi ile ölçme kesinliği için en iyi performans gösterdiği söylenebilir. KN temelli madde seçme yöntemlerinin KY temelli madde seçme yöntemlerine göre ölçme kesinliği bakımından üç kategorili sınıflamada daha iyi performans gösterdiği araştırmanın bulguları arasındadır. Dört kategorili sınıflamada AOOT sınıflama kriterinin KY temelli madde seçme yöntemi ile, GA sınıflama kriterinin ise KN temelli madde seçme yöntemi ile ölçme kesinliğinde etkili olduğu, GOO sınıflama kriterinin ise ölçme kesinliği olarak performansının en düşük olduğu araştırmanın bulgularındandır. MOK yetenek kestirimi ile çok kategorili sınıflamada sınıflama kriterleri ve madde seçme yöntemlerinin

çaprazlandığı BBST örneğine literatürde rastlanmadığı için literatürle tartışılmamıştır. Bu çalışmaya özgü bulgulardır sonraki alan çalışmalarına katkı sağlaması beklenmektedir.

İkinci Alt Probleme Ait Bulgular Ve Yorumlar

Araştırmanın ikinci alt probleminde iki, üç, dört kategorili BBST uygulamasında yetenek kestirimi AOK olduğunda MFB (KY-KN) ve KLB (KY-KN) madde seçme yöntemleri ile GOO (FB:0.1), AOOT (FB:0.1) ve GA (%90) sınıflama kriterlerine göre OTU, OSD, yanlılık, RMSE, OMH, ve gerçek yetenekler ile kestirilen yetenekler arasındaki korelasyonun nasıl değiştiği araştırılmıştır.

Tablo 8'de OTU, OSD ve ölçme kesinliği için RMSE, OMH, r, yanlılık değerleri iki, üç, dört kategorili sınıflamalar için gösterilmektedir. İki kategorili sınıflamada Tablo 8'e göre yeteneğin AOK yöntemi ile kestirildiği durumda MFB (KY-KN) madde seçme yöntemleriyle GOO sınıflama kriteri birlikte kullanıldığında ortalama 18-18 madde ile 0.833 ve 0.862 OSD ile sınıflama yapılmıştır. GA sınıflama kriteriyle birlikte 16-18 madde 0.824-0.857 sınıflama doğruluğuyla, AOOT sınıflama kriteri ile birlikte analize dahil edildiğinde 25-27 madde ve 0.833-0.866 sınıflama doğruluğuyla sınıflama yapıldığı anlaşılmaktadır. Tablo 8'e göre AOK yetenek kestirimi kullanıldığında MFB-KY temelli madde seçme yönteminin GA sınıflama kriteri ile en az madde ile testi sonlandırdığı anlaşılmaktadır. AOOT sınıflama kriteri ile oluşturulan koşullarda en yüksek sayıda madde ile sınıflamanın gerçekleştiği söylenebilir. KN temelli madde seçme yöntemi kullanıldığında OSD ve OTU için GOO sınıflama kriteri diğer sınıflama kriterlerinden daha iyi performans göstermiştir. Her iki madde seçme yöntemi için de AOOT sınıflama kriterinin en yüksek madde ile testi sonlandırıp sınıflama yaptığı söylenebilir. Bunun yanı sıra KN temelli madde seçme yöntemi KY temelli madde seçme yöntemine göre daha yüksek sınıflama doğruluğuyla daha yüksek performans gösterdiği söylenebilir. Özetle Tablo 8'de OTU ve OSD değişkenleri için hesaplanan değerler incelendiğinde OTU olarak GA kriterinin OSD olarak AOOT kriterinin KY temelli madde seçme yöntemi ile, GOO sınıflama kriterinin ise KN temelli madde seçme yöntemi ile birlikte performansının yüksek olduğu söylenebilir.

Tablo 8

Yetenek Kestirim Yöntemi AOK Olduğunda İki, Üç, Dört Kategorili Sınıflamaya Ait Test Etkilliliği Ve Ölçme Kesinliği Değerleri

Koşullar		Bağımlı						
Madde Seçme Yöntemleri	Sınıflama Kriterleri	SKS	OTU	OSD	r	Yanlılık	RMSE	OMH
MFB-KY	GOO(FB:0.1)	İki	18	0.833	0.902	-0.128	0.479	0.325
		Üç	24.25	0.783	0.913	-0.122	0.461	0.302
		Dört	27.64	0.727	0.919	-0.117	0.445	0.288
	GA(%90)	İki	16	0.824	0.897	-0.128	0.489	0.335
		Üç	22.56	0.778	0.910	-0.124	0.466	0.308
		Dört	26.95	0.725	0.917	-0.12	0.45	0.293
	AOOT(FB:0.1)	İki	25	0.833	0.922	-0.127	0.435	0.282
		Üç	30.84	0.786	0.929	-0.134	0.416	0.264
		Dört	33.01	0.725	0.932	-0.132	0.409	0.257
MFB-KN	GOO(FB:0.1)	İki	18	0.862	0.896	-0.061	0.42	0.305
		Üç	24.73	0.801	0.923	-0.102	0.42	0.289
		Dört	28.20	0.721	0.937	-0.073	0.337	0.246
	GA(%90)	İki	18	0.857	0.896	-0.076	0.424	0.308
		Üç	23.39	0.795	0.923	-0.105	0.419	0.289
		Dört	28.87	0.719	0.938	-0.077	0.336	0.245
	AOOT(FB:0.1)	İki	27	0.866	0.927	-0.067	0.357	0.255
		Üç	31.37	0.802	0.943	-0.117	0.365	0.243
		Dört	33.94	0.718	0.946	-0.078	0.314	0.228
KLB-KY	GOO(FB:0.1)	İki	18	0.823	0.9	-0.133	0.485	0.33
		Üç	24.32	0.772	0.910	-0.129	0.469	0.307
		Dört	27.63	0.716	0.916	-0.121	0.456	0.296
	GA(%90)	İki	16	0.82	0.896	-0.136	0.496	0.339
		Üç	22.78	0.766	0.908	-0.136	0.474	0.315
		Dört	27.03	0.712	0.915	-0.128	0.457	0.297
	AOOT(FB:0.1)	İki	26	0.82	0.919	-0.144	0.443	0.29
		Üç	30.77	0.773	0.923	-0.139	0.433	0.274
		Dört	32.96	0.715	0.927	-0.145	0.424	0.265
KLB-KN	GOO(FB:0.1)	İki	18	0.857	0.895	-0.072	0.425	0.305
		Üç	24.71	0.789	0.92	-0.116	0.438	0.297
		Dört	28.15	0.721	0.932	-0.088	0.352	0.253
	GA(%90)	İki	18	0.853	0.893	-0.083	0.43	0.31
		Üç	23.52	0.786	0.922	-0.121	0.431	0.297
		Dört	29.03	0.718	0.933	-0.098	0.353	0.251
	AOOT(FB:0.1)	İki	27	0.862	0.922	-0.074	0.371	0.264
		Üç	31.27	0.794	0.937	-0.133	0.391	0.255
		Dört	33.87	0.719	0.945	-0.091	0.32	0.231

SKS: Sınıflama kategori sayısı, OTU: Ortalama test uzunluğu, OSD: Ortalama sınıflama doğruluğu, OMH: Ortalama mutlak hata, r: Gerçek yeteneklerle kestirilen yetenekler arası korelasyon

İki kategorili sınıflamada Tablo 8'de bireylerin gerçek yetenekleri ile kestirilen yetenekleri arasındaki korelasyonların GOO sınıflama kriteri MFB (KY-KN) madde seçme yöntemleriyle birlikte kullanıldığında 0.902-0.896 değerlerinde, GA sınıflama kriteriyle 0.910-0.896, AOOT sınıflama kriteriyle 0.932-0.946 değerlerindedir. Bu bulguya göre korelasyon (r) değeri bakımından AOOT sınıflama kriterinin KY- KN temelli madde seçme yöntemlerinin her ikisi ile beraber en iyi performansı gösterdiği genel anlamda KY temelli madde seçme yönteminin KN temelli yöntemin olduğu koşullardan daha yüksek korelasyonların belirlendiği görülmektedir. İki kategorili sınıflama için araştırmmanın bu bulguları Gündeğer (2017) ve Demir (2019) araştırma sonuçlarıyla uyumludur. Tablo 8 incelendiğinde MFB (KY-KN) madde seçme yöntemleriyle GOO sınıflama kriteri birlikte kullanıldığında -0.128,-0.061 yanlılık değerlerinin, GA sınıflama kriteriyle -0.128,-0.076 yanlılık değerlerinin, AOOT sınıflama kriteriyle birlikte -0.127,-0.067 değerlerinin elde edildiği görülmektedir. KY temelli madde seçme yöntemleri kullanıldığında KN temelli madde seçme yöntemlerine göre daha düşük yanlılık değerlerinin hesaplandığı söylenebilir. Tablo 8'de yanlılıkla beraber standart hata değerlerini gösteren RMSE ve OMH değerlerine göre GOO sınıflama kriteri MFB (KY-KN) madde seçme yöntemleriyle 0.479-0.42 RMSE, 0.325-0.305 OMH değerleri, GA sınıflama kriteri ile 0.489-0.424 RMSE, 0.335-0.308 OMH, AOOT sınıflama kriteri ile 0.435-0.357 RMSE ve 0.282-0.255 OMH değerleri hesaplanmıştır. Bu bulguya göre AOOT sınıflama kriterinin her iki madde seçme yöntemi ile kullanıldığı koşullarda en düşük hata değerleriyle sınıflama yaptığı söylenebilir. GA sınıflama kriteri ise en yüksek hata değerleri ile sınıflama yaptığı söylenebilir. KY temelli madde seçme yöntemi ile KN temelli madde seçme yöntemine göre daha düşük RMSE ve OMH değerleri elde edilmiştir. Araştırmmanın bu bulgusu Gündeğer (2017) ile uyumlu olmakla beraber GOO sınıflama kriterinin GA sınıflama kriterine göre daha başarılı performans göstermesi bu araştırmaya ait bulgular arasındadır. Bu bulgulara göre GOO sınıflama kriterinin KN temelli madde seçme yöntemi ile birlikte test etkililiği ve ölçme kesinliği olarak uygulamada avantaj sağlayacağı yorumu düşünülebilir.

İki kategorili sınıflamada KLB (KY-KN) madde seçme yöntemi ile GOO sınıflama kriteri kullanıldığında ortama 18-18 madde ile 0.823-0.857 sınıflama doğruluğuyla, GA sınıflama kriteri ile 16-18 madde, 0.82-0.853 OSD ile, AOOT sınıflama kriteri ile birlikte 26-27 madde, 0.82-0.862 OSD ile sınıflama yapıldığı anlaşılmaktadır. Bu bulguya göre test etkililiği için en iyi performansı en düşük madde ve en yüksek sınıflama doğruluğunu GA sınıflama kriteri KY ve KN madde seçme yöntemleriyle gösterdiği söylenebilir. En fazla madde gerektiren sınıflama yönteminin ise AOOT sınıflama kriteri olduğu görülmektedir. Bununla birlikte GOO sınıflama kriteri KY temelli madde seçme yöntemi ile birlikte kullanıldığında GA sınıflama kriterine kıyasla ortalama test uzunluğu olarak benzer performans gösterse de OSD olarak daha iyi performans gösterdiği görülmektedir. GOO sınıflama kriteri AOOT nin modifiye edilmiş daha genel halidir. GOO sınıflama kriteri güçlük parametresi kesme puanıyla eşleşmeyen daha uzaktaki maddeleri de verimli hale getirdiği için madde havuzu daha etkin kullanıldığından AOOT sınıflama kriterine göre daha az madde ile sınıflama yaptığı yorumu yapılabilir. GOO sınıflama kriterinin KLB madde seçme yöntemi ile KN temelli madde seçiminde AOOT yerine daha az madde ile sınıflama yaptığı için uygulamada avantajlı olabileceği söylenebilir. OSD için AOOT sınıflama kriterinin KN temelli madde seçme yöntemiyle diğer yöntemlere göre daha yüksek sınıflama doğruluğu gösterdiği görülmektedir. KN temelli madde seçme yönteminin OSD için KY temelli madde seçme yönteminin test uzunluğu olarak daha iyi performans gösterdiği araştırmanın bulguları arasındadır. Nydick ve diğerleri (2012), Thompson (2009) ile Thompson ve Ro'nun (2007) araştırma sonuçlarıyla araştırmanın bu bulgusu benzerlik göstermektedir. Tablo 8 de gerçek yetenekler ile kestirilen yetenek düzeyleri arasındaki korelasyon ve yanlılık değerleri için GOO sınıflama kriteri KLB (KY-KN) madde seçme yöntemleriyle birlikte kullanıldığında 0.9-0.895 korelasyon,-0.133,-0.144 yanlılık, GA sınıflama kriteri ile 0.896-0.893 korelasyon -0.136,-0.083 yanlılık, AOOT sınıflama kriteri ile birlikte 0.919-0.922 korelasyon,-0.144,-0.074 yanlılık değerleri görülmektedir. Bu bulguya göre en yüksek korelasyon değerleri AOOT sınıflama kriteri ile birlikte her iki madde seçme yöntemi çaprazlandığında hesaplandığı anlaşılmaktadır. En düşük yanlılık değeri ise KN temelli

madde seçme yöntemi ile GOO sınıflama kriteri, KY madde seçme yöntemi ile AOOT sınıflama kriteri kullanıldığında görülmektedir. Yanlılık değerlerinin madde seçme yöntemleri ve sınıflama kriterlerine göre değiştiği söylenebilir. Tablo 8 de kestirimin standart hatasını dikkate alan RMSE ve OMH değerlerine göre KLB (KY-KN) madde seçme yöntemleriyle GOO sınıflama kriteri birlikte kullanıldığında 0.485-0.425 RMSE, 0.33-0.305 OHM değerleri, GA sınıflama kriteri ile 0.496-0.43 RMSE, 0.339-0.31 OMH, AOOT sınıflama kriteri ile 0.443-0.371 RMSE, 0.29-0.264 OMH değerleri elde edilmiştir. Bu bulguya göre en düşük hata değerleri her iki madde seçme yöntemiyle AOOT sınıflama kriteri birlikte kullanıldığında elde edilmiştir. Bu bulgu Gündeğer (2017) ve Demir (2019) araştırma sonuçlarıyla da desteklenmektedir.

Özetle yetenek kestiriminin AOK olduğu MFB ve KLB madde seçme yöntemleriyle yapılan iki kategorili sınıflamada test etkililiği için en yüksek sınıflama doğruluğu ve en az madde ile testi sonlandıran sınıflama koşulunun MFB-KY madde seçme yöntemiyle GA sınıflama kriterinin çaprazlandığı koşula ait performansın yüksek olduğu söylenebilir. GA sınıflama kriteri belirlenen yetenek düzeylerini kullanarak belirlenen güven aralığını kesme puanı ile karşılaştırdığı için KY temelli madde seçme yöntemleri ile daha başarılı performans gösterdiği yorumu yapılabilir. KN temelli madde seçme yöntemleriyle yapılan sınıflamada KY temelli madde seçme yöntemlerine göre OSD daha yüksektir. Ölçme kesinliği için hata değerlerinin yanlılık değerlerinin düşük, kestirilen yetenek ile gerçek yetenekler arası korelasyonun yüksek olması beklenir. AOOT sınıflama kriteri MFB (KY-KN) madde seçme yöntemleriyle birlikte kullanıldığında en yüksek korelasyon değerleri hesaplandığı görülmektedir. Aynı şekilde AOOT sınıflama kriteri KLB-KN madde seçme yöntemiyle en düşük hata değerlerinde sınıflama yaptığı ölçme kesinliğinin yüksek olduğu araştırmanın bulgularındandır. AOOT sınıflama kriterinin farksızlık bölgesi 0.10 olarak belirlenmiştir. Böylece telafi edilebilir hata oranları düşüktür. Küçük farksızlık bölgesi nedeniyle yüksek doğrulukta daha fazla madde ile sınıflama yaptığı AOOT sınıflama kriterinin özelliği ile araştırmanın bu bulgusunun örtüştüğü yorumu yapılabilir. Ayrıca KN temelli madde seçme

yöntemleri ve GOO sınıflama kriteri AOOT den daha az madde ile sınıflama yaptığı için uygulamada avantaj sağlayabileceği yorumu yapılabilir. Bu bulgunun da GOO sınıflama kriterinin madde havuzunda kesme noktasından çok uzak güçlük değerlerine sahip maddeleri verimli olarak kullandığı için AOOT den daha az madde ile sınıflama yapma özelliği ile örtüşmektedir.

Üç kategorili sınıflamada MFB (KY-KN) madde seçme yöntemleriyle GOO sınıflama kriteri birlikte kullanıldığında ortalama 24-24 madde ile 0.783-0.801 sınıflama doğruluğuyla, GA sınıflama kriteri ile birlikte 22-23 madde, 0.778-0.795 OSD ve AOOT sınıflama kriteriyle 30-31 madde, 0.786-0.802 OSD ile sınıflama yapılmıştır. Bu bulgulara göre MFB(KY-KN) her iki madde seçme yöntemleriyle yapılan üç kategorili sınıflamada GA sınıflama kriterinin en az maddeyle sınıflama yaptığı görülmektedir. AOOT sınıflama kriterinin ise en fazla maddeyle sınıflama yaptığı görülmektedir. OSD için AOOT sınıflama kriterinin KN temelli madde seçme yöntemiyle birlikte performansının daha yüksek olduğu tablodan anlaşılmaktadır. Bununla beraber KY temelli madde seçme yöntemi KN temelli madde seçme yöntemine göre az maddeyle sınıflama yaparken KN temelli madde seçme yöntemi ile birlikte de daha yüksek sınıflama doğruluğuyla sınıflama yapılmıştır. Tablo 8'e göre kestirilen yeteneklerle gerçek yetenekler arası korelasyon ve yanlılık için üç kategorili sınıflamada MFB (KY-KN) temelli madde seçme yöntemiyle birlikte GOO sınıflama kriteri 0.913-0.923 korelasyon, -0.122,-0.102 yanlılık değerleri, GA sınıflama kriteriyle birlikte 0.910-0.923 korelasyon,-0.124,-0.105 yanlılık değerleri, AOOT sınıflama kriteri 0.929-0.943 korelasyon, -0.134,-0.117 yanlılık değerleri hesaplanmıştır. Hesaplanan bulgulara göre KY ve KN her iki madde seçme yöntemiyle de AOOT sınıflama kriteri kullanıldığında yanlılık değerleri en düşük ve gerçek yeteneklere en yakın yetenekler hesaplanmıştır. Farklı bir ifade ile korelasyon ve yanlılık olarak en iyi performansı AOOT sınıflama kriterinin gösterdiği anlaşılmaktadır. GA sınıflama kriteri ile en düşük korelasyon değerleri hesaplanmıştır. Bununla beraber KN temelli madde seçme yöntemi kullanıldığında KY temelli madde seçme yöntemine göre yetenekler arası korelasyon değerleri ve yanlılık değerlerinin daha yüksek

olduğu araştırmanın bulguları arasındadır. Tablo 8’de kestirimin standart hatasını gösteren RMSE ve OMH değerlerine göre üç kategorili sınıflamada MFB (KY-KN) madde seçme yöntemleri ile GOO sınıflama kriteri kullanıldığında 0.461-0.42 RMSE, 0.302-0.305 OMH değerleri, GA sınıflama kriteri ile birlikte 0.466-0.419 RMSE ve 0.308-0.289 OMH değerleri, AOOT sınıflama kriteriyle 0.416-0.365 RMSE ve 0.264-0.243 OMH değerleri hesaplanmıştır. Bu bulgulara göre üç kategorili sınıflamada AOOT sınıflama kriterinin en az hata ile sınıflama yapıldığı söylenebilir. Bununla birlikte MFB-KY madde seçme yönteminin kullanıldığı koşullarda MFB-KN madde seçme yönteminin kullanıldığı koşullara kıyasla daha yüksek OMH ve RMSE değerleri elde edilmiştir.

Üç kategorili sınıflamada KLB (KY-KN) madde seçme yöntemi kullanıldığında GOO sınıflama kriteri ile birlikte ortalama 24-24 madde 0.772-0.789 OSD ile, GA sınıflama kriteri ile birlikte 22-23 madde 0.766-0.786 OSD ile, AOOT sınıflama kriteri ile ortalama 30-31 madde ve 0.82-0.862 ortalama sınıflama doğruluğu ile sınıflama yapıldığı anlaşılmaktadır. Bu bulgulara göre test etkililiği açısından en az madde ile sınıflamanın GA sınıflama kriteri ile yapıldığı AOOT sınıflama kriterinin ise en yüksek madde sayısı ile sınıflama yaptığı söylenebilir. AOOT sınıflama kriteri ile yapılan sınıflamada en yüksek ortalama sınıflama doğruluğu hesaplandığı GA sınıflama kriteri ile en az sınıflama doğruluğunun hesaplandığı araştırmanın bulguları arasındadır bunun yanı sıra KY temelli madde seçme yöntemi KN temelli madde seçme yöntemine göre OSD olarak daha yüksek performans göstermiştir. Test etkililiği için OSD yüksek ortalama madde sayısının az olması beklenir. GOO sınıflama kriteri de GA sınıflama kriterine benzer sayıda madde ile daha yüksek sınıflama doğruluğu ile sınıflama yaptığı için uygulayıcılara önerilebilir. Tablo 8’de gerçek yetenekler ile kestirilen yetenekler arası korelasyon ve yanlılık değerlerine bakıldığında KLB (KY-KN) madde seçme yöntemleriyle GOO sınıflama kriteri 0.910-0.92 korelasyon -0.129,-0.116 yanlılık, GA sınıflama kriteri ile 0.908-0.922 korelasyon ve -0.136,-0.121 yanlılık, AOOT sınıflama kriteri ile 0.923-0.937 korelasyon ve -0.139,-0.133 yanlılık değerleri ile sınıflama yapıldığı tablodan anlaşılmaktadır. Buna göre gerçek yeteneklerle kestirilen yetenekler arası en

yüksek korelasyonların (r) ve en düşük yanlılığın AOOT sınıflama kriterinin kullanıldığı durumlarda gözlemlendiği, GOO sınıflama kriterinin kullanıldığı koşullarda da en düşük korelasyon en yüksek yanlılık değerleri hesaplandığı söylenebilir. Her iki madde seçme yöntemi için de tüm sınıflama kriterleri için KN temelli madde seçme yöntemi kullanıldığında korelasyon değerlerinin daha yüksek olduğu araştırmanın bulguları arasındadır. Tablo 8'de KLB (KY-KN) madde seçme yöntemleri ile sınıflama kriterlerinin çaprazlandığı üç kategorili sınıflama koşullarında kestirimin standart hatasını dikkate alan RMSE ve OMH değerlerinin GOO sınıflama kriteriyle 0.469-0.438 RMSE, 0.307-0.297 OMH değeri, GA sınıflama kriteri ile 0.474-0.431 RMSE, 0.315-0.297 OMH değerleri, AOOT sınıflama kriteri ile birlikte 0.433-0.391 RMSE, 0.274-0.255 OMH değerleri hesaplanmıştır. Elde edilen bulgulara göre en düşük RMSE ve OMH değerleri AOOT sınıflama kriterinin her iki madde seçme yöntemi ile birlikte kullanıldığı koşullarda hesaplandığı söylenebilir. Benzer şekilde en yüksek RMSE ve OMH değerleri KY temelli madde seçme yöntemi ile GA sınıflama kriterinin birlikte kullanıldığı koşullarda hesaplanmıştır. Araştırmanın bu bulgusu Demir'in (2019) araştırmasıyla da desteklenmektedir.

Özetle Tablo 8'deki değerler dikkate alınarak gerçek veri seti ile yapılan BBST simülasyonunda yetenek kestirim yöntemi AOK iken MFB ve KLB (KY-KN) madde seçme yöntemleriyle birlikte test etkililiği için BBST simülasyonunda en az madde ile sınıflama yapılması ve sınıflama doğruluğunun en yüksek olması beklenir. Üç kategorili sınıflamada GA sınıflama kriterinin tüm madde seçme yöntemiyle de en az sayıda madde ile sınıflama yaptığı, AOOT sınıflama kriterinin OSD için KN temelli madde seçme yöntemi ile birlikte kullanıldığında daha yüksek değer hesaplanmış olsa da GA sınıflama kriteri de OSD için yüksek değerlere sahip olduğu için uygulamada özellikle KN temelli MFB madde seçme yöntemi ile birlikte OSD yüksek olması nedeniyle avantaj sağlayacağı anlaşılmaktadır. MFB madde seçme yönteminin KLB madde seçme yönteminden daha iyi performans gösterdiği araştırmanın bulguları arasındadır. BBST simülasyonunda ölçme kesinliği için gerçek yeteneklerle kestirilen yetenekler arası korelasyonun yüksek yanlılık ve hata değerlerinin

düşük olması beklenir. Ölçme kesinliği için AOOT sınıflama kriterinin KN temelli MFB madde seçme yöntemiyle birlikte kullanıldığında en düşük hata yanlılık değerleri hesaplanmıştır. Bu bulgulara rağmen Demir (2019) da araştırmasında AOOT sınıflama kriterini ölçme kesinliğinde önerse de KY temelli madde seçme yöntemiyle performansının yüksek olduğu kanaatine varmıştır. Madde havuzunu $\theta = 0$ kesme noktası etrafında yüksek bilgi veren maddelerden oluşturmuş olmasını bu durumun kaynağı olarak göstermiştir. Bu araştırmada madde havuzu gerçek verilerden oluşmaktadır. Herhangi bir manipülasyon yapılmamıştır.

Dört kategorili sınıflamada MFB (KY-KN) madde seçme yöntemi ile GOO sınıflama kriteri birlikte kullanıldığında ortalama 27-28 madde ve 0.727-0.716 OSD, GA sınıflama kriteri ile birlikte kullanıldığında 26-28 madde ve 0.725-0.719 OSD, AOOT sınıflama kriteriyle birlikte 33-33 madde ve 0.725-0.718 OSD ile sınıflama yapıldığı görülmektedir. Tablo 8'e göre test etkililiğinde dört kategorili sınıflamada KN ve KY temelli her iki madde seçme yöntemi ile GA sınıflama kriterinin daha düşük madde ve daha yüksek sınıflama doğruluğu ile sınıflama yaptığı görülmektedir. AOOT sınıflama kriteri ise en yüksek madde sayısı ile sınıflama yaptığı anlaşılmaktadır. Ayrıca KY temelli madde seçme yönteminin KN temelli madde seçme yöntemine göre daha iyi performans gösterdiği görülmektedir. Test etkililiğinin madde seçme yöntemi ve sınıflama kriterlerine göre değiştiği yorumu yapılabilir. Dört kategorili sınıflamada bireylerin gerçek yetenekleri ile kestirilen yetenekleri arasındaki korelasyon yanlılık kestirimin standart hata değerlerine baktığımızda MFB (KY-KN) madde seçme yöntemleri ile GOO sınıflama kriteri birlikte kullanıldığında 0.919-0.937 korelasyon, -0.117,-0.073 yanlılık, 0.445 -0.337 RMSE, 0.228-0.246 OMH değerleri, GA sınıflama kriteri ile 0.917-0.938 korelasyon,-0.12,-0.077 yanlılık, 0.45-0.336 RMSE, 0.293-0.245 OMH değerleri, AOOT sınıflama kriteri ile birlikte 0.932-0.946 korelasyon, -0.132,-0.078 yanlılık, 0.409-0.314 RMSE, 0.257-0.228 OMH değerleri hesaplanmıştır. Bu bulgulara göre ölçme kesinliği için gerçek yeteneklerle kestirilen yetenekler arası korelasyonun yüksek yanlılık ve standart hatayı gösteren RMSE ve OMH değerlerinin düşük olması BBST için beklenen

durumdur. En yüksek korelasyon ile en düşük yanlışlık ve hata değerleri AOOT sınıflama kriteri KY-KN temelli madde seçme yöntemlerinin her ikisi ile birlikte kullanıldığında hesaplanmıştır. Buna göre dört kategorili sınıflamada ölçme kesinliği için AOOT sınıflama kriteri diğer yöntemlere göre daha iyi performans gösterdiği için uygulamada avantajlı olabileceği yorumu yapılabilir. Bununla birlikte KN temelli madde seçme yönteminin KY temelli madde seçme yöntemine göre daha iyi performans gösterdiği söylenebilir. Bu bulgu Eggen (1999); Eggen ve Straetmans (2000) çalışmalarıyla uyumludur.

Dört kategorili sınıflamada KLB (KY-KN) madde seçme yöntemleriyle GOO sınıflama kriterinin çaprazlandığı koşullarda ortalama 27-28 madde, 0.716-0.721 OSD, GA sınıflama kriteriyle birlikte ortalama 27-29 madde 0.712-0.718 OSD ve AOOT sınıflama kriteri ile birlikte 32-33 madde ve 0.715-0.719 OSD ile sınıflama yapıldığı görülmektedir. Bu bulgulara göre KLB madde seçme yöntemi ile GOO sınıflama kriteri diğer sınıflama kriterlerine göre daha iyi performans gösterdiği en az madde ve en yüksek sınıflama doğruluğu ile sınıflama yaptığı söylenebilir. AOOT sınıflama kriterinin ise en yüksek madde ile sınıflama yaptığı anlaşılmaktadır. KN temelli madde seçme yöntemi sınıflama doğruluğu açısından KY temelli madde seçme yöntemi ise ortalama madde sayısı olarak daha iyi performans gösterdiği araştırmanın bulguları arasındadır. Tablo 8'de KLB (KY-KN) madde seçme yöntemleriyle sınıflama kriterlerinin birlikte kullanıldığı koşullara ait korelasyon, yanlışlık, RMSE ve OMH değerlerine bakıldığında GOO sınıflama kriteri ile birlikte 0.916-0.932 korelasyon, -0.121,-0.088 yanlışlık, 0.456-0.352 RMSE ve 0.296-0.253 OMH değerleri, GA sınıflama kriteri ile birlikte kullanıldığı koşullarda 0.915-0.933 korelasyon, -0.128,-0.098 yanlışlık, 0.454-0.353 RMSE ve 0.297-0.251 OMH değerleri, AOOT sınıflama kriteri ile birlikte kullanıldığında 0.927-0.945 korelasyon, -0.145,-0.091 yanlışlık, 0.424-0.32 RMSE ve 0.265-0.231 OMH değerleri hesaplanmıştır. Bu bulgulara göre ölçme kesinliği için gerçek yetenekler ile kestirilen yetenekler arası en yüksek korelasyon, en düşük yanlışlık değerleri ve düşük standart hata değerleri her iki madde seçme yöntemiyle birlikte AOOT sınıflama kriteri kullanıldığında hesaplandığı görülmektedir. Ayrıca KN temelli madde

seçme yönteminin KY temelli madde seçme yöntemine göre ölçme kesinliği açısından daha iyi performans gösterdiği söylenebilir.

Özetle dört kategorili sınıflamada yetenek kestirimi AOK iken KLB ve MFB madde seçme yöntemleri ve GOO, GA, AOOT sınıflama kriterleriyle yapılan çaprazlamada MFB (KY-KN) madde seçme yöntemi ile test etkililiği olarak en az madde yüksek sınıflama doğruluğu ile GA sınıflama kriterinin sınıflama yaptığı, KLB (KY-KN) madde seçme yöntemiyle ise GOO sınıflama kriterinin test etkililiği olarak en iyi performansı gösterdiği söylenebilir. GOO sınıflama kriteri AOOT' nin modifiye edilmiş daha genel bir halidir ve dört kategorili sınıflamada KLB madde seçme yöntemi ile uygulamada avantajlı olabileceği yorumu yapılabilir. Bunun yanı sıra MFB madde seçme yönteminde KY temelli madde seçme yönteminin KN temelli madde seçme yöntemine göre daha iyi performans gösterdiği, KLB madde seçme yönteminde ise KN temelli madde seçme yöntemi sınıflama doğruluğu açısından KY temelli madde seçme yöntemi ise ortalama madde sayısı olarak daha iyi performans gösterdiği araştırmanın bulguları arasındadır. Ölçme kesinliği açısından AOOT sınıflama kriteri MFB-KLB her iki madde seçme yöntemiyle de en iyi performansı gösterdiği MFB de ölçme kesinliği için KN temelli madde seçme yöntemleri KY temelli madde seçme yöntemine göre daha iyi performans gösterdiği söylenebilir. Bu bulgu Eggen (1999) çalışmasıyla da uyumludur. KLB için KY temelli madde seçme yönteminin KN temelli madde seçme yöntemine göre ölçme kesinliği açısından daha iyi performans gösterdiği söylenebilir. Ayrıca araştırmada KLB'ın MFB'ye benzer bazen MFB den daha iyi performans gösterdiği görülmektedir. Bu bulgu da Eggen (1999) çalışmasıyla da desteklenmektedir.

Tablo 8'deki tüm koşullarda sınıflama kategori sayısı arttıkça OTU artmış OSD ise azaldığı yani test etkililiğinin düştüğü anlaşılmaktadır. Ölçme kesinliği için sınıflama kategori sayısı arttıkça korelasyon değerleri artmış yanlışlık, RMSE ve OMH değerleri genel olarak düşmüş olması araştırmanın önemli bulgularındandır. Sınıflama kategori sayısı arttıkça sınıflamak için gereken madde sayısı arttığı için bireylerin son yetenek düzeyleri daha hassas kestirildiği yorumu bu bulguya göre yapılabilir ve Demir (2019) sınıflama kategori

sayısı arttıkça korelasyonların benzer değerler gösterdiğini bulgulamış bu araştırmanın bulgularıyla nispeten uyumludur. Araştırmadaki tüm koşullara göre test etkililiği açısından en iyi sonuçların (OTU: 16; OSD:0.824) GA sınıflama kriteri ile MFB-KY madde seçme yöntemi ile birlikte kullanıldığı sınıflamanın iki kategoride yapıldığı koşullar iken, en düşük sonuçların (OTU:33.94; OSD:0.718) MFB-KN madde seçme yöntemi ile AOOT sınıflama kriterinin birlikte kullanıldığı sınıflamanın dört kategoride yapıldığı koşullarda hesaplandığı söylenebilir. Ölçme kesinliği için AOOT sınıflama kriterinin diğer sınıflama kriterlerine göre daha iyi performans gösterdiği dört kategorili sınıflama ve MFB-KN madde seçme yönteminin performansının yüksek olduğu yorumu Tablo 8'e göre yapılabilir.

Üçüncü Alt Probleme Ait Bulgular Ve Yorumlar

Araştırmanın üçüncü alt probleminde iki, üç ve dört kategorili BBST uygulamasında BSD yetenek kestirim yönteminde, MFB(KY-KN) ve KLB(KY-KN) madde seçme yöntemleri ile iki, üç, dört kategorili sınıflamada GOO (FB:0.1), GA (%90), AOOT (FB: 0,1) sınıflama kriterlerine göre OTU, OSD ve ölçme kesinliği nasıl değiştiği araştırılmıştır.

Tablo 9'da araştırma probleminde belirtilen koşulları karşılaştırmada kullanılan bağımlı değişkenler olan ortalama test uzunluğu (OTU), ortalama sınıflama doğruluğu (OSD), gerçek yetenekler ile kestirilen yetenekler arası korelasyon (r), yanlılık, RMSE ve ortalama mutlak hataya ait değerler yer almaktadır. Tüm koşullar için 25 tekrarın ortalaması alınarak değerler hesaplanmıştır.

İki kategorili sınıflamada Tablo 9'da Yetenek kestirimi BSD olduğunda MFB (KY-KN) madde seçme yöntemlerinin birlikte kullanıldığı koşullarda GOO sınıflama kriteri kullanıldığında 18-18 ortalama test uzunluğu, 0.842-0.864 OSD, GA sınıflama kriteri kullanıldığında 16-16 OTU ve 0.837-0.855 OSD ile AOOT sınıflama kriteri ile birlikte 26-27 OTU ve 0.845-0.871 OSD ile sınıflama yapılmıştır. Bu bulgulara göre GA sınıflama kriteri KY ve KN temelli madde seçme yöntemlerinin ikisi için de en az madde ile sınıflama yapmıştır.

Tablo 9

Yetenek Kestirim Yöntemi BSD Olduğunda İki, Üç, Dört Kategorili Sınıflamaya Ait Test Etkilliliği Ve Ölçme Kesinliği Değerleri

Koşullar			Bağımlı					
Madde seçme yöntemleri	Sınıflama kriterleri	SKS	OTU	OSD	r	Yanlılık	RMSE	OMH
MFB-KY	GOO(FB:0.1)	İki	18	0.842	0.907	-0.107	0.424	0.294
		Üç	24.29	0.789	0.919	-0.109	0.403	0.269
		Dört	27.73	0.73	0.927	-0.107	0.386	0.256
	GA(%90)	İki	16	0.837	0.903	-0.107	0.433	0.302
		Üç	22.99	0.790	0.917	-0.113	0.409	0.276
		Dört	27.25	0.726	0.926	-0.107	0.389	0.258
	AOOT(FB:0.1)	İki	26	0.845	0.926	-0.112	0.39	0.257
		Üç	30.72	0.793	0.934	-0.109	0.372	0.243
		Dört	32.92	0.728	0.935	-0.11	0.372	0.237
MFB-KN	GOO(FB:0.1)	İki	18	0.864	0.897	-0.051	0.418	0.306
		Üç	24.58	0.809	0.937	-0.077	0.342	0.242
		Dört	28.16	0.735	0.941	-0.063	0.323	0.238
	GA(%90)	İki	16	0.855	0.889	-0.049	0.433	0.317
		Üç	23.43	0.803	0.933	-0.079	0.351	0.25
		Dört	27.65	0.734	0.94	-0.065	0.327	0.241
	AOOT(FB:0.1)	İki	27	0.871	0.93	-0.066	0.351	0.252
		Üç	31.27	0.81	0.948	-0.083	0.313	0.216
		Dört	33.92	0.733	0.952	-0.071	0.295	0.216
KLB-KY	GOO(FB:0.1)	İki	18	0.834	0.907	-0.113	0.428	0.294
		Üç	24.49	0.788	0.920	-0.115	0.403	0.268
		Dört	27.69	0.728	0.926	-0.111	0.393	0.257
	GA(%90)	İki	16	0.835	0.903	-0.113	0.436	0.301
		Üç	23.13	0.783	0.918	-0.116	0.406	0.273
		Dört	27.30	0.727	0.926	-0.109	0.388	0.257
	AOOT(FB:0.1)	İki	26	0.837	0.925	-0.114	0.396	0.262
		Üç	30.70	0.792	0.931	-0.112	0.384	0.249
		Dört	32.90	0.724	0.933	-0.115	0.378	0.243
KLB-KN	GOO(FB:0.1)	İki	18	0.862	0.895	-0.058	0.423	0.306
		Üç	24.60	0.802	0.936	-0.089	0.354	0.251
		Dört	28.12	0.732	0.935	-0.074	0.343	0.246
	GA(%90)	İki	17	0.859	0.89	-0.058	0.432	0.312
		Üç	23.30	0.794	0.933	-0.09	0.36	0.256
		Dört	27.72	0.733	0.934	-0.075	0.344	0.246
	AOOT(FB:0.1)	İki	27	0.866	0.924	-0.071	0.365	0.26
		Üç	31.14	0.805	0.947	-0.096	0.32	0.225
		Dört	33.86	0.733	0.949	-0.081	0.305	0.221

SKS: Sınıflama kategori sayısı, OTU: ortalama test uzunluğu, OSD: Ortalama sınıflama doğruluğu, OMH: Ortalama mutlak hata, r: Gerçek yeteneklerle kestirilen yetenekler arası korelasyon

Bireyi sınıflamak testi sonlandırmak için gereken en az madde sayısının GA sınıflama kriterinin kullanıldığı koşullardan elde edildiği görülmektedir. Bu araştırmada AOOT sınıflama kriteri ile birlikte en yüksek sınıflama doğruluğu ile sınıflama yapılmıştır. Farklı bir ifade ile GA sınıflama kriterine göre OSD bakımından AOOT'nin daha iyi performans gösterdiği söylenebilir. Araştırmanın bu bulgusu Reckase (1983) ve Thompson (2011) bulgularıyla uyumludur. Tablo 9'da bireylerin simülasyon öncesi hesaplanan gerçek yetenekleri ile BBST simülasyonu sonucu kestirilen son yetenek düzeyleri arasındaki korelasyon (r), yanlışlık, RMSE, OMH değerleri için MFB (KY-KN) madde seçme yöntemleriyle birlikte GOO sınıflama kriteri çaprazlandığında 0.907-0.897 korelasyon,-0.107,-0.051 yanlışlık, 0.424-0.418 RMSE ve 0.294-0.306 OMH değerleri hesaplanmıştır. GA sınıflama kriteri ile birlikte 0.903-0.897 korelasyon, -0.107,-0.049 yanlışlık,0.433-0.433 RMSE ve 0.302-0.317 OMH değerleri hesaplanmıştır. AOOT sınıflama kriteri ile birlikte 0.926-0.93 korelasyon, -0.112,-0.066 yanlışlık, 0.39-0.351 RMSE ve 0.257-0.252 OMH değerleri hesaplanmıştır. Bu bulgulara göre AOOT sınıflama kriteri ile birlikte en düşük yanlışlık en yüksek korelasyon değerleri olduğu söylenebilir. GA sınıflama kriteri ile en düşük korelasyon ve en yüksek yanlışlık değerlerinin hesaplandığı söylenebilir. Kestirimin standart hata değerlerini gösteren RMSE ve OMH değerleri için de AOOT sınıflama kriteri ile yapılan sınıflamada en düşük hata değerlerinin hesaplandığı anlaşılmaktadır. GA sınıflama kriteri ile en yüksek hata değerlerinin hesaplandığı söylenebilir. Buna göre ölçme kesinliği için (yanlışlık, standart hata ve korelasyon değerleri) AOOT sınıflama kriteri ile yapılan sınıflamanın performansının daha iyi olduğu test etkililiği (ortalama sınıflama doğruluğu ve ortalama test uzunluğu) için ise, OTU için GA sınıflama kriteri OSD için ise AOOT sınıflama kriterinin daha iyi performans gösterdiği yorumu yapılabilir. Madde seçme yöntemlerine göre KY temelli madde seçme yönteminin KN temelli yöntemle göre tüm koşullar için daha iyi performans gösterdiği, OSD için KN temelli madde seçme yönteminin performansının daha iyi olduğu tablodan anlaşılmaktadır. Araştırmanın bu bulgusu, Rackase (1983) ve Thompson (2011) araştırma sonuçlarıyla uyumludur. Buna karşılık

Gündeğer (2017) yaptığı çalışmada GA sınıflama kriterinin OTU açısından daha iyi OSD açısından AOOT ile benzer performans gösterdiği sonucuna ulaşmıştır.

İki kategorili sınıflamada KLB (KY-KN) madde seçme yöntemleri kullanıldığında GOO sınıflama kriteri ile birlikte 18-18 OTU ve 0.834-0.862 OSD ile sınıflama yapılmıştır. GA sınıflama kriteri ile birlikte 16-17 OTU ve 0.835-0.859 OSD ile AOOT sınıflama kriteri ile birlikte 26-27 OTU ve 0.837-0.866 OSD ile sınıflama yapıldığı anlaşılmaktadır. Bu bulgulara göre GA sınıflama kriterinin her iki madde seçme yöntemi ile en az madde ile sınıflama yaptığı söylenebilir. AOOT sınıflama kriterinin ise en çok maddeyle sınıflama yaptığı anlaşılmaktadır. OSD bakımından AOOT sınıflama kriterinin en yüksek doğrulukla sınıflama yaptığı araştırmanın bulguları arasındadır. KY temelli madde seçme yönteminin OTU olarak KN temelli madde seçme yönteminin ise OSD olarak daha iyi performans gösterdiği anlaşılmaktadır. Tablo 9'a göre gerçek yeteneklerle kestirilen yetenekler arası korelasyon yanlılık, kestirimin standart hataları RMSE ve OMH değerleri için GOO sınıflama kriteri ile yapılan sınıflamada 0.907-0.895 korelasyon, -0.113,-0.058 yanlılık, 0.428-0.423 RMSE ve 0.294-0.306 OMH değerleri hesaplanmıştır. GA sınıflama kriteri ile birlikte yapılan sınıflamada 0.918-0.89 korelasyon, -0.113,-0.058 yanlılık, 0.436-0.432 RMSE ve 0.301-0.312 OMH değerleri hesaplanmıştır. AOOT sınıflama kriteri ile birlikte 0.925-0.924 korelasyon, -0.114,-0.071 yanlılık, 0.396-0.365 RMSE ve 0.262-0.26 OMH değerleri hesaplanmıştır. Bu bulgulara göre ölçme kesinliği açısından en iyi performansı en yüksek korelasyon, en düşük yanlılık ve hata değerleri ile AOOT sınıflama kriterinin gösterdiği söylenebilir. GA sınıflama kriteri kullanıldığında standart hata olarak GOO sınıflama kriterinden daha yüksek değerler hesaplandığı yanlılık ve gerçek yetenekler ile kestirilen yetenekler arası korelasyon olarak benzer performans gösterdikleri söylenebilir. Ayrıca kestirilen yetenek temelli madde seçme yönteminin performansının KN temelli madde seçme yöntemine göre tüm koşullar için daha iyi olduğu araştırmanın bulguları arasındadır.

İki kategorili sınıflamada Tablo 9'da görüldüğü üzere gerçek veri seti üzerinden gerçekleştirilen BBST simülasyonunda BSD ile yeteneği kestirdiğimizde madde seçme

yöntemi farketmeksizin oluşturulan tüm koşullarda test etkililiği için GA sınıflama kriterinin diğer sınıflama kriterlerine göre daha başarılı performans gösterdiği söylenebilir. AOOT sınıflama kriterinin test etkililiği açısından OSD olarak performansı diğer sınıflama kriterlerine göre daha iyi OTU olarak en düşük olduğu araştırmanın bulguları arasındadır. Ölçme kesinliği için AOOT sınıflama kriterinin diğer sınıflama kriterlerine göre daha iyi performans gösterdiği araştırmanın bulguları arasındadır. Bununla birlikte MFB madde seçme yönteminin tüm koşullar için KLB madde seçme yöntemine göre iki kategorili sınıflamada daha iyi performans gösterdiği söylenebilir.

Üç kategorili sınıflamada Tablo 9'da BSD yetenek kestirimi kullanıldığında MFB (KY-KN) madde seçme yöntemleriyle sınıflama kriterlerinin çaprazlandığı koşullarda GOO sınıflama kriteri ile birlikte ortalama 24-24 madde ve 0.789-0.809 OSD, GA sınıflama kriteri ile 22-23 madde 0.790-0.803 OSD ve AOOT sınıflama kriteri ile birlikte 30-31 madde ve 0.793-0.81 OSD ile sınıflama yapıldığı anlaşılmaktadır. Bu bulgulara göre üç kategorili sınıflamada GA sınıflama kriterinin en az madde ile sınıflama yaptığı söylenebilir. KY temelli madde seçme yöntemi ile AOOT sınıflama kriteri, KN temelli madde seçme yöntemi ile GOO sınıflama kriterinin daha yüksek OSD ile sınıflama yaptığı anlaşılmaktadır. Bununla birlikte KN temelli madde seçme yönteminin daha yüksek OSD ile sınıflama yaptığı görülmektedir. Ölçme kesinliği için gerçek yeteneklerle kestirilen yetenekler arası korelasyon, yanlılık, kestirimin standart hatasını gösteren RMSE ve OMH değerlerine göre MFB (KY-KN) madde seçme yöntemleri ile GOO sınıflama kriteri çaprazlandığında 0.919-0.937 korelasyon, -0.109, -0.077 yanlılık, 0.403-0.342 RMSE ve 0.269-0.242 OMH değerleri, GA sınıflama kriteri kullanıldığında 0.917-0.933 korelasyon, -0.113, -0.079 yanlılık, 0.409-0.351 RMSE ve 0.276-0.25 OMH değerleri, AOOT sınıflama kriteri ile 0.934-0.948 korelasyon, -0.109, -0.083 yanlılık, 0.372-0.313 RMSE ve 0.243-0.216 OMH değerleri hesaplanmıştır. Bu bulgulara göre MFB (KY-KN) madde seçme yöntemleriyle AOOT sınıflama kriteri en yüksek korelasyon en düşük yanlılık ve standart hata değerleri ile ölçme kesinliği açısından en yüksek performansı gösterdiği söylenebilir. GA sınıflama kriterinin

hata değerlerinin diğer sınıflama kriterlerine göre en yüksek olduğu anlaşılmaktadır. Bunun yanı sıra KY temelli madde seçme yöntemleriyle hesaplanan korelasyonlar ve yanlılık değerleri KN temelli madde seçme yöntemlerine göre daha düşük, hata değerleri ise daha yüksektir. Yani KN temelli madde seçme yöntemleri hata ve korelasyon değerleri olarak daha iyi performans KY temelli madde seçme yöntemleri ise yanlılık değerleri olarak daha iyi performans gösterdiği araştırmanın bulguları arasındadır.

KLB (KY-KN) madde seçme yöntemleriyle sınıflama kriterlerinin birlikte kullanıldığı yeteneğin BSD yöntemiyle kestirildiği üç kategorili BBST için GOO sınıflama kriteri ile 24-24 OTU, 0.788-0.802 OSD ile sınıflama yapılmıştır. GA sınıflama kriteri ile 23-23 OTU, 0.783-0.794 OSD, AOOT sınıflama kriteri ile ortalama 30-31 madde ve 0.792-0.805 OSD ile sınıflama yapıldığı anlaşılmaktadır. Bu bulgulara göre GA sınıflama kriteri kullanıldığında en az maddeyle, AOOT sınıflama kriteri kullanıldığında en yüksek sınıflama doğruluğu ile sınıflama yapıldığı anlaşılmaktadır. KN temelli madde seçme yönteminin OSD için daha iyi performans gösterdiği OTU için KY temelli madde seçme yöntemi ile benzer performans gösterdiği araştırmanın bulguları arasındadır. Üç kategorili sınıflamada KLB madde seçme yöntemi ile test etkililiği için GA sınıflama kriterinin ve KN temelli madde seçme yönteminin daha iyi performans gösterdiği söylenebilir. Araştırmada ölçme kesinliğini gösteren korelasyon, yanlılık, kestirimin standart hata değerleri için ise GOO sınıflama kriteri ile birlikte 0.920-0.936 korelasyon, -0.115,-0.089 yanlılık, 0.403-0.354 RMSE ve 0.268-0.251 OMH değerleri, GA sınıflama kriteri ile birlikte 0.918-0.933 korelasyon, -0.116,-0.09 yanlılık, 0.406-0.36 RMSE ve 0.273-0.256 OMH değerleri, AOOT sınıflama kriteri ile birlikte 0.931-0.947 korelasyon, -0.112,-0.096 yanlılık, 0.384-0.32 RMSE ve 0.249-0.225 OMH değerleri hesaplandığı anlaşılmaktadır. Tablo 9'a göre yanlılık, hata değerlerinin düşük korelasyon değerlerinin yüksek olması ölçme kesinliği için beklenen durumdur. AOOT sınıflama kriterinin diğer sınıflama kriterlerine göre ölçme kesinliği açısından daha iyi performans gösterdiği bulgulara göre söylenebilir. GA sınıflama kriterinin ölçme kesinliği olarak performansı diğer sınıflama kriterlerine göre daha düşük olduğu görülmektedir. KN temelli

madde seçme yönteminin performansı KY temelli madde seçme yöntemine göre kestirilen yeteneklerle gerçek yetenekler arası korelasyon olarak daha yüksek, yanlışlık, hata değerleri olarak daha düşük olduğu anlaşılmaktadır.

Özetle Tablo 9'da BSD yetenek kestirimi kullanıldığında MFB-KLB (KY-KN) madde seçme yöntemleriyle ile yapılan üç kategorili sınıflamada MF-KN temelli madde seçme yöntemi ile GA sınıflama kriterinin çaprazlanmasından oluşan koşulun test etkililiği yüksek sınıflama doğruluğu ile sınıflama yaptığı görülmektedir. MFB-KY madde seçme yöntemi ile en az OTU ile sınıflama yapılırsa da MFB-KN madde seçme yöntemi ile OTU açısından çok fark olmadığından MFB-KN madde seçme yöntemi ile yapılan sınıflamanın daha etkili olduğu söylenebilir. Ölçme kesinliği için AOOT sınıflama kriterinin MFB-KN madde seçme yöntemi ile çaprazlandığı koşullara ait sınıflamanın performansının yüksek olduğu araştırmamızın bulgularındandır. Buna ek olarak tüm koşullar için MFB-KN madde seçme yönteminin kullanıldığı koşullarda hesaplanan değerlerin diğer madde seçme yöntemlerine göre daha etkili olduğu anlaşılmaktadır. Araştırmamızın bu bulgusu Eggen (1999) araştırma sonuçlarıyla da desteklenmektedir.

Dört kategorili sınıflamada Tablo 9'da BSD yetenek kestirimi ile MFB (KY-KN) madde seçme yöntemleri ile sınıflama kriterlerinin çaprazlandığı koşullara ait GOO sınıflama kriteri ile ortalama 27-27 madde ile 0.73-0.728 OSD, GA sınıflama kriteri ile 27-27 OTU ve 0.726-0.733 OSD, AOOT sınıflama kriteri ile 32-33 OTU ve 0.728-0.733 OSD ile sınıflama yapıldığı anlaşılmaktadır. Bu bulgular doğrultusunda GOO sınıflama kriteri KY temelli madde seçme yöntemiyle GA sınıflama kriteri ise KN temelli madde seçme yöntemi ile test etkililiği açısından en az madde ve en yüksek sınıflama doğruluğu ile en iyi performansı gösterdiği söylenebilir. AOOT sınıflama kriteri ise diğer sınıflama kriterlerine göre daha yüksek sayıda madde ile sınıflama yaptığı OSD doğruluğu olarak benzer performans gösterdiği yorumu yapılabilir. KY temelli madde seçme yöntemi ile KN temelli madde seçme yöntemi OTU olarak benzer performans gösterdiği OSD için KN temelli madde seçme yönteminin daha yüksek sınıflama doğruluğu ile sınıflama yaptığı

bulgulardan anlaşılmaktadır. Tablo 9'da gerçek yetenekler ile kestirilen yetenekler arası korelasyon yanlılık değerleri kestirimin standart hatasını gösteren RMSE ve OMH değerleri ise MFB (KY-KN) madde seçme yöntemleriyle GOO sınıflama kriteri çaprazlandığında 0.927-0.941 korelasyon, -0.107,-0.063 yanlılık, 0.386-0.323 RMSE, 0.256-0.238 OMH değerleri, GA sınıflama kriteri ile 0.926-0.94 korelasyon, -0.107,-0.065 yanlılık, 0.389-0.327 RMSE ve 0.258-0.241 OMH değerleri, AOOT sınıflama kriteri ile 0.935-0.952 korelasyon, -0.11,-0.071 yanlılık, 0.372-0.295 RMSE ve 0.237-0.216 OMH değerleri hesaplanmıştır. Bu bulgulara göre AOOT sınıflama kriteri ile KY ve KN temelli her iki madde seçme yönteminin oluşturulduğu koşullarda gerçek yeteneklerle kestirilen yetenekler arası korelasyonlar en yüksek, yanlılık, RMSE, OMH değerleri diğer sınıflama kriterleri ile oluşturulan koşullara göre en düşüktür. AOOT sınıflama kriterinin ölçme kesinliği için diğer kriterlere göre daha iyi performans gösterdiği söylenebilir. Ayrıca KN temelli madde seçme yönteminin ölçme kesinliği açısından KY temelli madde seçme yöntemine göre daha iyi performans gösterdiği araştırmanın bulgularından anlaşılmaktadır. GA sınıflama kriterinin ise diğer kriterlere göre dört kategorili sınıflamada ölçme kesinliği olarak performansının en düşük olduğu araştırmanın bulgularına göre söylenebilir.

KLB (KY-KN) madde seçme yöntemleriyle sınıflama kriterlerinin kullanıldığı koşullara ait dört kategorili sınıflamada GOO sınıflama kriteri ile 27-28 OTU ve 0.728-0.732 OSD değerleri, GA sınıflama kriteri ile 27-27 OTU ve 0.727-0.733 OSD değerleri, AOOT sınıflama kriteri ile 32-33 OTU ve 0.724-0.733 OSD değerleri hesaplanmıştır. GA sınıflama kriterinin diğer sınıflama kriterlerine göre daha az madde ile daha yüksek sınıflama doğruluğu ile sınıflama yaptığı, test etkililiğinin daha yüksek olduğu araştırmanın bulgularından anlaşılmaktadır. AOOT sınıflama kriterinin diğer kriterlere göre daha yüksek sayıda madde ile testi sonlandırdığı yorumu yapılabilir. Ayrıca KN temelli madde seçme yönteminin daha yüksek OSD ile sınıflama yaptığı KY temelli madde seçme yönteminin KN temelli madde seçme yöntemine göre daha az madde ile sınıflama yaptığı araştırmanın bulgularına göre söylenebilir. Tablo 9'da bireylerin gerçek yetenek düzeyleri ile kestirilen

yetenek düzeyleri arasındaki korelasyon, yanlışlık, kestirimin standart hata değerlerini gösteren RMSE ve OMH değerleri olarak GOO sınıflama kriteri ile yapılan sınıflamada 0.926-0.935 korelasyon, -0.111,-0.074 yanlışlık, 0.393-0.343 RMSE, 0.257-0.246 OHM, GA sınıflama kriteri ile 0.926-0.934 korelasyon, -0.109,-0.075 yanlışlık, 0.388-0.344 RMSE ve 0.257-0.246 OMH, AOOT sınıflama kriteri ile 0.933-0.949 korelasyon, -0.115,-0.081 yanlışlık, 0.378-0.305 RMSE ve 0.243-0.221 OMH değerleri hesaplanmıştır. Buna göre gerçek yetenekler ile kestirilen yetenekler arasındaki en yüksek korelasyon, en düşük yanlışlık ve standart hata değerlerinin AOOT sınıflama kriterinin kullanıldığı koşullarda hesaplandığı anlaşılmaktadır. Bu bulgulara göre AOOT sınıflama kriterinin ölçme kesinliği için en iyi performansı gösteren yöntem olduğu söylenebilir. Ayrıca KN temelli madde seçme yönteminin korelasyon ve standart hata değerleri için KY temelli madde seçme yönteminden daha iyi performans gösterdiği, KY temelli madde seçme yönteminin ise daha düşük yanlışlıkla testi sonlandırdığı ya da sınıflama yaptığı yorumu yapılabilir.

Özetle Tablo 9'daki değerlere göre BBST simülasyonunda yetenek kestirim yöntemi BSD olduğunda MFB-KLB (KY-KN) madde seçme yöntemleriyle sınıflama kriterlerinin çaprazlandığı tüm koşullarda sınıflama kategori sayısı arttıkça testin sonlanması için gereken madde sayısı artmış OSD azalmıştır. Bu durumda kategori sayısı arttıkça madde havuzunda madde sayısı azaldığı için test etkililiğinin düşmüş olduğu söylenebilir. Bununla birlikte sınıflama kategori sayısı arttıkça RMSE, OMH değerleri düşmüş ancak korelasyon değeri yükselmiştir. Bu durumda sınıflama kategori sayısı arttıkça testin sonlanması için madde sayısı artmış olması nedeniyle bireylerin son yetenek düzeylerinin daha hassas kestirildiği ölçme kesinliğinin arttığı söylenebilir. İki kategorili sınıflamada GA sınıflama kriterinin ölçme kesinliği olarak, AOOT sınıflama kriterinin ise test etkililiği olarak en iyi performansı gösterdiği anlaşılmaktadır. Ayrıca MFB madde seçme yönteminin KLB madde seçme yöntemine göre tüm koşullarda daha iyi performans gösterdiği araştırmanın bulgularındandır. Üç kategorili sınıflamada MFB-KN madde seçme yöntemi ile GA sınıflama kriterinin test etkililiği bakımından, AOOT sınıflama kriteri ile çaprazlanmasının ölçme

kesinliđi olarak başarılı performans gösterdiđi anlaşılmaktadır. Dört kategorili sınıflamada GA sınıflama kriterinin KN temelli madde seçme yöntemiyle, GOO sınıflama kriterinin KY temelli madde seçme yöntemiyle test etkililiđi olarak başarılı performans gösterdiđi anlaşılmaktadır. AOOT sınıflama kriterinin ise dört kategorili sınıflamada tüm koşullar için ölçme kesinliđi olarak en iyi performansı gösterdiđi yorumu yapılabilir. MFB için KY ve KN temelli madde seçme yöntemlerinin OTU için benzer OSD için KN temelli madde seçme yönteminin daha iyi performans gösterdiđi anlaşılmaktadır. KLB için KN temelli madde seçme yönteminin OSD için, KY temelli madde seçme yönteminin ise OTU için daha iyi performans gösterdiđi yorumu arařtırmanın bulgularına göre yapılabilir.

Dördüncü Alt Probleme Ait Bulgular Ve Yorumlar

Arařtırmanın dördüncü alt probleminde ortalama test uzunluđu, ortalama sınıflama dođruluđu ve ölçme kesinliđinin iki, üç, dört kategorili sınıflamada madde seçme yöntemlerine göre nasıl deđiřtiđi incelenmiřtir. Tüm deđerler 25 tekrarın ortalaması alınarak Tablo 10'da sunulmuřtur.

Tablo 10'a göre madde seçme yöntemleri dıřındaki bađımsız deđiřkenler dikkate alınmadan iki üç ve dört kategorili sınıflamada sınıflama kategori sayısı arttıkça OTU arttıđı, OSD azaldıđı anlaşılmaktadır. MFB-KY madde seçme yöntemi ortalama 19 madde 0.836 OSD ile sınıflama yaptıđı test etkililiđi olarak diđer madde seçme yöntemlerine göre daha etkili olduđu söylenebilir. MFB-KN madde seçme yönteminin iki kategorili sınıflamada en yüksek sınıflama dođruluđu ile sınıflama yaptıđı arařtırmanın bulguları arasındadır. MFB madde seçme yönteminin KLB madde seçme yöntemine göre tüm kategorilerde daha az madde daha yüksek sınıflama dođruluđu ile sınıflama yaptıđı görölmektedir. Madde seçme yöntemlerinden en iyi performansı MFB-KY'in gösterdiđi bunu KLB-KY'in takip ettiđi MFB-KN'ın KLB-KN dan daha iyi performans gösterdiđi görölmektedir. Bununla birlikte bireylerin gerçek yetenekleri ve kestirilen yetenekleri arasındaki korelasyon yanlılık ve hata deđerleri için sınıflama kategori sayısı arttıkça korelasyon deđerlerinin yükseldiđi hata deđerlerinin düřtüđu görölmektedir.

Tablo 10

Madde Seçme Yöntemlerine Göre İki, Üç, Dört Kategorili Sınıflamaya Ait Test Etkliliği Ve Ölçme Kesinliği Değerleri

Koşullar		Bağımlı değişkenler					
Madde seçme yöntemleri	SKS	OTU	OSD	r	Yanlılık	RMSE	OMH
MFB-KY	İki	19.98	0.836	0.91	-0.123	0.435	0.296
	Üç	25.95	0.787	0.921	-0.121	0.412	0.272
	Dört	29.28	0.728	0.927	-0.118	0.399	0.26
MFB-KN	İki	20.84	0.862	0.906	-0.06	0.401	0.289
	Üç	26.45	0.803	0.932	-0.096	0.366	0.252
	Dört	30.09	0.727	0.943	-0.073	0.321	0.234
KLB-KY	İki	20.18	0.83	0.908	-0.13	0.441	0.299
	Üç	26.02	0.779	0.918	-0.128	0.421	0.277
	Dört	29.28	0.721	0.924	-0.125	0.409	0.266
KLB-KN	İki	20.83	0.859	0.903	-0.067	0.408	0.292
	Üç	26.41	0.795	0.931	-0.105	0.374	0.257
	Dört	30.21	0.727	0.938	-0.085	0.335	0.24

SKS: Sınıflama kategori sayısı, OTU: Ortalama test uzunluğu, OSD: Ortalama sınıflama doğruluğu, OMH: Ortalama mutlak hata, r: Gerçek yeteneklerle kestirilen yetenekler arası korelasyon

Literatürde bu karşılaştırmaya çok kategorili sınıflama için fazla sayıda çalışmada yer verilmediği için tartışılmamıştır. Bunlar çalışmaya özgü bulgulardır. Daha sonraki alan çalışmalarına katkı sağlaması beklenmektedir. İki kategorili sınıflama sonuçları Eggen ve Straetmans'ın (2000) çalışmasıyla çok uyumlu değildir. Eggen (1998) tam tersi bulguyla MFB-KN temelli madde seçme yönteminin KY temelli yöntemine göre daha başarılı olduğu yorumunu yapmıştır. MFB-KN madde seçme yöntemi ile yapılan dört kategorili sınıflamanın ölçme kesinliği açısından daha kullanışlı olacağı söylenebilir. İki kategorili sınıflamada MF-KN madde seçme yönteminin en düşük yanlılıkla sınıflama yaptığı araştırmanın bulguları arasındadır. MFB madde seçme yönteminin KLB madde seçme yöntemine göre daha iyi performans gösterdiği görülmektedir. Buna göre BBST uygulamalarında MFB-KY madde

seçme yöntemi tercih edildiğinde test etkililiği, ölçme kesinliği bakımından tüm koşullar için daha kullanışlı olduğu söylenebilir. Thompson (2007, 2009) iki kategorili sınıflamada kestirilen yetenek düzeyinde en çok bilgi veren maddelerin seçilmesi durumunda testin kıaldığını belirtmektedir. Thompson'un bu bulgusu araştırmının bulgularını desteklemektedir. Çok kategorili sınıflama için literatürde fazla çalışma olmadığı için ortak bir sonuca ulaşılamamıştır.

Beşinci Alt Probleme Ait Bulgular Ve Yorumlar

Araştırmının beşinci alt probleminde test uzunluğu, sınıflamanın doğruluğu ve ölçmenin kesinliği yetenek kestirim yöntemlerine göre nasıl değiştiği incelenmiştir.

Araştırmının beşinci alt probleminde gerçek veri seti ile yapılan BBST simülasyonunda yetenek kestirim yöntemleri dışındaki bağımsız değişkenler dikkate alınmadan ortalama test uzunluğu, ortalama sınıflama doğruluğu, gerçek yetenekler ile kestirilen yetenekler arası korelasyon, yanlılık, RMSE ve ortalama mutlak hata değerlerinin iki üç dört kategorili sınıflamada yetenek kestirim yöntemlerine göre nasıl değiştiği incelenmiştir. Tüm değerler için 25 tekrarın ortalaması alınarak hesaplanan değerler Tablo 11'de sunulmuştur.

Tablo 11'de Tüm yetenek kestirim yöntemleri için sınıflama kategori sayısı arttıkça OTU, korelasyon, yanlılık değerlerinin arttığı, OSD, RMSE ve OMH değerlerinin azaldığı görülmektedir. Buna göre BSD yetenek kestiriminin iki kategorili sınıflamada en az madde ve en yüksek sınıflama doğruluğu ile sınıflama yaptığı görülmektedir. BSD yetenek kestiriminin tüm koşullarda MOK ve AOK yetenek kestirim yöntemlerine göre test etkililiği bakımından daha iyi performans gösterdiği anlaşılmaktadır. MOK yetenek kestiriminin AOK yetenek kestirimine göre OTU olarak benzer OSD olarak daha yüksek performans gösterdiği araştırmının bulguları arasındadır. Yi, Wang ve Ban (2000) araştırmasında elde edilen iki kategorili sınıflama için AOK'un BSD'dan daha fazla sayıda madde ile sınıflama yaptığı bulgusuyla araştırmının bulgusu uyumaktadır. MOK yetenek kestiriminin performansının AOK'dan daha etkili olduğu bu araştırmada elde edilen bulgudur. Hata ve korelasyon değerleri olarak BSD yetenek kestiriminin en iyi performansı gösterdiği AOK

yetenek kestiriminin korelasyon ve hata deęerleri olarak MOK yetenek kestiriminden daha iyi performans gösterdięi sylenebilir.

Tablo 11

Yetenek Kestirimlerine Gre Test Etkililięi Ve lme Kesinlięi Deęerleri

Koşullar		Bağımlı Deęişkenler					
Yetenek Kestirim Yöntemleri	SKS	OTU	OSD	r	Yanlılık	RMSE	OMH
MOK	İki	21.46	0.864	0.846	-0.056	0.688	0.439
		26.31	0.801	0.907	-0.155	0.583	0.374
	Drt	30.25	0.726	0.903	-0.045	0.479	0.303
AOK	İki	21.19	0.859	0.905	-0.072	0.405	0.291
		26.50	0.794	0.928	-0.116	0.411	0.278
	Drt	30.33	0.718	0.939	-0.084	0.335	0.242
BSD	İki	20.55	0.863	0.904	-0.059	0.405	0.292
		26.42	0.804	0.939	-0.086	0.341	0.24
	Drt	29.91	0.734	0.942	-0.072	0.324	0.235
BMK	İki	20.75	0.859	0.907	-0.06	0.404	0.288
		26.38	0.799	0.934	-0.1	0.352	0.244
	Drt	30.08	0.73	0.942	-0.081	0.324	0.234

SKS: Sınıflama kategori sayısı, OTU: Ortalama test uzunluęu, OSD: Ortalama sınıflama doęruluęu, OMH: Ortalama mutlak hata, r: Gerek yeteneklerle kestirilen yetenekler arası korelasyon

Yanlılık olarak AOK yetenek kestiriminin daha dşk yanlılık deęerleri ile sınıflama yaptığı grlmektedir. zetle Tablo 11'de BSD test etkililięi ve lme kesinlięi olarak en iyi performansı gsterdięi ve MOK'un test uzunluęu ve sınıflama doęruluęu aısından AOK'dan daha yksek performans gsterdięi, AOK'un korelasyon ve hata deęerleri yani lme kesinlięi olarak MOK'dan daha etkili olduęu anlaşılmaktadır. Literatrde daha ok iki kategorili sınıflama iin BSD ve AOK yetenek kestiriminin performansının karşılaştırdığı az

sayıda çalışma mevcuttur. Örneğin Gündeğer (2017) iki kategorili sınıflama için AOK'un BSD yetenek kestiriminden daha fazla sayıda madde ile sınıflama yaptığı sonucuna ulaşmıştır ve bu araştırmayla uyumludur. Çok kategorili sınıflamaya ait MOK ve BSD yetenek kestirimlerinin performanslarına ait çalışmaya rastlanmadığı için literatürle tartışılmamıştır. Sonraki alan çalışmalarına katkı sağlayacağı düşünülmektedir.

Altıncı Alt Probleme Ait Bulgular Ve Yorumlar

Araştırmanın altıncı alt probleminde gerçek veri seti ile yapılan BBST simülasyonunda sınıflama kriterleri dışındaki bağımsız değişkenler dikkate alınmadan ortalama test uzunluğu, ortalama sınıflama doğruluğu, gerçek yetenekler ile kestirilen yetenekler arası korelasyon, yanlılık, RMSE ve ortalama mutlak hata değerlerinin iki üç dört kategorili sınıflamada sınıflama kriterlerine göre nasıl değiştiği incelenmiştir. Tüm değerler için 25 tekrarın ortalaması alınarak hesaplanan değerler Tablo 12'de sunulmuştur.

Tablo 12

Sınıflama Kriterlerine Göre Test Etkililiği Ve Ölçme Kesinliği Değerleri

Sınıflama kriterleri	Bağımlı değişkenler						
	SKS	OTU	OSD	r	Yanlılık	RMSE	OMH
GOO (FB:0.1)	İki	18.07	0.861	0.896	-0.06	0.421	0.304
	Üç	24.68	0.799	0.927	-0.097	0.38	0.264
	Dört	28.14	0.727	0.937	-0.077	0.337	0.244
GA (%90)	İki	17.31	0.855	0.891	-0.063	0.429	0.31
	Üç	23.39	0.796	0.925	-0.101	0.385	0.268
	Dört	28.28	0.727	0.936	-0.08	0.338	0.244
AOOT(FB:0.1)	İki	27.11	0.866	0.926	-0.067	0.361	0.257
	Üç	31.22	0.802	0.943	-0.104	0.344	0.232
	Dört	33.89	0.728	0.949	-0.08	0.307	0.222

SKS: Sınıflama kategori sayısı, OTU: Ortalama test uzunluğu, OSD: Ortalama sınıflama doğruluğu, r: Gerçek yeteneklerle kestirilen yetenekler arası korelasyon, OMH: Ortalama mutlak hata

Tablo 12'de en az sayıda yaklaşık 17 madde ile GA sınıflama kriterinin iki kategorili sınıflamada sınıflama yaptığı görülmektedir. OTU bakımından GA sınıflama kriterinin en iyi

performansı gösterdiği bunu GOO ve AOOT sınıflama kriterlerinin takip ettiği anlaşılmaktadır. OSD olarak AOOT sınıflama kriterinin iki kategorili sınıflamada 0.866 ile en yüksek performansı gösterdiği görülmektedir. GOO sınıflama kriteri GA sınıflama kriterinden OSD olarak daha iyi performans gösterdiği anlaşılmaktadır. Buna göre BBST için az sayıda madde ile yüksek sınıflama doğruluğu ile sınıflama yapılması amaçlandığından GA ve GOO sınıflama kriterinin gerçek BBST uygulamalarında kullanılmasının uygun olacağı anlaşılmaktadır. Gerçek yeteneklerle kestirilen yetenekler arası korelasyon yanlılık, RMSE VE OMH değerlerine göre başka bir ifade ile ölçme kesinliği için sınıflama kategori sayısı arttıkça korelasyon değerlerinin arttığı, yanlılık ve hata değerlerinin azaldığı araştırmanın bulgularından anlaşılmaktadır. AOOT sınıflama kriterinin iki, üç ve dört kategorili tüm sınıflama koşullarında diğer sınıflama kriterlerine göre ölçme kesinliği olarak daha yüksek performansla sınıflama yaptığı anlaşılmaktadır. BBST uygulamalarında AOOT sınıflama kriteri dört kategorili sınıflamada en yüksek ölçme kesinliği performansı gösterse de 33 madde ile sınıflama yaptığı için yöntemin kullanışlı olmayacağı nedeniyle GOO sınıflama kriterinin 18 madde ve 0.861 OSD ile sınıflama yapması hata değerlerinin de GA sınıflama kriteriyle hesaplanan hata değerlerinden daha düşük olması sebebiyle daha kullanışlı bir yöntem olduğu söylenebilir. Bu bulgu iki kategorili sınıflama için Thompson (2011), Nozawa ve Zhu'nun (2012) araştırma sonuçlarıyla örtüşmektedir. Çok kategorili sınıflama için literatürde Demir (2019) GA ve AOOT sınıflama kriterinin performansını simülasyon verisi ile pratik kısıtlamalar altında araştırmıştır. GA ve AOOT sınıflama kriterinin performansı ile Demir'in (2019) çalışması nispeten uyumlu olmakla birlikte GOO sınıflama kriterinin performansının da dahil olduğu çalışmaya rastlanmadığı için literatürle çok kategorili sınıflama için tartışılmamıştır. Daha sonraki çalışmalarda alana katkı sağlayacağı düşünülmektedir. Sonuç olarak sınıflama kategori sayısı arttıkça test etkililiğinin azaldığı ölçme kesinliğinin arttığı anlaşılmaktadır.

Yedinci Alt Probleme Ait Bulgular Ve Yorumlar

Araştırmanın yedinci alt probleminde gerçek veri seti ile yapılan BBST simülasyonunda sınıflama kategori sayısı dışındaki bağımsız değişkenler dikkate alınmadan test etkililiği ve ölçme kesinliğinin nasıl değiştiği incelenmiştir. Tüm koşullar için 25 tekrarın ortalaması alınarak hesaplanan değerler Tablo 13'te sunulmuştur.

Tablo 13

Sınıflama Kategori Sayısına Göre Test Etkililiği Ve Ölçme Kesinliği Değerleri

Koşullar	Bağımlı değişkenler					
	Sınıflama kategori sayısı	OTU	OSD	r	Yanlılık	RMSE
İki kategorili	20.83	0.861	0.904	-0.064	0.405	0.29
Üç kategorili	26.43	0.799	0.932	-0.1	0.37	0.254
Dört kategorili	30.11	0.727	0.941	-0.079	0.328	0.237

OTU: Ortalama test uzunluğu, OSD: Ortalama sınıflama doğruluğu, r: gerçek yeteneklerle kestirilen yetenekler arası korelasyon

Tablo 13'e göre sınıflama kategori sayısı arttıkça test uzunluğunun arttığı ortalama sınıflama doğruluğunun azaldığı söylenebilir. Bu durumda test etkililiğinin kategori sayısı arttıkça düştüğü görülmektedir. Çok kategorili sınıflamada hata değerlerinin azaldığı, korelasyon değerlerinin arttığı anlaşılmaktadır. Ölçme kesinliğinin arttığı görülmektedir. Sınıflamak için kategori sayısı arttıkça madde havuzundaki madde sayısı azaldığı için OTU arttığı ve daha hassas ölçme yapıldığı bulgulara göre söylenebilir. BBST uygulamalarında az sayıda madde ile yüksek doğrulukla sınıflama yapmak amaç olduğu için az kategoride yapılan sınıflamaların daha avantajlı olduğu düşünülebilir. Eggen (1999) ve Nydick ve diğerleri (2012) tarafından yapılan çalışmalarda da kategori sayısının artması OTU'nu artırırken OSD 'nu azalttığı görülmüştür.

Sekizinci Alt Probleme Ait Bulgular Ve Yorumlar

Araştırmanın sekizinci alt probleminde iki, üç, dört kategorili BBST uygulamasına ait tüm analiz bulguları doğrultusunda en uygun BBST testi algoritmasının nasıl olduğu araştırılmıştır.

Yapılan analizlere göre BBST uygulaması ile yapılan sınıflamada az madde ile yüksek doğrulukta sınıflama yapılması farklı bir ifade ile test etkililiğinin yüksek olması ve sınıflamanın standart hata değerlerinin ve yanlılığın düşük, kestirilen yeteneklerle gerçek yetenekler arası korelasyonun yüksek olması başka bir ifade ile ölçme kesinliğinin yüksek olması beklenir. Bu doğrultuda yapılan analizlerde Maksimum olabilirlik yetenek kestirimi ile yeteneğin kestirildiği iki, üç, dört kategorili sınıflamaların tamamı için GA sınıflama kriteri ile MFB-KY temelli madde seçme yönteminin çaprazlandığı koşullarla yapılan sınıflamanın tüm kategoriler için en uygun test algoritması olduğu analiz bulgularından anlaşılmaktadır.

Ağırlıklı olabilirlik yetenek kestirimi ile yeteneğin kestirildiği BBST uygulamasında GA sınıflama kriteri daha az maddeyle sınıflama yapsa da GOO sınıflama kriterinin MFB-KN temelli madde seçme yöntemi ile sınıflamanın doğruluğu ve kestirimin standart hata değeri için daha etkili olduğu yapılan analizlerden anlaşılmaktadır. Bu bulgu doğrultusunda AOK yetenek kestirimi ile GOO sınıflama kriteri ve MFB-KN madde seçme yönteminin oluşturduğu koşulların iki ve üç kategorili sınıflama için en uygun test algoritması olduğu söylenebilir. Dört kategorili sınıflamada GA sınıflama kriterinin MFB-KN temelli madde seçme yöntemi ile oluşturduğu koşulların BBST için en uygun test algoritması olduğu analiz bulgularından anlaşılmaktadır.

BSD yetenek kestirimi ile yeteneğin kestirildiği test algoritmasında GOO sınıflama kriteri GA sınıflama kriterine göre biraz daha fazla madde ile sınıflama yapsa da sınıflamanın doğruluğunun yüksek olması ve hata değerlerinin daha düşük olması nedeni ile KLB-KY madde seçme yöntemi ile oluşturulan algoritmanın iki kategorili sınıflama için en uygun koşul olduğu anlaşılmaktadır. Üç ve dört kategorili sınıflamalar için GOO sınıflama kriteri MFB-KN temelli madde seçme yöntemi ile en uygun test algoritmasını oluşturduğu analizlere göre söylenebilir.

Bölüm 5

Sonuç ve Öneriler

Beşinci bölümde araştırmanın bulgularına göre ulaşılan sonuçlara ve bu sonuçlar ile geliştirilen önerilere yer verilmiştir.

Sonuçlar

Bu araştırmada 919 bireye uygulanan 256 maddeye ait İngilizce seviye belirleme sınavının gerçek veri seti üzerinde BBST uygulanabilirliği araştırılmıştır. Bu amaçla sınıflama kriterleri AOOT, GA, GOO, madde seçme yöntemleri MFB(KY-KN), KLB (KY-KN) sınıflama kategori sayıları iki, üç ve dört, MOK, AOK, BSD yetenek kestirim yöntemlerinin kullanıldığı durumlarda OTU, OSD, bireylerin gerçek yetenekleri ile kestirilen yetenekleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri gerçek veriye dayalı simülasyon aracılığı ile incelenmiştir. Oluşturulan 108 koşul için araştırma yapılmış ve araştırmanın sekiz alt problemine yanıt aranmıştır. Araştırmanın sonuçları aşağıda sunulmuştur.

1. Araştırmada MOK yetenek kestirimi kullanıldığında GA sınıflama kriteri iki kategorili sınıflamada MFB(KY-KN) madde seçme yöntemi ile, GOO sınıflama kriterinin ise her iki madde seçme yönteminde KN temelli madde seçme yöntemi ile test etkililiği (OSD, OTU) için daha iyi performans gösterdiği sonucuna ulaşılmıştır. AOOT sınıflama kriterinin ise MFB-KY madde seçme yöntemi ile ölçme kesinliği farklı bir ifade ile gerçek yetenekler ile kestirilen yetenekler arası korelasyon, yanlılık, RMSE, OMH değerleri olarak en iyi performansı gösterdiği sonucuna ulaşılmıştır.

2. Üç kategorili sınıflamada MFB-KN madde seçme yöntemi ile GA sınıflama kriterinin OTU ve OSD için yüksek performansa sahip olduğu, AOOT sınıflama kriterinin ise MFB-KY madde seçme yöntemi ile çaprazlandığında ölçme kesinliği yüksek sınıflama yaptığı belirlenmiştir. GOO sınıflama kriteri MFB-KN madde seçme yöntemi ile etkili performans göstermesiyle birlikte sınıflamanın standart hata değerleri farklı bir ifade ile ölçme kesinliğinin düşük olduğu sonucuna ulaşılmıştır.

3. Dört kategorili sınıflamada GA sınıflama kriterinin MFB (KY-KN) her iki madde seçme yöntemiyle de OTU ve OSD için en yüksek performansı göstermiştir. Ölçme kesinliği için KN temelli madde seçme yöntemleriyle GA sınıflama kriterinin KY temelli madde seçme yöntemleriyle AOOT sınıflama kriterinin yüksek performans gösterdiği sonucuna ulaşılmıştır. GOO sınıflama kriterinin ise ölçme kesinliği olarak performansının en düşük olduğu sonuçlanmıştır.

4. Madde seçme yöntemleri ile sınıflama kriterlerinin çaprazlandığı tüm koşullar için MOK yetenek kestirimi kullanıldığında KN temelli madde seçme yöntemlerinin OTU ve OSD için daha yüksek performans gösterdiği doğru sınıflama oranının yüksek olduğu belirlenmiştir.

5. Yetenek kestirimi MOK olduğu sınıflamada sınıflama kategori sayısı arttıkça OTU'nun arttığı ve OSD düştüğü, hata değerlerinin azaldığı farklı bir ifade ile test etkililiğinin düştüğü ölçme kesinliğinin yükseldiği çok kategorili sınıflamada ölçmenin daha hassas olduğu belirlenmiştir.

6. Yetenek kestirimi AOK olduğunda iki kategorili sınıflamada MFB-KY madde seçme yöntemi ile GA sınıflama kriteri en az madde en yüksek sınıflama doğruluğu ile sınıflama yaptığı belirlenmiştir. KN temelli madde seçme yöntemleriyle yapılan sınıflamada KY temelli madde seçme yöntemlerine göre OSD'nun daha yüksek olduğu belirlenmiştir. AOOT sınıflama kriterinin MFB (KY-KN) madde seçme yöntemleriyle birlikte kullanıldığında en düşük hata en yüksek korelasyonla sınıflama yaptığı sonucuna ulaşılmıştır.

7. KN temelli madde seçme yöntemleriyle GOO sınıflama kriterinin AOOT sınıflama kriterinden daha az madde ile sınıflama yaptığı sonucuna ulaşılmıştır.

8. Üç kategorili sınıflamada tüm madde seçme yöntemleriyle GA sınıflama kriterinin en az sayıda madde ile sınıflama yaptığı belirlenmiştir. Özellikle MFB-KN madde seçme yöntemi ile başarılı performans gösterdiği sonucuna ulaşılmıştır. AOOT sınıflama kriterinin ise üç kategorili sınıflamada MFB-KN madde seçme yöntemiyle birlikte en düşük hata ve

yanlılık değerleriyle sınıflama yaptığı sonucuna ulaşılmıştır. Üç kategorili sınıflamada MFB madde seçme yöntemlerinin KLB madde seçme yöntemlerinden daha iyi performans gösterdiği belirlenmiştir.

9. Dört kategorili sınıflamada GA sınıflama kriterinin MFB (KY-KN) madde seçme yöntemlerinin ikisiyle de OTU ve OSD için yüksek performans gösterdiği, GOO sınıflama kriterinin de KLB (KY-KN) madde seçme yöntemleriyle OTU ve OSD için yüksek performans gösterdiği sonucuna ulaşılmıştır.

10. AOK yetenek kestirimi kullanıldığında MFB madde seçme yönteminin KLB madde seçme yöntemine göre daha iyi performans gösterdiği belirlenmiştir. Dört kategorili sınıflamada MFB madde seçme yönteminde KY temelli madde seçme yönteminin, KLB madde seçme yönteminde ise KN temelli yöntemin ortalama madde sayısı olarak daha iyi performans gösterdiği anlaşılmıştır.

11. Tüm koşullarda AOK yetenek kestirimi ile sınıflama yapıldığında sınıflama kategori sayısı arttıkça OTU arttığı OSD ise azaldığı farklı bir ifade ile test etkililiğinin düştüğü, ölçme kesinliğinin yükseldiği sonuçlarına ulaşılmıştır.

12. BSD yetenek kestirimi kullanıldığında madde seçme yöntemi fark etmeksizin ortalama test uzunluğu, ortalama sınıflama doğruluğu için iki kategorili sınıflamada GA sınıflama kriterinin en iyi performansı gösterdiği, AOOT sınıflama kriterinin ise OSD performansının yüksek olduğu anlaşılmıştır. Ölçme kesinliği için AOOT sınıflama kriterinin iki kategorili sınıflamada performansının yüksek olduğu belirlenmiştir.

13. Tüm koşullar için MFB madde seçme yönteminin KLB madde seçme yönteminden daha etkili sınıflama yaptığı anlaşılmıştır.

14. Üç kategorili sınıflamada MFB-KN madde seçme yöntemi ile GA sınıflama kriterinin test etkililiği olarak, AOOT sınıflama kriterinin de ölçme kesinliği olarak yüksek performans gösterdiği sonucuna ulaşılmıştır.

15. Dört kategorili sınıflamada GA sınıflama kriterinin KN temelli madde seçme yöntemi ile, GOO sınıflama kriterinin KY temelli madde seçme yöntemiyle test etkililiği olarak AOOT sınıflama kriterinin ölçme kesinliği olarak en iyi performansı gösterdiği sonucuna ulaşılmıştır.

16. Tüm yetenek kestirimleri ile oluşturulan koşullarda artan kategori sayısının test etkililiğini düşürdüğü yani OTU'nu artırırken OSD'nu düşürdüğü görülmüştür. Bununla birlikte RMSE ve OMH değerlerinin ise genel olarak düştüğü daha hassas ölçmenin gerçekleştiği ölçme kesinliğinin arttığı belirlenmiştir.

17. Madde seçme yöntemi ve yetenek kestirim yöntemi fark etmeksizin üç sınıflama kriterinin de bireyleri doğru kategoriye ataması bakımından iyi performans gösterdikleri belirlenmiştir.

18. BSD yetenek kestirim yönteminin test etkililiği ve ölçme kesinliği olarak en iyi performansı gösterdiği, MOK'un OTU ve OSD olarak AOK'dan, AOK'un korelasyon ve hata değerleri başka bir ifade ile ölçme kesinliği olarak MOK'dan daha etkili olduğu sonucuna ulaşılmıştır. Buna göre BBST uygulamasında BSD yetenek kestiriminin diğer yöntemlere göre gücü ve MOK'un AOK'a, AOK'un da MOK'a karşı güçleri ortaya koyulmuştur.

19. BBST için az sayıda madde ile yüksek sınıflama doğruluğu ile sınıflama yapılması amaçlandığından GA ve GOO sınıflama kriterinin gerçek BBST uygulamalarında kullanılmasının uygun olacağı sonucuna ulaşılmıştır. Daha hassas ölçme için daha düşük hata ve yanlılık değerleriyle AOOT sınıflama kriterinin sınıflama yaptığı belirlenmiştir.

20. BBST uygulamalarında sınıflama kategori sayılarına göre fark etse de MFB madde seçme yönteminin KLB madde seçme yöntemine göre daha iyi performans gösterdiği görülmektedir. Buna göre MFB-KY madde seçme yöntemi test etkililiği, MFB-KN temelli madde seçme yöntemi de ölçme kesinliği bakımından tüm koşullar için daha kullanışlı olduğu sonucuna ulaşılmıştır.

21. Tüm kořullarda kategori sayısının artması test uzunluęunu artırdığı sınıflamanın doęruluęunu azalttığı ve ölçmenin hata deęerlerini düşürdüęü daha hasas ölçme yapıldığı sonucunu oluşturmuştur.

Öneriler

Araştırma bulguları ve sonuçlarına göre arařtırmacılara ve uygulayıcılara yön vermesi için önerilerde bulunulmuştur.

Arařtırmacılara Öneriler

1. Arařtırmada yetenek kestirimi olarak BSD, MOK, AOK yetenek kestirimlerinin performansı incelenmiştir. MSD ve BYK gibi yetenek kestirim yöntemlerinin performansı arařtırılabilir.

2. Arařtırmada madde havuzu ikili puanlanan tek boyutlu maddelerden oluşmaktadır. Çok kategorili puanlanan ve çok boyutlu maddelerden oluşan madde havuzlarıyla arařtırma yapılabilir.

3. Arařtırma gerçek veriye dayanan simülasyon çalışmasıdır. Arařtırmanın sonuçları kullanılarak gerçek BBST uygulaması da yapılabilir.

4. Arařtırma 2PLM e göre kalibre edilmiş madde havuzu ile gerçekleştirilmiştir. Madde havuzunda setlerde sınava giren birey sayısı fazla olduğunda yakınsama problemi yaşanmadığında 3PLM ile ve 1PLM ile kalibre edilip MTK modellerine göre de karşılaştırma yapılabilir.

5. Arařtırmada 256 madde ile oluşturulan madde havuzu kullanılmıştır. Farklı büyüklükte madde havuzlarıyla arařtırma tekrarlanıp madde havuzu büyüklüğünün BBST"ne etkisi karşılaştırılabilir.

6. BKK sınıflama kriterinin performansı da yeni arařtırmalarda incelenebilir.

7. Arařtırmada kesme noktası ve kestirilen yetenek temelli madde seçme yöntemi olarak KLB ve MFB madde seçme yöntemleri kullanılmıştır. Global madde seçme yönteminin de iki kategorili sınıflama ile performansı arařtırılabilir.

8. Model veri uyumu sağlanamadığında, tek boyutluluk özelliği sağlanamadığında sınıflama kriterlerinin performansı araştırılabilir.

9. Bu çalışmada içerik dengeleme ve madde kullanım sıklığı kontrolü yöntemlerinin performansı araştırılmamıştır. Gerçek veri ile bu yöntemlerin performansı da araştırılabilir.

10. Bu çalışmada GOO ve AOOT sınıflama kriteri için $FB= 0.10$ GA sınıflama kriteri için %90 güven aralığı belirlenmiştir. 0.20, 0.30 gibi farksızlık bölgesi değerleri ve %95 gibi güven aralığı değerleri için de koşullar belirlenebilir ve yöntemlerin performansları araştırılabilir. FB değeri ve güven değerinin sınıflamaya etkisi araştırılabilir.

Uygulayıcılar İçin Öneriler

1. Test etkililiği için GA ve GOO sınıflama kriterini uygulayıcılara önerilebilir.

2. AOOT sınıflama kriteri ölçme kesinliği yüksek fakat OTU sınıflama yaptığı için uygulayıcılara yüksek riskli testlerde örneğin tıp alanında önerilebilir.

3. Sınıflama kategori sayısı arttıkça ölçme kesinliğinin arttığı görülmektedir. Özellikle Sağlık alanları yüksek risk taşıdığından çok kategorili sınıflama ile daha hasas ölçme yapıldığından uygulayıcılara önerilebilir.

4. MFB madde seçme yöntemi KLB madde seçme yönteminden daha etkili performans gösterdiğinden özellikle ölçme kesinliği için MFB-KN madde seçme yöntemi, test etkililiği için MFB-KY madde seçme yönteminin kullanılması önerilebilir.

5. MOK yetenek kestirimi kullanıldığında iki ve üç kategorili sınıflamalarda GA sınıflama kriterinin özellikle MFB-KN madde seçme yöntemi ile uygulayıcılara önerilebilir.

6. MOK yetenek kestirimi kullanıldığında ölçme kesinliği yüksek sınıflamalar için AOOT sınıflama kriteri MFB-KY madde seçme yöntemi ile kullanılması uygulayıcılara önerilebilir.

7. GOO sınıflama kriteri test etkililiği için MFB-KN madde seçme yöntemi ile birlikte uygulayıcılara önerilse de ölçme kesinliğinin önemli olduğu durumlar için uygulamada tercih edilebilir.

8. MOK yetenek kestirimi kullanıldığında dört kategorili sınıflamada GA sınıflama kriteri KN temelli madde seçme yöntemi ile ölçme kesinliği için uygulayıcılara

önerilmektedir. GOO sınıflama kriterinin dört kategorili sınıflamada ölçme kesinliği için performansı düşük olduğundan uygulamada önerilmemektedir.

9. AOK yetenek kestirimi ile birlikte tüm koşullar için iki kategorili sınıflamada GA sınıflama kriteri ile MFB-KY temelli madde seçme yöntemi ile test etkililiği için önerilmektedir.

10. KN temelli madde seçme yöntemleriyle iki kategorili sınıflamada GOO sınıflama kriteri AOOT sınıflama kriterine göre daha az madde ile sınıflama yaptığı için kullanılması önerilmektedir.

11. Üç kategorili sınıflamada GA sınıflama kriteri tüm madde seçme yöntemleriyle en az sayıda madde ile sınıflama yaptığı için özellikle MFB-KN madde seçme yöntemi ile uygulayıcılara önerilmektedir.

12. MFB madde seçme yöntemi AOK yetenek kestirimi kullanıldığında KLB madde seçme yöntemine göre daha iyi performans gösterdiğinden üç kategorili sınıflamada kullanılması önerilmektedir.

13. Ölçme kesinliği için AOOT sınıflama kriteri üç kategorili sınıflamada MFB-KN temelli madde seçme yöntemleriyle uygulanması önerilmektedir.

14. Dört kategorili sınıflamada AOK yetenek kestirimi kullanıldığında GA sınıflama kriteri MFB (KY-KN) madde seçme yöntemleriyle GOO sınıflama kriteri ise KLB (KY-KN) madde seçme yöntemleriyle test etkililiği için kullanılması önerilmektedir.

15. Ölçme kesinliği için AOOT sınıflama kriteri tüm madde seçme yöntemleriyle dört kategorili sınıflamada kullanımı önerilmektedir.

16. BSD yetenek kestirimi ile yeteneğin kestirdiği iki kategorili sınıflamada GA sınıflama kriteri tüm madde seçme yöntemleriyle OTU ve OSD olarak en iyi performansı gösterdiği için uygulayıcılara kullanımı önerilmektedir. Ölçme kesinliği olarak AOOT sınıflama kriteri özellikle MFB madde seçme yöntemleriyle uygulamada tercih edilebilir.

17. Üç kategorili sınıflamada GA sınıflama kriteri MFB-KN madde seçme yöntemi ile test etkililiği için BBST uygulamasında kullanılması önerilmektedir.

18. AOOT sınıflama kriteri ise ölçme kesinliği olarak üç kategorili sınıflamada MFB-KN madde seçme yöntemiyle birlikte kullanımı önerilmektedir.

19. Dört kategorili sınıflamada GOO sınıflama kriteri KY temelli madde seçme yöntemleriyle, GA sınıflama kriteri ise KN temelli madde seçme yöntemleriyle birlikte kullanımı test etkililiği için önerilmektedir. AOOT sınıflama kriteri ise tüm madde seçme yöntemleriyle dört kategorili sınıflamada ölçme kesinliği için kullanımı önerilmektedir.

20. BSD yetenek kestirimi ile yeteneğin kestirdiği uygulamalarda MFB madde seçme yöntemi için KN temelli madde seçme yöntemi, KLB madde seçme yöntemi için KN temelli OSD için, KY temelli OTU için uygulayıcılara önerilmektedir.

21. BSD yetenek kestirimi tüm koşullar için MOK ve AOK yetenek kestirimlerine göre test etkililiği ve ölçme kesinliği olarak daha iyi performans gösterdiği için uygulamada kullanılabilir.

22. MOK yetenek kestirimi ise AOK yetenek kestiriminden OSD olarak daha yüksek performans gösterdiği için uygulamada önerilmektedir. AOK yetenek kestirimi ise ölçme kesinliği olarak MOK yerine uygulamada tercih edilebilir.

23. GA ve GOO sınıflama kriterleri BBST uygulamalarında OSD ve OTU olarak iyi performans gösterdiklerinden uygulamada kullanımı tercih edilebilir.

24. Tüm koşullar için test etkililiğinin kategori sayısı arttıkça düştüğü ölçme kesinliğinin arttığı görülmektedir. Az kategoride yapılan sınıflamaların avantajlı olduğu düşünülerek uygulayıcılara yüksek riskli testlerde hata oranı düşük sınıflama için çok kategorili sınıflama önerilebilir.

Kaynaklar

- Bartroff, J., Finkelman, M., & Lai, T. L. (2008). Modern sequential analysis and its applications to computerized adaptive testing. *Psychometrika*, 73, [473-486](#).
<http://doi.org/10.1007/S11336-007-9053-9>
- Brown, T.A. (2006). Confirmatory factor analysis for applied research. *New York: The Guilford Press*
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.). *Statistical theories of mental test scores*, 397- 472. *Massachusetts: Addison-Wesley*.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*, 46(4), 443-459
[doi: http://doi.org/10.5926/jjep1953.41.1_22](http://doi.org/10.5926/jjep1953.41.1_22)
- Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö., Karadeniz, Ş. ve Demirel, F. (2008). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Yayınları.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29.
- Davey, T., & Pitoniak, M. J. (2006). *Designing computerized adaptive tests*. *Handbook of Test Development*, 543-574. Routledge.
- De Ayala, R.J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- Demars, C. (2010). *Item response theory*. New York: Oxford University Press, Inc.
- Demir, S. (2019). *Bireyselleştirilmiş Bilgisayarlı Sınıflama Testlerinde Sınıflama Doğruluğunun İncelenmesi*. (Yayımlanmış Doktora Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara. <https://tez.yok.gov.tr/UlusalTezMerkezi/>

- Demir, S. (2022). A Comparison of Computerized Classification Testing Methods over Different Size Item Pools. *International Online Journal of Educational Sciences*, 14(3). <https://doi.org/10.15345/iojes.2022.03.010>
- Diao, H., & Sireci, S. G. (2018). Item response theory based methods for estimating classification accuracy and consistency. *Journal of Applied Testing Technology*, 19(1), 20-25. <https://doi.org/10.1111/j.1745-3984.2009.00096.x>.
- Doğan, N., & Ozturk, N. (2015). Investigating Item Exposure Control Methods in Computerized Adaptive Testing. *Educational Sciences: Theory & Practice*, 1, 85-98.
- Dooley, K. (2002). Simulation research methods. In J. Baum (Ed.). *Companion to organizations*, 829-848. London: Blackwell.
- Eckes, T. (2014). Examining testlet effects in the TestDaF listening section: A testlet response theory modeling approach. *Language Testing*, 31(1), 39-61.
- Eggen, T. J. H. M. (1999). Item selection in adaptive testing with the sequential probability ratio test. *Applied Psychological Measurement*, 23, 249-261. <https://doi.org/10.1177/01466219922031365>
- Eggen, T. J. H. M., & Straetmans, G. J. J. M. (2000). Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological Measurement*, 60(5), 713-734. <https://doi.org/10.1177/00131640021970862>
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates Publishers.
- Eroğlu, M. & Kelecioğlu, H. (2015). Bireyselleştirilmiş bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının ölçme kesinliği ve test uzunluğu açısından karşılaştırılması. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, 28(1), 31-52.

- Ferguson, R. L. (1969). Computer-assisted criterion-referenced measurement (Working PaperNo. 41). *Pittsburgh PA: University of Pittsburgh Learning and Research Development Center.*(ERIC Documentation Reproduction No. ED 037 089).
- Gelbal, S., & Şahin, M. D. (2020). Development of a multidimensional computerized adaptive test based on the bifactor model. *International Journal of Assessment Tools in Education*, 323–342. <https://doi.org/10.21449/ijate.707199>
- Gündeğer, C. (2017). *Bireyselleştirilmiş bilgisayarlı sınıflama testi kriterlerinin sınıflama Doğruluğu ve test uzunluğu açısından karşılaştırılması.*(Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
<https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer Nijhoff Publishing.
- Haring, S. H. (2014). *A comparison of three statistical testing procedures for computerized classification testing with multiple cutscores and item selection methods* (Doctoral dissertation University of Texas at Austin). <http://hdl.handle.net/2152/24838>
- Huang, W. (2004). Stepwise likelihood ratio statistics in sequential studies. *Journal of the Royal Statistical Society*, 66, 401-409.
- Huebner, A. (2012). Item overexposure in computerized classification tests using sequential item selection. *Practical Assessment, Research & Evaluation*, 17(12), 1-9. <https://pareonline.net/getvn.asp?v=17&n=12>
- Huebner, A., & Wang, C., Chen, P. (2020). Stopping rules for multi-category computerized classification testing. *British Journal of Mathematical and Statistical Psychology*, 74(2),184-202.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Scientific software international.

- Jiao, H., & Lau, A. C. (2003, Nisan). The effects of model misfit in computerized classification test. The National Council of Educational Measurement toplantısında sunulan bildiri, Chicago, IL. <http://iacat.org/sites/default/files/biblio/ji03-01.pdf>
- Kaptan, S. (1995). *Bilimsel araştırma teknikleri ve istatistik teknikleri*. Rehber Yayınevi.
- Kaya, E. (2022). *A Comparability and Classification Analysis of Computerized Adaptive and Conventional Paper-Based Versions of an English Language Proficiency Reading Subtest* (Yayımlanmamış Doktora Tezi, Bilkent Üniversitesi), Ankara.
- Kline, T. J. (2005). *Psychological testing: A practical approach to design and evaluation*.
- Kingsbury, G. G., & Weiss, D. J. (1980). A Comparison of Adaptive, Sequential and Conventional Testing Strategies for Mastery Decisions. (Research Report 80-4). University of Minnesota, Minneapolis: MN. <http://iacat.org/sites/default/files/biblio/ki80-04.pdf> adresinden erişilmiştir.
- Kingsbury, G.G., & Weiss, D.J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In D. J. Weiss 113 (Ed.), *New horizons in testing: Latent trait theory and computerized adaptive testing*, 237-254. New York: Academic Pres
- Kim,J., Lee, W.C., Kim,D. & Kelley, K. (2009). Investigation of Vertical Scaling Using the Rasch Model. *National Council on Measurement in Education: April 2009*.
- Kolen, J. M. (2004). Linking Assessments: Concept and History. *Applied Psychological Measurement*, 28(4), 219- 226.
- Kolen, M. J., & Brennan, R. L. (2004). Test equating, scaling, and linking: Methods and practices (2nd ed.). *New York: Springer-Verlag*.
- Lau, C. A. (1996). *Robustness of a unidimensional computerized testing mastery procedure with multidimensional testing data*. (Doctoral Dissertation, University of Iowa).
- Lau, C. A., & Wang, T. (1999, Nisan). Computerized classification testing under practical

constraints with a polytomous model. American Educational Research Association (AERA) toplantısında sunulan bildiri. Montreal, Canada. <http://iacat.org/sites/default/files/biblio/la99-01.pdf>

Lau, C. A., & Wang, T. (1999, Nisan). *Computerized classification testing under practical constraints with a polytomous model*. American Educational Research Association (AERA) toplantısında sunulan bildiri. Montreal, Canada. <http://iacat.org/sites/default/files/biblio/la99-01.pdf> adresinden erişilmiştir

Lathrop, Q. N. (2015). Practical issues in estimating classification accuracy and consistency with R package caclRT. *Practical Assessment, Research & Evaluation*, 20(18), 1-5. <https://doi.org/10.7275/43vm-p442>

Lewis, C., & Sheehan, K. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement*, 14, 367- 386. <https://doi.org/10.1177/014662169001400404>

Liu, T. H., Chen, C. T., Cheng, C. P., & Shih, C. L. (2022). Improving test efficiency for a grid multidimensional computerized classification test by the application of a conditional latent-trait distribution to a sequential probability ratio test. *Behavior Research Methods*, 1-14.

Lin, C. J., & Spray, J. (2000). *Effects of Item-Selection Criteria on Classification Testing with the Sequential Probability Ratio Test*. ACT (Research Report 2000-8). Iowa city, IA: ACT Research Report Series. <https://eric.ed.gov/?id=ED445066>

Luecht, R. M. (1998). Computer-assisted test assembly using optimization heuristics. *Applied Psychological Measurement*, 22(3), 224-236.

Lord, F. M. (1952). Notes on a problem of multiple classification. *Psychometrika*, 17(3), 297-304.

- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates Publishers
- Lord, F. M. (1986). Maximum likelihood and Bayesian parameter estimation in item response theory. *Journal of Educational Measurement*, 157-162.
- Meng, H (2007). *A comparison study of IRT calibration methods for mixed-format tests in vertical scaling*. (Doctoral Dissertation, University of Iowa).
- Miller, I. & Miller, M. (2004). *John E. Freund's Mathematical Statistics with Applications*. (7th Edition). New Jersey: Prentice Hall.F
- Nydick, S. & Weiss, D. J. (2009). *A hybrid simulation procedure for developments of CATs*. Paper presented at the Item Pool Development Paper session at the 2009 GMACConference on Computerized Adaptive Testing.
- Nydick, S. W., Nozawa, Y., & Zhu, R. (2012, Nisan). *Accuracy and efficiency in classifying examinees using computerized adaptive tests: An application to a large scale test*. The National Council on Measurement in Education (NCME) toplansında sunulan bildiri, Vancouver, British Columbia, Canada.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.476.3381&rep=rep1&type=pdf>
- Nydick, S. W. (2013). *Multidimensional mastery testing with CAT*. (Doctoral Dissertation, theUniversity of Minnesota). Available from ProQuest Dissertations and Theses database. (UMI No. 3607925)
- Nydick, S. W. (2014). *catirt: An R Package for Simulating IRT-Based Computerized Adaptive Tests*. <https://cran.rproject.org/web/packages/catlrt/catlrt.pdf>
- Nurakhmetov, D. (2019). Reinforcement learning applied to adaptive classification testing. *Theoretical and Practical Advances in Computer-based Educational Measurement*, 325-336.

- Owen, R. J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association*, 70, 351-356. <https://doi.org/10.2307/2285821>
- Parshall, C. G., Spray, J. A., Kalohn, J. C., Davey, T., Parshall, C. G., Spray, J. A., ... & Davey, T. (2002). *Considerations in Computer-Based Testing* (pp. 1-12). Springer New York.
- Patsula, L. N., & Steffen, M. (1997). Maintaining Item and Test Security in a CAT Environment: A Simulation Study. *Laboratory of Psychometric and Evaluative Research Report No. 309*.
- Paek, I., & Cole, K. (2019). *Using R for item response theory model applications*. Routledge.
- R Core Team (2013). *R: A language and environment for statistical computing*, (Version 3.0.1), Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/> adresinden erişilmiştir.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. University of Chicago Press
- Reckase, M. D. (1983). A procedure for decision making using tailored testing. In D. J. Weiss (Ed.). *New horizons in testing: latent trait theory and computerized adaptive testing*, 237-254. New York: Academic Press.
- R Core Team (2013). R: A language and environment for statistical computing, (Version 3.0.1) [Computer software], Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.Rproject.org/>
- Revelle, W. (2020). How to: Use the psych package for factor analysis and data reduction. *Northwestern University, Department of Psychology: Evanston, IL, USA*.
- Rudner, L. M. (2002, Nisan). *An examination of decision-theory adaptive testing*

procedures. American Educational Research Association (AERA) toplantısında sunulan bildiri, New Orleans, LA. <http://iacat.org/sites/default/files/biblio/ru02-01.pdf> adresinden erişilmiştir.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores.

Psychometrika, Monograph No. 17.

Spray, J. A. (1993). Multiple-category classification using a sequential probability ratio test

(Report No. ACT-RR-93-7). *American College Testing*

Spray J. A., & Reckase M. D. (1994, April 5-7). *The selection of test items for decision*

Making with a computer adaptive test [Paper presentation]. National Meeting of the National Council on Measurement in Education, New Orleans, LA.

<https://eric.ed.gov/?id=ED372078>

Spray, J. A., & Reckase, M. D. (1996). Comparison of SPRT and sequential bayes

procedures for classifying examinees into two categories using a computerized test.

Journal of Educational and Behavioral Statistics, 21(4), 405-414.

<https://doi.org/10.3102/10769986021004405>

Spray, J. A., Abdel-Fattah, A., Huang, C., & Lau, C. A. (1997). *Unidimensional*

Approximations for a Computerized Classification Test When the Item Pool and

Latent Space Are Multidimensional. (Research Report 97-5). Iowa city, IA: ACT

Research Report Series. <https://eric.ed.gov/?id=ED414298>.

Sümer, N. (2000). Yapısal Eşitlik Modelleri: Temel Kavramlar ve Örnek Uygulamalar. *Türk*

Psikoloji Yazilari.

Stocking, M. L. (1994). Three practical issues for modern adaptive testing item pools 1. *ETS*

Research Report Series, 1994(1), i-34.

Tabachnick, B. G., & Fidel, L. S. (2001). *Using multivariate statistics*. MA: Allyn ve Bacon.

- Thompson, N. A. (2009). Item selection in computerized classification testing. *Educational and Psychological Measurement*, 69(5), 778-793.
<https://doi.org/10.1177/0013164408324460>
- Thompson, N. A., & Ro, S. (2007). Computerized classification testing with composite hypotheses. In D. J. Weiss (Ed.). *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*. http://www.iacat.org/sites/default/files/biblio/cat07nt_hompson.pdf.
- Thompson, N. A., & Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research, and Evaluation*, 16(1).
<https://doi.org/10.7275/wqzt-9427>
- Thompson, N. A. (2011). Termination criteria for computerized classification testing. *Practical Assessment, Research & Evaluation*, 16(4), 1-7.
<https://doi.org/10.7275/wq8m-zk25>
- Thompson, N. A. (2007). A practitioner's guide for variable-length computerized classification testing. *Practical Assessment Research & Evaluation*, 12(1), 1-13.
<http://www.iacat.org/sites/default/files/biblio/th07-01.pdf>.
- van Groen, M. M., Eggen, T. J., & Veldkamp, B. P. (2016). Multidimensional computerized adaptive testing for classifying examinees with within-dimensionality. *Applied Psychological Measurement*, 40(6), 387-404.
- Wald, A. (1947). *Sequential analysis*. New York: John Wiley.
<https://psycnet.apa.org/record/1947-04217-000>
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427-450. <https://doi.org/10.1007/BF02294627>
- Wang, T. & Vispoel, W. P. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(2), 109-135.

- Wang, T., Hanson, B. A. & Lau, C. A. (1999). Reducing bias in CAT trait estimation: a comparison of approaches. *Applied Psychological Measurement*, 23(3), 263-278.
- Wang, S., & Wang, T. (2001). Precision of warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement*, 25(4), 317–331. <https://doi.org/10.1177/01466210122032163>
- Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational measurement*, 24(3), 185-201.
- Wainer, H., & Thissen, D. (1987). Estimating ability with the wrong model. *Journal of Educational Statistics*, 12(4), 339-368.
- Wainer, H. & Mislevy, R. J. (2000). Item response theory, item calibration and proficiency estimation. In Wainer, H. (Ed.). *Computerized adaptive testing: a primer*, 61-100. New Jersey: Lawrence Erlbaum Associates Publishers.
- Way, W. D. (2006). *Practical questions in introducing computerized adaptive testing to K-12 assessments*. Retrieved from http://www.pearsonassessments.com/NR/rdonlyres/EC965AB8-EE70-46E5B1A5036BE41AB899/0/RR_05_03.Pdf?WT.mc_id=TMRS_Practical_Question_in_Introducing_Computerized.
- Weiss, D. J. (Ed.). (2014). *New horizon testing: Latent trait test theory and computerized adaptive testing*. Elsevier.
- Weiss, D. J., & Kingsbury, G. 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>.
- Weissman, A. (2006). A Feedback Control Strategy for Enhancing Item Selection Efficiency

in Computerized Adaptive Testing. *Applied Psychological Measurement*, 30, 84-99.

<https://doi.org/10.1177/0146621605282774>

Yang, X., Poggio, J. C., & Glasnapp, D. R. (2006). Effects of estimation bias on multiple category classification with an IRT-based adaptive classification procedure.

Educational and Psychological Measurement, 66(4), 545-564.

<https://doi.org/10.1177/0013164405284031>

Yoes, M. (1995). An updated comparison of micro-computer based item parameter

estimation procedures used with the 3-parameter IRT model. *Saint Paul, MN:*

Assessment Systems Corporation

Yi, Q., Wang, T. & Ban, J. (2000). Effects of Scale Transformation and Test Termination

Rule on the Precision of Ability Estimates in CAT. ACT Research Report Series,

2000-2. <https://doi/10.1111/j.1745-3984.2001>.

EK-A: Etik Komisyonu Onay Bildirimi

Evrak Tarih ve Sayısı: 05.01.2021-566



T.C.
KAHRAMANMARAŞ ST İMAM NİVERSİTESİ
REKTRLĖ
Sosyal ve Beşeri Bilimler Etik Kurulu

Sayı : E-72321963-605.01-566
Konu : Veri Toplama

EĖİTİM BİLİMLERİ BLM BAŞKANLIĖINA

İlgi : 29.12.2020 tarihli ve 11930 sayılı yazı,

Kurulumuzdan istemiş olduğunuz etik kurul kararı ekte gönderilmiştir.
Gereğini bilgilerinize arz/rica ederim

Prof. Dr. Nusret GKSU
Kurul Başkanı

Ek: Sosyal ve Beşeri Bilimler Etik Kurulu 04.01.2021 Tarih ve 2021-1 Sayılı Kararı (2 sayfa)

T.C.

KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ
SOSYAL VE BEŞERİ BİLİMLER ETİK KURUL KARARLARI

Toplantı Tarihi: 04.01.2021

Toplantı Sayısı: 2020/1

Üniversitemiz Sosyal ve Beşeri Bilimleri Etik Kurulu, Kurul Başkanı Prof. Dr. Nusret GÖKSU'nun Başkanlığında toplanarak aşağıdaki kararları almıştır,

KARAR 1 :

Dr. Öğr. Üyesi Levent YAKAR'ın 29.12.2020 Tarihli dilekçesi ve ekleri Prof. Dr. Nail YILDIRIM (raportör) 31.12.2020 tarih ve E. 49759 sayılı raporu görüşüldü.

Üniversitemiz Öğretim Üyelerinden Dr. Öğr. Üyesi Levent YAKAR'ın "İngilizce Hazırlık Sınıflarında Bilgisayarda Bireyselleştirilmiş Test Kullanımı" başlıklı çalışması kapsamında kullanmak için uygulanacak açık uçlu görüşme formu (ankette yer alan soru, önerme ve/veya ölçeklerin alıntılanmasına yönelik akademik alıntılanma kurallarına uyma zorunluluğu araştırmayı yapan akademisyen ve varsa danışmanlığını yaptığı öğrenciye ait olmak koşuluyla) üniversite öğrencilerine uygulama talebinin uygun olduğuna oylama sonucunda oy birliği ile;

Karar verildi

BAŞKAN (İMZA)
Prof. Dr. Nusret GÖKSU

ÜYE (İMZA) Prof. Dr. Salih YEŞİL	ÜYE (İMZA) Prof. Dr. Ahmet NALÇACI	ÜYE (İMZA) Prof. Dr. Nail YILDIRIM
ÜYE (İMZA) Prof. Dr. Kemal TİMUR	ÜYE (İMZA) Prof. Dr. Faruk ÇİFTÇİ	

EK-B:Veri Kullanım İzin Belgesi

02.08.2022

İlgili Makama

Yürütücüsü olduğum “İngilizce Hazırlık Sınıflarında Bilgisayarda Bireyselleştirilmiş Test Kullanımı” başlıklı Tübitak projesi kapsamında Eylül 2022 itibari ile veri toplanmasına başlanacaktır. Projede bursiyer olarak görevli Hacettepe Üniversitesi Eğitimde Ölçme Değerlendirme Bölümünde doktora öğrencisi Demet ALKAN proje kapsamında topladığı gerçek verileri doktora tezinde kullanabilir. Bilgilerinize saygılarımla arz ederim.

Kahramanmaraş Sütçü İmam Üniversitesi

Doç. Dr. Levent YAKAR

EK-C:Etik komisyon İzni



T.C.
HACETTEPE ÜNİVERSİTESİ REKTÖRLÜĞÜ
Eğitim Bilimleri Enstitüsü Müdürlüğü

Tarih: 09/06/2023 15:04
Sayı: E-51944218-300-00002892897



00002892897

Sayı : E-51944218-300-00002892897
Konu : Etik Komisyonu İzni (Demet ALKAN)

9.06.2023

EĞİTİM BİLİMLERİ ANABİLİM DALI BAŞKANLIĞINA

İlgi : 05.06.2023 tarihli ve E-52695194-300-00002883193 sayılı yazınız.

Ana Bilim Dalınız Eğitimde Ölçme ve Değerlendirme programı öğretim üyesi Prof. Dr. Nuri DOĞAN'ın danışmanlığını yürüttüğü öğrencilerden Demet ALKAN'ın, Kahramanmaraş Sütçü İmam Üniversitesi Dr. Öğr. Üyesi Levent YAKAR'ın "İngilizce Hazırlık Sınıflarında Bilgisayarda Bireyselleştirilmiş Test Kullanımı" başlıklı çalışma kapsamında alınan etik komisyonu izninin adı geçen öğrencinin tez çalışmasında geçerli sayılması isteği, çalışma için gerekli izinlerin alınması ve izinlerle ilgili belgelerin öğrencilerin tezlerinde bulunması koşuluyla uygun bulunmuştur. Bilgilerinizi ve gereğini rica ederim.

Prof. Dr. Selahattin GELBAL
Enstitü Müdürü

Bu belge güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu: 47BE7B7A-8C20-4DE3-8784-36D96521B32E
Adres: Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğü 06800
Beytepe-ANKARA
E-posta: ebe@hacettepe.edu.tr Elektronik Ağ: www.hacettepe.edu.tr
Telefon: (0 312) 297 85 70-71 Faks:(0 312) 299 85 66
Kep:

Belge Doğrulama Adresi: <https://www.turkiye.gov.tr/hu-ebys>

Bilgi için: Aysun ALTUN

Bilgisayar İşletmeni

Telefon: -



EK-Ç: 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.

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

Demet ALKAN

EK-D: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu

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

HACETTEPE ÜNİVERSİTESİ

Eğitim Bilimleri Enstitüsü

Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı : İngilizce Seviye Belirleme Sınav Sonuçları Üzerinde Bilgisayarda
Bireyselleştirilmiş Sınıflama Testi Yaklaşımının Uygulanması

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
01/12/2023.	113	184058	20/12 /2023	%9	2244098216

Uygulanan filtreler:

1. Kaynaklar hariç
2. Alıntılar dâhil
3. 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: **Demet ALKAN**

Öğrenci No.: **N18248213**

Ana Bilim Dalı: **Eğitim Bilimleri**

Programı: **Eğitimde Ölçme ve Değerlendirme**

Statüsü: **Doktora**

İmza

DANIŞMAN ONAYI

UYGUNDUR.
Prof.Dr. Nuri DOĞAN

İmza

EK-E: Thesis/Dissertation Originality Report

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

HACETTEPE UNIVERSITY

Graduate School of Educational Sciences

To The Department of Educational Sciences.

Thesis Title: Application Of Computerized Adaptive Classification Test Approach On
English Placement Test Results

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
01/12 /2023.	113	184058	20/12 /2023	%9	2244098216

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: Demet ALKAN
Student No.: N18248213
Department: Educational Sciences
Program: Measurement and Evaluation in Education
Status: Ph.D.

Signature

ADVISOR APPROVAL

APPROVED
Prof.Dr. Nuri Doğan

Signature

EK-F: Yayınlama 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 fikrî mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin 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 İlişkin Yönerge**" kapsamında tezim 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 tarihimden itibaren ... ay ertelenmiştir. ⁽²⁾
- Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

..... / /

(imza)

Demet ALKAN

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin 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 tezinerişime 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 tezinerişime 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 tezlere 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

