



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

BİREYSELLEŞTİRİLMİŞ BİLGİSAYARLI SINIFLAMA TESTLERİNDE
SINIFLAMA DOĞRULUĞUNUN İNCELENMESİ

Seda DEMİR

Doktora Tezi

Ankara, 2019

Liderlik, arařtırma, inovasyon, kaliteli eęitim ve deęiřim ile

Daha ileriye ... En İyiyeye ...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

BİREYSELLEŞTİRİLMİŞ BİLGİSAYARLI SINIFLAMA TESTLERİNDE
SINIFLAMA DOĞRULUĞUNUN İNCELENMESİ

INVESTIGATION OF CLASSIFICATION ACCURACY AT COMPUTERIZED
ADAPTIVE CLASSIFICATION TESTS

Seda DEMİR

Doktora Tezi

Ankara, 2019

Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Seda DEMİR'in hazırladıđı "Bireyselleřtirilmiř Bilgisayarlı Sınıflama Testlerinde Sınıflama Dođruluđunun İncelenmesi" bařlıklı bu alıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiřtir.

J¼ri Bařkanı

Prof. Dr. Selahattin GELBAL

İmza

J¼ri Üyesi (Danıřman)

Do. Dr. Burcu ATAR

İmza

J¼ri Üyesi

Prof. Dr. Nuri DOĐAN

İmza

J¼ri Üyesi

Prof. Dr. řeref TAN

İmza

J¼ri Üyesi

Do. Dr. Dilara BAKAN
KALAYCIOĐLU

İmza

Bu tez Hacettepe Üniversitesi Lisans¼st¼ Eđitim, Öğretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından 24 / 10 / 2019 tarihinde uygun gör¼lm¼ř ve Enstit¼ Yönetim Kurulunca / / tarihinde kabul edilmiřtir.

Prof. Dr. Ali Ekber řAHİN
Eđitim Bilimleri Enstitüsü M¼d¼r¼

Öz

Bu arařtırmada, Bireyselleřtirilmiř Bilgisayarlı Sınıflama Testlerinde (BBST) iki kategorili puanlanan maddelerden oluřan tek boyutlu madde havuzu üzerinden sınıflamanın iki, üç veya dört kategoride yapıldığı durumlarda, yetenek kestirim yöntemi Ağırlıklandırılmıř Olabilirlik Kestirimi (AOK) olduėunda, ortalama test uzunluėu (OTU), ortalama sınıflama doėruluėu (OSD) ve ölçme kesinliėinin; Ardıřık Olasılık Oran Testi (AOOT) ve Güven Aralıėı (GA) sınıflama kriterleri ile Kestirilen Yetenek Temelinde Maksimum Fisher Bilgisi (MFB-KY) ve Kesim Noktası Temelinde Maksimum Fisher Bilgisi (MFB-KN) madde seėme yöntemlerine göre içerik dengeleme ve madde kullanım sıklığı kontrolü gibi pratik kısıtlamalar altında nasıl deėiřtiėinin incelenmesi amaėlanmıřtır. Arařtırmada içerik dengeleme yöntemlerinden Kısıtlanmıř Bireyselleřtirilmiř Bilgisayarlı Test (KBBT) ve Modifiye Edilmiř Multinomial Model (MMM), madde kullanım sıklığı kontrol yöntemlerinden ise Sympson-Hetter Yöntemi (SH) ve Madde Uygunluk Yöntemi (MU) incelenmiřtir. Arařtırma kapsamında, R yazılımı kullanılarak türetilen 500 madde ve 5000 bireyden oluřan veriye dayanan Monte Carlo (MC) simülasyonu için toplam 48 kořul oluřturulmuř ve elde edilen deėerler her bir kořul için 30 tekrarın ortalaması alınarak hesaplanmıřtır. alıřma sonucunda, OTU bakımından GA'nın, OSD bakımından ise AOOT'nin daha bařarılı performans gösterdiėi; MFB-KY madde seėme yönteminin MFB-KN'den daha kullanıřlı olduėu; korelasyon, yanlılık, RMSE ve Ortalama Mutlak Hata (OMH) deėerleri bakımından AOOT'nin daha etkili olduėu görülmüřtür. Bununla birlikte, istenen içerik dengesini saėlamada MMM yöntemi daha bařarılıyken, test etkililiėi (OTU ve OSD) bakımından KBBT'nin daha avantajlı olduėu; madde kullanım sıklığı kontrolünde ise MU yöntemi daha iyi performans gösterirken, test etkililiėinde SH'nin daha avantajlı olduėu belirlenmiřtir. Artan kategori sayısının ise OTU'yu artırırken OSD'yi azalttıėı; korelasyon, yanlılık, RMSE ve OMH deėerlerinde iyileřme saėladıėı görülmüřtür.

Anahtar sözcükler: bireyselleřtirilmiř bilgisayarlı sınıflama testi, içerik dengeleme, madde kullanım sıklığı kontrolü, sınıflama kriterleri, madde seėme yöntemleri, sınıflama kategori sayısı

Abstract

The aim of this study is to compare Sequential Probability Ratio Test (SPRT) and Confidence Interval (CI) classification criteria, Maximum Fisher Information method on the basis of estimated-ability (MFI-EB) and cut-point (MFI-CB) item selection methods while ability estimation method is Weighted Likelihood Estimation (WLE) in Computerized Adaptive Classification Testing (CACT), according to the average test length (ATL), average classification accuracy (ACA), and measurement precision under content balancing (Constrained Computerized Adaptive Testing: CCAT and Modified Multinomial Model: MMM) and item exposure control (Sympton-Hetter Method: SH and Item Eligibility Method: IE) when the classification is done based on two, three, or four categories for unidimensional pool of dichotomous items. 48 conditions are created in Monte Carlo (MC) simulation for the data, generated in R software, including 500 items and 5000 examinees, and the results are calculated over 30 replications. As a result of the study, it was observed that CI performs better in terms of ATL and SPRT performs better in ACA and correlation, bias, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) values, respectively; MFI-EB is more useful than MFI-KN. It was also seen that MMM is more successful in content balancing whereas CCAT is better in terms of test efficiency (ATL and ACA) and IE is superior in terms of item exposure control though SH is more beneficial in test efficiency. Besides, increasing the number of classification categories increases ATL but decreases ACA and it gives better results in terms of the correlation, bias, RMSE, and MAE values.

Keywords: computerized adaptive classification testing, content balancing, item exposure control, classification criteria, item selection methods, number of classification categories

Teşekkür

Doktora eğitimim ve tez çalışmamın her aşamasında yanımda olup heyecanımı paylaşan, güler yüzlü rehberliği ve değerli görüşleriyle yoluma ışık tutan danışmanım Doç. Dr. Burcu ATAR'a çalışmama sağladığı katkılardan dolayı sonsuz teşekkürlerimi sunarım.

Doktora eğitimim ve tez çalışmam süresince, karşılaştığım bütün zorluklar karşısında, öğrendiğim her yeni bilginin ruhuma kattığı mutluluk, kararlılık ve azimle ilerlediğim bu yolda gelişimime büyük katkılar sağlayan değerli hocalarım Prof. Dr. Selahattin GELBAL, Prof. Dr. Hülya KELECİOĞLU ve Prof. Dr. Nuri DOĞAN başta olmak üzere bugünlere gelmemde katkısı olan tüm hocalarıma teşekkürlerimi sunarım.

Tez çalışmam süresince özellikle yazdığım kodlarla ilgili karşılaştığım sorunlarda, verdiği fikirlerle çözüme ulaşmamda ilham kaynağı olan Prof. Dr. Alan Huebner'a gösterdiği samimi ve içten tavırlarından dolayı sonsuz teşekkür ederim.

Tez izleme komitemde ve savunma jürimde yer alarak kıymetli fikir ve önerileriyle çalışmamı zenginleştirmeme katkı sağlayan Prof. Dr. Şeref TAN ve savunma jürimde bulunan Dilara BAKAN KALAYCIOĞLU'na çalışmamla ilgili yapıcı önerileri ve katkıları için sonsuz teşekkür ederim.

Bugünlere gelmemde sayısız emekleri olan, desteklerini ve sabırlarını benden hiçbir zaman eksik etmeyen, tez çalışmamın her aşamasında benimle birlikte çaba, emek ve titizlik gösteren, evlatları olmaktan gurur duyduğum canım annem Neriman DEMİR'e ve canım babam Mehmet DEMİR'e bu zorlu, heyecan ve emek dolu yolda benimle birlikte yürüdükleri için gönülden teşekkür ederim. Ayrıca bana verdikleri destek ve mükemmel kardeşlikleriyle motivasyonumu her zaman yükselten ablam Eda DEMİR ve kardeşim Ali Aytuğ DEMİR'e çok teşekkür ederim.

Son olarak tez çalışmam süresince manevi desteklerini her zaman yanımda hissettiğim kıymetli dostlarıma teşekkür ederim.

Böyle bir tez çalışmasının ortaya çıkışını mümkün kılan tüm bu isimlere kalbimin derinliklerinden gelen sonsuz teşekkürlerimi sunarım.

İçindekiler

Öz.....	ii
Abstract.....	iii
Teşekkür.....	iv
Tablolar Dizini.....	vii
Şekiller Dizini.....	viii
Simgeler ve Kısaltmalar Dizini.....	ix
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	8
Araştırma Problemi.....	10
Sayıltılar.....	11
Sınırlılıklar.....	12
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	13
Bireyselleştirilmiş Bilgisayarlı Sınıflama Testi (BBST).....	13
Pratik Kısıtlamalar.....	35
İçerik Dengeleme.....	35
Madde Kullanım Sıklığı.....	39
İlgili Araştırmalar.....	44
Bölüm 3 Yöntem.....	58
Araştırma Yöntemi.....	58
Verilerin Türetilmesi.....	58
BBST Simülasyon Koşulları.....	63
Verilerin Analizi.....	64
Bölüm 4 Bulgular ve Yorumlar.....	66
Alt Problem 1'e İlişkin Bulgular ve Yorumlar.....	66
Alt Problem 2'ye İlişkin Bulgular ve Yorumlar.....	91

Alt Problem 3'e İlişkin Bulgular ve Yorumlar	93
Alt Problem 4'e İlişkin Bulgular ve Yorumlar	95
Alt Problem 5'e İlişkin Bulgular ve Yorumlar	97
Alt Problem 6'ya İlişkin Bulgular ve Yorumlar.....	100
Bölüm 5 Sonuç, Tartışma ve Öneriler	103
Sonuçlar.....	103
Öneriler	105
Kaynaklar	109
EK-A: Türetilen Madde Parametrelerinden Rastgele Seçilmiş Bir Tanesine Ait Betimsel İstatistikler.....	120
EK-B: Türetilen Yetenek Parametrelerinden Rastgele Seçilmiş Bir Tanesine Ait Betimsel İstatistikler.....	121
EK-C: Etik Komisyonu Onay Bildirimi	122
EK-Ç: Etik Beyanı.....	123
EK-D: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu	124
EK-E: Thesis/Dissertation Originality Report.....	125
EK-F: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	126

Tablolar Dizini

Tablo 1 <i>Araştırma Kapsamında İncelenen Değişkenler</i>	8
Tablo 2 <i>BBST Bileşenleri</i>	14
Tablo 3 <i>MC Veri Setine Ait Açıklanan Varyans ve Özdeğerler</i>	61
Tablo 4 <i>Yetenek Kestirim Yöntemi AOK Olduğunda Alt Problem 1.1. Koşullarına Ait OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri</i>	67
Tablo 5 <i>Yetenek Kestirim Yöntemi AOK Olduğunda Alt Problem 1.2. Koşullarına Ait Uygulanan İçerik Oranı, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri</i>	72
Tablo 6 <i>Yetenek Kestirim Yöntemi AOK Olduğunda Alt Problem 1.3. Koşullarına Ait MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri</i>	78
Tablo 7 <i>Yetenek Kestirim Yöntemi AOK Olduğunda Alt Problem 1.4. Koşullarına Ait Uygulanan İçerik Oranı, MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri</i>	84
Tablo 8 <i>AOOT ve GA Sınıflama Kriterlerinin Uygulanan İçerik Oranı, MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri</i>	92
Tablo 9 <i>MFB-KY ve MFB-KN Madde Seçme Yöntemlerinin Uygulanan İçerik Oranı, MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri</i> ...	94
Tablo 10 <i>KBBT ve MMM İçerik Dengeleme Yöntemlerinin Uygulanan İçerik Oranı, MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri</i>	96
Tablo 11 <i>SH ve MU Madde Kullanım Sıklığı Kontrol Yöntemlerinin Uygulanan İçerik Oranı, MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri</i>	98
Tablo 12 <i>İki, Üç ve Dört Kategoride Yapılan Sınıflamaların Uygulanan İçerik Oranı, MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri</i>	101

Şekiller Dizini

Şekil 1. İki kategoride yapılacak sınıflamalar için sınıflama kararı.....	28
Şekil 2. Üç kategoride yapılacak sınıflamalar için sınıflama kararı.....	29
Şekil 3. Dört kategoride yapılacak sınıflamalar için sınıflama kararı.....	31
Şekil 4. Test bilgi fonksiyonu.....	59
Şekil 5. Yetenek parametrelerinin dağılım grafiği.....	60
Şekil 6. Yamaç-birikinti grafiği.....	62
Şekil 7. Sınıflama kriterleri ve madde seçme yöntemlerinin çaprazlandığı OTU değişim grafiği.....	68
Şekil 8. Sınıflama kriterleri ve madde seçme yöntemlerinin çaprazlandığı OSD değişim grafiği.....	68
Şekil 9. Sınıflama kriterleri, madde seçme yöntemleri ve içerik dengeleme yöntemlerinin çaprazlandığı OTU değişim grafiği.....	74
Şekil 10. Sınıflama kriterleri, madde seçme yöntemleri ve içerik dengeleme yöntemlerinin çaprazlandığı OSD değişim grafiği.....	74
Şekil 11. Sınıflama kriterleri, madde seçme yöntemleri ve madde kullanım sıklığı kontrol yöntemlerinin çaprazlandığı OTU değişim grafiği.....	80
Şekil 12. Sınıflama kriterleri, madde seçme yöntemleri ve madde kullanım sıklığı kontrol yöntemlerinin çaprazlandığı OSD değişim grafiği.....	81
Şekil 13. Sınıflama kriterleri, madde seçme yöntemleri, içerik dengeleme yöntemleri ve madde kullanım sıklığı kontrol yöntemlerinin çaprazlandığı OTU değişim grafiği.....	87
Şekil 14. Sınıflama kriterleri, madde seçme yöntemleri, içerik dengeleme yöntemleri ve madde kullanım sıklığı kontrol yöntemlerinin çaprazlandığı OSD değişim grafiği.....	88

Simgeler ve Kısaltmalar Dizini

AOK: Ağırlıklandırılmış Olabilirlik Kestirimi

AOOT: Ardışık Olasılık Oran Testi

BBST: Bireyselleştirilmiş Bilgisayarlı Sınıflama Testi

BBT: Bireyselleştirilmiş Bilgisayarlı Test

FB: Farksızlık Bölgesi

GA: Güven Aralığı

KBBT: Kısıtlanmış Bireyselleştirilmiş Bilgisayarlı Test (İçerik Dengeleme Yöntemi)

MFB-KY: Kestirilen Yetenek Temelinde Maksimum Fisher Bilgisi

MFB-KN: Kesim Noktası Temelinde Maksimum Fisher Bilgisi

MKSAMY: Maksimum Madde Kullanım Sıklığı Oranını Aşan Maddelerin Yüzdesi

MKSAMO: Maksimum Madde Kullanım Sıklığı Oranını Aşan Maddelerin Madde Kullanım Sıklıkları Ortalaması

MMM: Modifiye Edilmiş Multinomial Model (İçerik Dengeleme Yöntemi)

MTK: Madde Tepki Kuramı

MU: Madde Uygunluk Yöntemi (Madde Kullanım Sıklığı Kontrol Yöntemi)

OMH: Ortalama Mutlak Hata

OSD: Ortalama Sınıflama Doğruluğu

OTU: Ortalama Test Uzunluğu

SH: Sympson-Hetter Yöntemi (Madde Kullanım Sıklığı Kontrol Yöntemi)

Bölüm 1

Giriş

Bu bölümde araştırmanın problem durumuna, amacına ve önemine, problem cümlesine, alt problemlerine, sayıltılarına ve sınırlılıklarına yer verilmiştir.

Problem Durumu

Günümüz toplumlarında giderek yaygınlaştığı görülen testlerin, kendini gösterdiği öncelikli alanlardan birisi de eğitimidir. Eğitimde test etmenin birçok farklı amacı olabilir. Bu amaçlar arasında eğitimin etkililiğini arttırma, öğrencileri bireysel olarak değerlendirme, seçme veya yerleştirme kararları alma, belgelendirme, öğrenmedeki ilerlemeyi izleme ve tanılayıcı amaçlı test etme sıralanabilir. Belirtilen bu amaçların gerçekleştirilebilmesinde, öğrencinin sahip olduğu yetenek düzeyi hakkında zamanında ve doğru bilgiye erişebilmenin kritik bir öneme sahip olduğu söylenebilir. Bu noktada, bilgi ve iletişim teknolojilerindeki gelişmelerin eğitim alanındaki en büyük yansımalarından biri olan Bireyselleştirilmiş Bilgisayarlı Testler (BBT; Computerized Adaptive Testing: CAT), giderek daha nitelikli ve etkili değerlendirmeler yapılmasına katkıda bulunmaktadır.

BBT'nin en büyük avantajı, test uzunluğunu ve dolayısıyla teste ayrılan süreyi azaltılıp ölçüm hassasiyetini artırarak daha etkili testler geliştirilebilmesine olanak sağlamasıdır (Thompson, 2009). Başka bir ifadeyle BBT yaklaşımı, test etkililiğini önemli ölçüde artırırken, testin geçerliğini ve güvenilirliğini de korumaktadır (Thompson, 2007b). Bunun yanı sıra bireyin son yetenek düzeyinin test sonlanır sonlanmaz kestirilebilmesi, bireyin kaydedilebilen bilgileri sayesinde test sonuçlarına ilave pek çok analiz yapılabilmesi ve testin istenilen zamanda uygulanabilmesi BBT'nin diğer avantajları arasında sıralanabilir. Bu avantajlarının yanında BBT'nin bazı dezavantajlarından da bahsetmek mümkündür. Genel olarak MTK temelinde gerçekleştirilen BBT'ler açık uçlu maddeler gibi bazı madde tipleri için uygun değildir ve ölçülen özelliğin tek boyutlu olma şartını sağlaması oldukça zordur. Bu nedenle her konu ya da ölçülecek özellik için uygulanabilir değildir. BBT uygulamalarında, her güçlük düzeyinde tekrarlanabilecek ölçümler için çok sayıda madde gerekmektedir. Dolayısıyla, büyük bir madde havuzuna ihtiyaç duyulmaktadır. Bunların yanı sıra, maddelere yanıt vermeden (boş bırakarak) geçmenin mümkün olmaması, sonraki maddeler önceki yanıtlara göre

seçildiğinden bireyin geri dönüp yanıtlarını değiştirememesi, sık kullanılan maddeler sebebiyle ortaya çıkabilecek güvenlik ve gizlilik sorunları, bilgisayar kullanma gerekliliği, bilgisayar donanımındaki sınırlılıklar ve maliyet de BBT'nin dezavantajları arasında gösterilebilir (Hambleton ve Swaminathan, 1985; Lord ve Stocking, 1988; Rudner, 1998). Sahip olduğu dezavantajlara karşılık sağladığı önemli avantajlar sebebiyle, BBT'ye olan ilgi her geçen gün artmakta ve BBT'nin daha yaygın olarak kullanılmasına yönelik çalışmalar önem kazanmaktadır.

Geleneksel kağıt-kalem testlerinden farklı olarak, BBT'de, farklı yetenek düzeylerine sahip olan bireyler, farklı test formlarıyla test edilmektedir. BBT'nin amacı, her bir bireyin gizil yeteneğini kestirmek ve bireyin mevcut performansına dayalı olarak madde havuzundan birey için en uygun (en çok bilgi sağlayan) test maddelerini seçmektir. BBT, belirli bir test algoritmasının uygulanmasını içerir. Bu algoritma iki ana bileşenden oluşur. İlk bileşen, bireylerin yeteneklerini maddelere verdikleri yanıtlara bağlı olarak belirleyen istatistiksel bir hesaplamayı içerir. İkinci bileşen ise her maddenin uygulanmasından sonra bireyin kestirilen geçici yetenek düzeyine en uygun madde seçiminin yapılmasıdır. Uygulanan algoritmayla kalibre edilmiş bir madde havuzu kullanılarak, BBT'nin başlama, devam etme ve sonlandırılması kontrol edilir (Eggen ve Straetmans, 2000). Süreç sonunda ise BBT'lerin geleneksel testlerden daha az sayıda maddeyle daha güvenilir yetenek kestirimleri sağladığını söylemek mümkündür (Fan, Wang, Chang ve Douglas, 2012; Thompson, 2009). BBT'nin sağladığı bu avantajlar, özellikle yurt dışında, Graduate Management Admission Test (GMAT), Graduate Record Examination (GRE), Adaptive Matrices Test (AMT), Measures of Academic Progress (MAP), the Armed Services Vocational Aptitude Battery (ASVAB), the National Assessment of Educational Progress (NAEP), the National Council of Licensure Examinations (NCLEX), North American Pharmacist Licensure Examination (NAPLEX) gibi büyük ölçekli BBT uygulamalarının tercih edilmesinin temel nedeni olarak görülebilir.

Bazen bireyleri test etmenin temel amacı etkili bir yetenek kestirimine ek olarak, geçti-kaldı gibi iki veya daha fazla kategoride yapılacak sınıflamaların doğruluğu olabilir. Bu durumda Bireyselleştirilmiş Bilgisayarlı Sınıflama Testlerinin (BBST; Computerized Adaptive Classification Testing: CACT) kullanımı tercih edilir. BBT'nin odak noktası bireyin yetenek kestirimi iken BBT'nin bir alt konu alanı

olarak düşünebileceğimiz BBST uygulamaları, yetenek kestiriminden ziyade bireyleri başarı, yetenek veya diğer özellikleri bakımından kategorilere doğru olarak sınıflama amacı taşımaktadır. Sınıflama sonucuna bağlı olarak alınabilecek önemli kararlar (örneğin, lise bitirme veya meslek seçimi) nedeniyle etkili ve doğru sınıflamalar yapabilmek, kritik öneme sahiptir (Thompson ve Ro, 2007). Bununla birlikte, eğitim öğretim sürecinin başında, süreç devam ederken ve süreç sonunda yapılacak doğru sınıflamaların bireyleri seçme, izleme veya yerleştirmede sağlayacağı katkılar da doğru sınıflamaların dolayısıyla da BBST uygulamalarının önemini artırmaktadır. Örneğin, tıp alanında uygulanan ve doğrudan insan hayatıyla ilgili olan yüksek riskli testler düşünüldüğünde BBST uygulamaları sonucu elde edilecek yüksek sınıflama doğruluğunun büyük öneme sahip olduğu söylenebilir. Buna bağlı olarak, BBST uygulamalarında, madde seçme yöntemi ve sonlandırma (sınıflama) kriterlerinin önemli bir rol oynadığı söylenebilir. Test maddelerinin seçimi, genellikle Maksimum Fisher Bilgisi (MFB) prensibine dayanarak yapılmaktadır. Sınıflama kriteri olarak ise alanyazında Ardışık Olasılık Oran Testine (AOOT) oldukça sık rastlanmaktadır (Sie, Finkelman, Riley ve Smits, 2015; Spray ve Reckase 1996; Thompson, 2011). Alanyazın incelendiğinde, BBST'de alınacak sınıflama kararları için istatistiksel hipotez testiyle verilen AOOT sınıflama kriterinin, istatistiksel kestirime dayanan yöntemlere iyi bir alternatif olduğunu gösteren birçok çalışma olduğu da görülmektedir (Eggen ve Straetmans, 2000; Lewis ve Sheenan, 1990; Reckase, 1983).

BBT ve BBST ile ilgili yapılan açıklamalar göz önüne alınarak eğitim alanında kullanılan testler, ölçme bakış açısıyla değerlendirilecek olursa, testlerin genellikle yetenek kestirimi ve sınıflama olarak ifade edilen iki büyük amacı arasında yapılacak ayırımın yeterli olacağı söylenebilir. Yetenek kestirimi yapılacak olması durumunda, test etmenin amacı, iyi tanımlanmış bir içerik alanında, bireyin yetenek, beceri veya yeterliliğini kestirmektir. Sınıflamadaki temel amaç ise, bireyin performansının tam/hassas (precisely) bir kestirimini elde etmekten çok, bireyin hangi kategoride sınıflanacağı konusunda bir karara varmaktır. Bunların yanı sıra “test etkililiği” hem BBT’ler hem de BBST’ler için önemli bir kavramdır. BBT’lerde bir test, hazırlanışı ve uygulanışı bilgisayara uyarlanarak bireyselleştirilir. BBT’ler, bir MTK modeli temelinde kalibre edilmiş bir madde havuzu gerektirmektedir. Tek boyutlu madde havuzuna sahip BBT

uygulamalarında bir, iki veya üç parametrelili lojistik MTK modellerinden biriyle kalibre edilmiş madde havuzu kullanılarak yapılan yetenek kestirimi için, mümkün olduğunca az madde ve düşük standart hatalar anlamına gelen yüksek etkililiğe sahip testler uygulanır (van der Linden ve Hambleton, 1996; aktaran, Thompson, 2009). BBST'lerde ise BBT'den farklı olarak, mümkün olduğunca az madde kullanılırken, düşük sınıflama hataları hedeflenerek test etkililiği sağlanmaktadır (Thompson, 2009). Dolayısıyla BBST'de, bireyselleştirilmiş madde seçimiyle hem testin ölçüm etkililiğini artırmak amaçlanır hem de BBT'lerin diğer avantajlarına sahip olunabilir (Eggen, 2011).

BBT ve BBST uygulamalarının ortak noktaları bulunmakla birlikte bazı temel farklılıkları da vardır. Her ikisi de genellikle madde tepki kuramına (MTK) dayandırılan bu iki uygulamadan, BBT'ler, bireyin yetenek parametresini (θ), mümkün olan en doğru ve kesin/hassas şekilde tahmin etmeye çalışırken; BBST bireyleri belirli kategorilerde sınıflamayı amaçlar. Ayrıca değişken uzunluktaki testler için madde seçme yöntemleri ve test sonlandırma kuralları da BBT ve BBST arasındaki farklara örnek olarak verilebilir. Kısaca ifade etmek gerekirse, BBST'de, BBT'den farklı olarak, birey doğru bir şekilde sınıflandığı sürece, bireyin yetenek düzeyinin belirlenmesi özel olarak ele alınan bir konu değildir (Wouda ve Eggen, 2009). Değişken uzunluktaki BBST'lerde AOOT'ye ya da farklı sınıflama kriterlerine göre birey önceden tanımlanmış kategorilerden birinde sınıflandığında test sonlandırılırken, değişken uzunluktaki BBT'lerde bireyin yeteneği (θ) önceden belirlenmiş duyarlılıkta ya da farklı sonlandırma kuralları için kestirildiğinde test sonlandırılmaktadır. BBT'ler için madde seçme yöntemleri, geçici (interim) yetenek kestiriminde Maksimum Fisher Bilgisi (MFB) veya Kullback-Liebler Bilgisi (KLB) gibi bazı bilgi indekslerinin en üst düzeye çıkarılmasına dayalı olarak bireyselleştirilmektedir. BBST'lerde ise madde seçimi yine maksimum bilgiyi kullanarak bireyselleştirilebilir, ancak yaygın olan uygulama geçti-kaldı kararlarını ayıran kesme noktasında (θ_0) bilgiyi en üst düzeye çıkaracak madde seçimini yapmaktır (Huebner, 2012). Bu yöntem BBST'ler için etkili ve kavramsal olarak tutarlı olarak görülmektedir (Spray ve Reckase, 1996).

BBST uygulamaları BBT'ye benzer şekilde; psikometrik model, kalibre edilmiş madde havuzu, başlama noktası, madde seçme yöntemi ve sonlandırma (sınıflama) kriterleri olmak üzere beş temel bileşenden oluşmaktadır (Thompson,

2007b). Bunların yanı sıra temel bileşenlerden olmamakla birlikte, daha az maddeyle geçerli ve güvenilir ölçümler yapabilmek için madde kullanım sıklığı ve içerik dengeleme gibi bazı pratik kısıtlamaların BBT uygulamalarında olduğu gibi BBST uygulamalarında da dikkate alınması büyük önem taşımaktadır. Madde kullanım sıklığı kontrolüyle test maddelerinin güvenliği sağlanmaya çalışılırken, içerik dengelemesiyle de geçerliği daha yüksek testler oluşturulmaya çalışılmaktadır. Ancak alanyazındaki birçok BBST çalışmasında, madde seçme yöntemleri arasındaki farklılıkların getirilen pratik kısıtlamalarla ortadan kalkacağına dair kaygılar nedeniyle madde kullanım sıklığı ve içerik dengeleme kontrolü yapılmadığı görülmektedir (Lin, 2011). Bu araştırmada, iki, üç veya dört kategoride yapılacak sınıflamalar için ortalama test uzunluğu (OTU), ortalama sınıflama doğruluğu (OSD) ve ölçme kesinliği bakımından daha gerçekçi bir karşılaştırma sağlamak amacıyla, farklı sınıflama kriterleri ve madde seçme yöntemlerinin yanı sıra farklı içerik dengeleme ve madde kullanım sıklığı kontrol yöntemleri gibi pratik kısıtlamalar da dikkate alınmıştır.

İçerik dengeleme, hem BBST'ler hem de BBT'ler için önemli bir sorundur. Eğitim ve öğretim süreçleri için belirlenen hedefler ve beceri alanları genelde zorlukları bakımından farklıdır ve bu nedenle içerik dengelemesi yapılmamış bir bireyselleştirilmiş testte yapılan madde seçimleriyle test amacına tam olarak ulaşamaz (Linn, 1989; aktaran, Kingsbury ve Zara, 1989). Bu da bireylerin geçerliği düşük testlerle değerlendirilmelerine sebep olur. Kingsbury ve Zara'ya (1989) göre büyük bir madde havuzunda, içerik dengelemesi yapılmadan BBT uygulaması gerçekleştirmek oldukça zordur. Uygulanacak bireyselleştirilmiş testte içerik dengelemesi yapıldığında, bireyler, içerik alanlarının her birini mümkün olduğu kadar uygun şekilde temsil eden, geçerliği daha yüksek bir testle karşılaşmaktadır. Alanyazın incelendiğinde, içerik dengelemesi yapılan BBST çalışmalarında kullanılan içerik dengeleme yöntemlerinin genel olarak Kingsbury ve Zara'nın (1989) önerdiği "spiral yöntem" (örneğin, Finkelman, 2008; Huebner, 2012) ve Kısıtlanmış BBT (KBBT) yöntemi (örneğin, Eggen ve Straetmans, 2000; Huebner ve Li, 2012) olduğu görülmektedir.

BBST'ler ve BBT'ler için bir diğer önemli sorun madde kullanım sıklığı kontrolüdür. Çok fazla kullanılmış (overexposed) maddeler, sınava giren bireylerin bu tür maddelerle ilgili bilgiyi sınava daha sonra girecek olan bireylerle paylaşma

olasılığında dolayı testlerin geçerliğini tehdit eden güvenlik ve gizlilik endişelerini ortaya çıkarmaktadır. Doğru yapılan bir "geçti-kaldı" sınıflamasında, yeterli olanın dışında bir maddeyi uygulamamak oldukça önemlidir. Çünkü daha kısa bir sınav, sınava giren kişinin test süresini azaltırken, testin kullanılabilirliğini artırmaktadır. Buna ek olarak, daha kısa bir test, bazı maddelerin kullanım sıklığı oranlarını azaltarak, güvenlik-gizlilik sorunlarının önüne geçer ve böylece madde havuzlarının çok sık yenilenmesi zorunluluğu da ortadan kalkar (Finkelman, 2008). Madde kullanım sıklığındaki artış, maddenin ele geçirilme olasılığını artırır ve bu da geçerliği düşürerek maddenin bireyin yeteneğini değerlendirmede daha az etkili olmasına neden olur (Barret, 2015). Bu durum özellikle çok düşük miktardaki hatayı tolere edebilen yüksek riskli testler için büyük bir sorundur.

Madde kullanım sıklığı problemi, ardışık madde seçimini kullanan BBST'lerde daha da büyük bir sorun haline gelmektedir. Bunun nedeni, ardışık madde seçiminin, bireylere uygulanacak maddeleri nasıl belirlediği göz önüne alınarak anlaşılabilir. Bu yöntemle, havuzdaki maddeler kesme noktasında (θ_0) verdikleri maksimum bilgiye göre en çok bilgiyi verenden en az bilgiyi verene doğru sıralanır. Herhangi bir madde kullanım sıklığı kontrol yönteminin bulunmadığı bir durumda bireye, havuzdaki en yüksek bilgiyi veren madde, daha sonra ikinci en yüksek bilgiyi veren madde uygulanacak ve bu şekildeki madde seçimi bir sınıflama kararı verilip test sonlandırılıncaya kadar devam edecektir. Böylece, sınava giren her bireyin havuzdaki en yüksek bilgiyi veren maddeleri görmesi kaçınılmaz bir sonuç olacaktır. Öte yandan, BBT'de bireylere, kestirilen geçici yetenek düzeylerinde maksimum bilgiyi veren maddeler uygulanır. Dolayısıyla BBT uygulamasına katılan belirli sayıdaki bireyin, θ yetenek düzeylerinin farklı olması nedeniyle tam olarak aynı madde setini görmeyecekleri olasılığı, bir tür doğal madde kullanım sıklığı kontrolü görevi yapmaktadır. Uygulanan madde kullanım sıklığı kontrolüyle, genellikle istenen maksimum madde kullanım sıklığı oranı (r^{max}) belirlenerek havuzdaki her maddenin kullanım sıklığını bu orana eşitlemek ya da bu orandan daha düşük tutmak amaçlanmaktadır. Alanyazın incelendiğinde, hem BBST'ler hem de BBT'ler için en sık kullanılan madde kullanım sıklığı kontrol yönteminin Sympson-Hetter (SH; Sympson ve Hetter, 1985) olduğu görülmektedir. Ancak, SH yönteminin olasılıklı doğasından dolayı, özellikle ardışık madde seçimi esnasında maddelerin büyük bir

bölümünde r^{max} oranını aşıldığının gözlenebilmesi mümkündür (Huebner, 2012). Bunun dışında Kısıtlayıcı Yöntem (KS; Restricted Method-RT; Revuelta ve Ponsoda, 1998) ve Madde Uygunluk Yöntemi (MU; Item Eligibility Method-IE; van der Linden ve Veldkamp, 2004) de madde kullanım sıklığı kontrolünü sağlamada kullanılan, benzer mantıkla çalışan diğer yöntemler arasındadır.

Sonuç olarak, alanyazın incelendiğinde bireylerin yeteneklerini etkili bir şekilde kestirebilmek amacıyla geliştirilmiş olan BBT'lerin, bireyleri sınıflamada yaşanan problemler için de oldukça yararlı olduğunu söylemek mümkündür. Alanyazındaki BBST çalışmalarında genellikle, "geçti-kaldı" gibi sadece iki kategorinin ve tek bir kesme noktasının bulunduğu; bazı sınıflamalarda ise birden çok kesme noktasının kullanıldığı görülmüştür. Huebner (2012), Lin ve Spray (2000), Reckase (1983), Sie ve diğerleri, (2015), Spray ve Reckase (1994), Thompson (2009), van Groen, Eggen ve Veldkamp'ın (2016) yaptıkları çalışmalarda belirledikleri tek kesme puanıyla bireyleri "geçti-kaldı" gibi iki kategoriden birine sınıflamayı; Eggen (1999), Thompson (2007a) çalışmalarında belirledikleri bir ve iki kesme puanıyla bireyleri iki ve üç kategoride sınıflamayı; Eggen ve Straetmans (2000) çalışmalarında belirledikleri iki kesme puanıyla bireyleri üç kategoride sınıflamayı; Nydick, Nozawa ve Zhu (2012) çalışmalarında belirledikleri iki ve dört kesme puanıyla bireyleri üç ve beş kategoride sınıflamayı; Yang, Poggio ve Glasnapp (2006) ise yaptıkları çalışmada belirledikleri dört kesme puanıyla bireyleri beş kategoriden birinde sınıflamayı amaçlamışlardır. Alanyazın incelendiğinde, sınıflama kategori sayısının OTU ve OSD üzerindeki etkileriyle ilgili tam bir netlik olmadığı göze çarpmaktadır. Örneğin, iki ve üç kategoride yapılan sınıflamalar sonucunda Thompson (2007a) kategori sayısı arttıkça testin uzadığını ve sınıflama doğruluğunun azaldığını ifade etmişken Eggen (1999) ise OTU ve OSD açısından benzer sonuçlar elde edildiğini ifade etmiştir. Bunun yanı sıra BBST olarak uygulanmasa da farklı kategorilerde sınıflamaların yapıldığı uluslararası sınavlar incelendiğinde, PISA'da (Programme for International Student Assessment) öğrencilerin yetenek düzeylerine göre altı kategoriye (1. Düzey, 2. Düzey, 3. Düzey, 4. Düzey, 5. Düzey, 6. Düzey) ayrıldığı, TIMSS'de (Trends in International Mathematics and Science Study) ise dört kategoriye (alt düzey, orta düzey, üst düzey, ileri düzey) ayrıldığı görülmektedir. Ulusal sınavlara bakıldığında ise ABİDE (Akademik Becerilerin İzlenmesi ve Değerlendirilmesi) projesiyle

öğrencileri yetenek düzeylerine göre beş kategoride (temel altı yetenek düzeyi, temel yetenek düzeyi, orta yetenek düzeyi, orta üstü yetenek düzeyi, ileri yetenek düzeyi) sınıflamanın amaçlandığı görülmektedir. Bu açıklamalar ışığında, ileride yapılacak araştırmalar için fikir sağlayıcı nitelikte olabilmesi amacıyla, mevcut BBST çalışmasında sınıflama kategori sayısı da dikkate alınması gereken bir değişken olarak düşünülmüştür. Dolayısıyla bu çalışmada yapılan diğer çalışmalardan farklı olarak; çeşitli sınıflama kriterleri ve madde seçme yöntemlerine göre içerik dengeleme ve madde kullanım sıklığı kontrolü altında bireyler iki, üç veya dört kategoride sınıflandırıldığında elde edilen sonuçlar OTU, OSD ve ölçme kesinliğinin bakımından karşılaştırılmıştır.

Araştırmanın Amacı ve Önemi

Bu araştırmanın temel amacı, BBST uygulamalarında iki kategorili puanlanan maddelerden oluşan tek boyutlu madde havuzu için simülasyonla üretilen veri seti üzerinden, kesme noktasının bir, iki veya üç (yani sınıflama yapılacak kategori sayısının; iki, üç veya dört) olması durumunda ortalama test uzunluğu (OTU), ortalama sınıflama doğruluğu (OSD) ve ölçme kesinliğinin; farklı sınıflama kriterlerine (Ardışık Olasılık Oran Testi: AOOT ve Güven Aralığı: GA) ve farklı madde seçme yöntemlerine (Kestirilen Yetenek Temelinde Maksimum Fisher Bilgisi: MFB-KY ve Kesim Noktası Temelinde Maksimum Fisher Bilgisi: MFB-KN) göre içerik dengeleme (Kısıtlanmış Bireyselleştirilmiş Bilgisayarlı Test: KBBT ve Modifiye Edilmiş Multinomial Model: MMM) ve madde kullanım sıklığı kontrolü (Sympson-Hetter Yöntemi: SH ve Madde Uygunluk Yöntemi: MU) gibi pratik kısıtlamalar altında nasıl değiştiğinin incelenmesidir. Çalışmada, yetenek kestirim yöntemi olarak tercih edilen Ağırlıklandırılmış Olabilirlik Kestirimi (AOK), sabit kalan bir simülasyon koşuludur. Problem durumunda açıklanan araştırma koşullarını oluşturan değişkenler Tablo 1’de özetlenmiştir.

Tablo 1

Araştırma Kapsamında İncelenen Değişkenler

Sınıflama Kriterleri	Madde Seçme Yöntemleri	Pratik Kısıtlamalar	Sınıflama Kategori Sayısı
AOOT ($\delta:0,20$) GA (%90)	MFB-KY MFB-KN	İçerik Dengeleme Yöntemleri	KBBT MMM İki Üç Dört
		Madde Kullanım Sıklığı Kontrol Yöntemleri	SH MU

Tablo 1’de görüldüğü gibi araştırmada toplam, *2 sınıflama kriteri x 2 madde seçme yöntemi x 2 içerik dengeleme yöntemi x 2 madde kullanım sıklığı kontrol yöntemi x 3 sınıflama kategori sayısı = 48 adet koşul* oluşturulmuştur.

Yurtiçi alanyazın incelendiğinde, BBST uygulamasına yönelik olarak sadece son yıllarda yapılmış ve sayıca çok az çalışmaya (Gündeğer, 2017; Gündeğer, 2018; Gündeğer ve Doğan, 2018) rastlanmaktadır. Gündeğer (2017) çalışmasında; farklı sınıflama kriterleri, yetenek kestirim yöntemleri ve madde seçme yöntemlerini çaprazlayarak iki kategoride yapılan sınıflamalardan elde edilen sonuçları OTU, OSD ve ölçme kesinliği bakımından karşılaştırmıştır. Bununla birlikte Gündeğer (2018) yaptığı başka bir çalışmada farklı büyüklükte ve çok kategorili maddelerden oluşan madde havuzlarında farklı yetenek kestirim yöntemlerinin etkililiğini incelemiştir. Gündeğer ve Doğan’ın (2018) çalışmalarında ise farklı madde seçme yöntemleri, farklı büyüklükteki sivri ve basık madde havuzları üzerinden OTU ve OSD bakımından karşılaştırılmıştır. Yurt dışında yapılan çalışmalar incelendiğindeyse, BBST ile ilgili daha fazla sayıda çalışma olduğu görülmektedir. Çalışmaların büyük çoğunluğunda, sınıflamanın sadece iki kategoride yapıldığı (örneğin, Lau, 1996; Reckase, 1983; Spray ve Reckase, 1996), içerik dengeleme ve madde kullanım sıklığı kontrolünün dikkate alınmadığı, genel olarak yalnızca sınıflama kriterlerinin (örneğin, Kingsbury ve Weiss, 1980; Nydick, 2013) veya madde seçme yöntemlerinin (örneğin, Eggen, 1999; Lin ve Spray, 2000) karşılaştırıldığı ya da farklı madde seçme yöntemleriyle çaprazlanan sınıflama kriterlerinin (örneğin, Eggen ve Straetmans, 2000; Thompson ve Ro, 2007) performanslarının incelendiği görülmektedir. Bunun yanı sıra sınıflama kriterinin performansının OTU ve OSD bakımından farklı madde kullanım sıklığı kontrol yöntemlerine göre karşılaştırıldığı (Huebner, 2012; Lau ve Wang, 1999), SH madde kullanım sıklığı kontrol yöntemiyle birlikte içerik dengeleme amacıyla spiral yöntemin kullanıldığı (Huebner ve Li, 2012) az sayıda çalışmaya da rastlanmaktadır.

Alanyazın incelendiğinde, tek boyutlu ve iki kategorili madde havuzunda; farklı sınıflama kriterleri ve farklı madde seçme yöntemlerinin yanı sıra farklı madde kullanım sıklığı kontrol yöntemleri, farklı içerik dengeleme yöntemleri ve farklı kategori sayılarının da çaprazlanmasıyla yapılan sınıflamaların OTU, OSD ve ölçme kesinliği bakımından incelendiği bir çalışmaya rastlanmamıştır. Çalışmanın

bu yönüyle yurtiçi ve yurtdışı alanyazına sağlayacağı katkı nedeniyle önemli olduğu düşünülmektedir. Ayrıca bilgi ve iletişim teknolojilerindeki gelişmelerin eğitim alanındaki yansımalarından biri olarak, zeka testleri de dahil olmak üzere özellikle büyük ölçekli sınavları uygulamada yapılmak istenen değişiklikler açısından da kurumlar, eğitimciler ve uygulayıcılar için fikir sağlayıcı nitelikte olacağı düşüncesi araştırmanın önemini artırmaktadır. Bununla birlikte, test sonuçlarına göre yapılacak doğru sınıflamaların bireyleri seçme, izleme veya yerleştirmede sağlayacağı katkılar göz önüne alındığında, özellikle yurtiçinde çok az sayıda olduğu görülen BBST çalışmalarına farklı araştırma desenleriyle yenilerinin katılmasının gerekli olduğu, ayrıca elde edilen sonuçların ölçme ve değerlendirme alanyazını açısından da önemli olabileceği düşünülmektedir.

Araştırma Problemi

İki kategorili puanlanan maddelerden oluşan tek boyutlu bir madde havuzunda gerçekleştirilen BBST uygulamalarında, yetenek kestirim yöntemi AOK iken sınıflama yapılacak kategori sayısının iki, üç veya dört olması durumunda ortalama test uzunluğu (OTU), ortalama sınıflama doğruluğu (OSD) ve ölçme kesinliği; sınıflama kriterlerine ve madde seçme yöntemlerine göre içerik dengeleme ve madde kullanım sıklığı kontrolü gibi pratik kısıtlamalar altında nasıl değişmektedir?

Alt problemler. Problem cümlesine ilişkin oluşturulan alt problemler aşağıda ifade edilmiştir.

1. AOOT sınıflama kriterinin farksızlık bölgesi (FB) sabiti δ : 0,20 düzeyi ve GA sınıflama kriterinin GA: %90 güven düzeyi için yetenek kestirim yöntemi AOK iken iki, üç veya dört kategorili sınıflamalarda;
 - 1.1. MFB-KY, MFB-KN madde seçme yöntemlerinin;
 - 1.2. MFB-KY, MFB-KN madde seçme yöntemleri ile KBBT ve MMM içerik dengeleme yöntemlerinin;
 - 1.3. MFB-KY, MFB-KN madde seçme yöntemleri ile SH ve MU madde kullanım sıklığı kontrolü yöntemlerinin;

- 1.4. MFB-KY, MFB-KN madde seçme yöntemleri ile KBBT, MMM içerik dengeleme yöntemleri ve SH, MU madde kullanım sıklığı kontrolü yöntemlerinin;

bir arada ele alındığı koşullarda OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

2. AOOT (δ : 0,20) ve GA (%90) sınıflama kriterlerine göre OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

3. MFB-KY ve MFB-KN madde seçme yöntemlerine göre OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

4. KBBT ve MMM içerik dengeleme yöntemlerine göre OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

5. SH ve MU madde kullanım sıklığı kontrol yöntemlerine göre OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

6. Sınıflama kategori sayısına göre OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

Sayıtlılar

1. Bu çalışmada MC simülasyon verisi kullanılmış olup; ikili puanlanan 500 maddeden oluşan tek boyutlu madde havuzu 3 PLM ile kalibre edilmiştir. Madde parametreleri; a parametresi orta ve yüksek ayırıcılık düzeylerini temsil edebilmesi amacıyla tek biçimli dağılımdan $U[0,5, 2,0]$, b ve c parametreleri ise gerçek uygulamadaki değerlere yakın olabilmesi amacıyla normal dağılımdan sırasıyla $N(-0,5, 1,5)$ ve $N(0,20, 0,05)$ olacak şekilde türetilmiştir. Bunun yanı sıra bireylerin yetenek parametreleri $(-3, +3)$ aralığında ve normal dağılımdan $(N(0,1))$ toplam 5000 birey için R yazılımında rastgele türetilmiştir. Çalışma kapsamında oluşturulan bu simülasyon koşullarının gerçek veri koşullarını sağladığı varsayılmıştır.

Sınırlılıklar

1. Bu çalışmada, tek boyutlu MTK'ya dayalı 3 PLM temel alınmıştır.
2. Çalışmadaki tüm koşullar için başlama kuralı $\theta = 0$ olarak belirlenmiştir.
3. Bu çalışma; AOOT (δ : 0,20) ve GA (%90) sınıflama kriterleri, MFB-KY ve MFB-KN madde seçme yöntemleri, KBBT ve MMM içerik dengeleme yöntemleri, SH ve MU madde kullanım sıklığı kontrol yöntemleri gibi belirli koşulları içeren MC simülasyonu ile sınırlıdır. Yapılan karşılaştırmalar yalnızca bu koşullar için gerçekleştirilmiştir.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde araştırmanın kuramsal temeli, ilgili araştırmalar ve ilgili araştırmaların özetine yer verilmiştir.

Bireyselleştirilmiş Bilgisayarlı Sınıflama Testi (BBST)

Eğitim alanında kullanılan testler ve psikolojik testlerin bireye uyarlanmış şekilde bilgisayarla uygulanması pek çok avantaj sunmaktadır. Psikometrik model olarak tercih edilebilecek MTK sayesinde oluşturulan değişken uzunluktaki testlerle (variable-length testing) birey kendisine çok kolay ya da çok zor gelecek maddelerle değil yetenek düzeyine en uygun maddelerle karşılaşır. Bu avantaj, madde havuzunda bulunan maddelerden hangisinin, belirli kısıtlamalar dahilinde, uygulanacak bir sonraki "en iyi" madde olduğunun belirlenmeye çalışıldığı akıllı madde seçim algoritmaları sayesinde elde edilir. Değişken uzunluktaki test, sınava giren her bireyin aynı sayıda maddeyi almadığı testleri ifade eder. Öyle ki, çok az sayıda madde kullanıldıktan sonra bile sınav amacına ulaşırsa test sonlandırılır (Vos, 2000). Yetenek (θ) kestirimi için en yaygın olarak kullanılan değişken uzunluktaki test yöntemi, BBT'lerdir (Weiss ve Kingsbury, 1984). Ancak testin amacı yetenek kestiriminden ziyade, bireyi belirli iki veya daha fazla kategoride sınıflamak olduğunda ise kullanılan yöntem BBST olarak adlandırılır (Lin ve Spray, 2000). Aynı maddelerden oluşan sabit test uzunluğuna sahip geleneksel kağıt-kalem testlerinin aksine BBST'lerde maddeler, BBT'lerde olduğu gibi, test sırasında ve uygulanan her bir maddenin ardından yeni bir maddenin seçimiyle gerçekleştirilir (Lin ve Spray, 2000; Spray ve Reckase, 1994). BBST için kullanılan madde seçme yöntemleri genel olarak, kestirilen yetenek temelli (KY) ve kesme noktası temelli (KN) olmak üzere iki türe ayrılmaktadır.

Weiss ve Kingsbury'ye (1984) göre bireyselleştirilmiş bir test geliştirmek için altı BBT bileşeni ile ilgili kararlar verilmelidir. Bunlar: (i) Psikometrik model, (ii) Madde havuzu, (iii) Başlama kuralı, (iv) Madde seçme yöntemi, (v) Yetenek kestirim yöntemi, (vi) Sonlandırma kuralıdır. Thompson'a (2007b) göre ise BBST uygulamaları, BBT'ye benzer olup beş temel bileşenden oluşmaktadır. Bu bileşenler; (i) Psikometrik model, (ii) Kalibre edilmiş madde havuzu, (iii) Başlama noktası, (iv) Madde seçme algoritması, (v) Sonlandırma (sınıflama) kriterleri olarak

sıralanmaktadır. İfade edilen bileşenler göz önüne alındığında BBST uygulamalarının, BBT'den yetenek kestirimi noktasında farklılaştığını söylemek mümkündür (Thompson, 2007b). Daha önceden değinildiği üzere, BBT'de test, bireyin yeteneğinin nokta kestirimi istenen kesinlikte yapıldığında sonlanır (Weiss ve Kingsbury, 1984), BBST'de test, birey önceden belirlenen kategorilerden birinde sınıflandığında sonlanmaktadır (Thompson, 2007b). Sonuç olarak, Thompson (2007b) ile Weiss ve Kingsbury'nin (1984) çalışmaları ışığında BBST bileşenleri genel olarak Tablo 2'deki gibi özetlenebilir.

Tablo 2

BBST Bileşenleri

BBST Bileşenleri	Mevcut Seçenekler	Örnek Çalışmalar
Psikometrik Model	KTK	Frick, 1992; Linn, Rock, ve Cleary, 1972; Rudner, 2002
	MTK	Eggen ve Straetmans, 2000; Kingsbury ve Weiss, 1983; Lau ve Wang, 1998; Reckase, 1983
Madde Havuzu	Sivri	Gündeğer ve Doğan, 2018; Xiao, 1999
	Basık	Finkelman, 2003 (aktaran, Thompson, 2007b); Kingsbury ve Weiss, 1983; Yang ve diğerleri 2006
Başlama Noktası	$\Theta = 0$ veya $LR=1$	Gündeğer, 2017; Gündeğer ve Doğan, 2018
	Ön bilgi	Yang ve diğerleri 2006
Madde Seçimi	Kestirim Temelli	Eggen, 1999; Kingsbury ve Weiss, 1983; Reckase, 1983
	Kesim Noktası Temelli	Eggen, 1999; Spray ve Reckase, 1994, 1996
	Global	Weissman, 2004 (aktaran, Thompson, 2007b)
Yetenek Kestirimi	MOK	Reckase, 1983; Yang ve diğerleri 2006
	AOK	Nydick ve diğerleri, 2012; Yang ve diğerleri 2006
	BSD	Gündeğer ve Doğan, 2018; Yang ve diğerleri 2006
	MSD	Yang ve diğerleri 2006
	Owen'in Bayesçi Yaklaşımı	Yang ve diğerleri 2006
Sonlandırma (sınıflama) Kriteri	AOOT	Eggen, 1999; Eggen ve Straetmans, 2000; Kingsbury ve Weiss 1980; Nydick ve diğerleri, 2012; Reckase, 1983; Spray ve Reckase, 1996; Thompson 2007a; Thompson, 2011
	GOO	Nydick ve diğerleri, 2012; Thompson, 2011
	BUT	Kingsbury ve Weiss 1980; Sie ve diğerleri, 2015; Spray ve Reckase, 1996; Thompson 2007a
	GA	Chang, 2005; Eggen ve Straetmans, 2000; Kingsbury ve Weiss, 1983; Nydick ve diğerleri, 2012; Thompson, 2011
	BKK	Glas ve Vos, 2006 ; Vos, 2000

Tablo 2'ye göre BBST uygulamaları, psikometrik model olarak KTK veya MTK yaklaşımı temelinde; sivri veya basık madde havuzları üzerinden; teste başlama kuralı olarak yetenek düzeyi $\theta = 0$ alınarak veya bireylere ait ön bilgiler kullanılarak; madde seçiminde kestirim temelli, kesme noktası temelli veya global (ortak) yöntemler tercih edilerek; yetenek kestiriminde MOK, AOK, veya Bayesçi yaklaşımlar (BSD, MSD, Owen'in Bayesçi Yaklaşımı) kullanılarak; ve testi sonlandırmada ise Thompson'ın (2009) ifade ettiği gibi temelde AOOT, BUT veya BKK olmak üzere toplam beş farklı sınıflama kriteri (AOOT, GOO, BUT, GA, BKK) seçeneğiyle gerçekleştirilebilmektedir.

Tablo 2'de görüldüğü gibi, altı farklı BBST temel bileşeninden ve bu bileşenlerin her biri için de farklı seçeneklerden bahsetmek mümkündür. Herhangi bir BBST uygulamasında, bu temel bileşenlerin sunulan seçeneklerden hangileriyle temsil edileceği, etkili bir BBST uygulaması için kritik öneme sahiptir. Çünkü BBST bileşenlerinin uygun şekilde seçimi ve tasarımı, testin etkililiğini ve güvenilirliğini belirleyen önemli unsurlardır (Thompson, 2009). Bunların yanı sıra temel bileşenlerden olmamakla birlikte, daha az maddeyle geçerli ve güvenilir ölçümler yapabilmek için pratik ek kısıtlamalar olarak ele alınabilecek madde kullanım sıklığı ve içerik dengelemenin BBT uygulamalarında olduğu gibi BBST uygulamalarında da dikkate alınması büyük önem taşımaktadır (Davey ve Parshall, 1995). Bu sebeple mevcut araştırmada madde kullanım sıklığı ve içerik dengeleme de ele alınmıştır.

Psikometrik model. BBST uygulamalarının ilk bileşeni, psikometrik modelin seçimidir. BBST'ler, Madde Tepki Kuramı (MTK) veya Klasik Test Kuramı (KTK) temelinde ele alınabilir (Parshall, Spray, Kalohn ve Davey, 2006). KTK yaklaşımı, kavramsal olarak daha basit olması ve küçük örneklemelere uygulanabilirlik avantajından (Rudner, 2002) dolayı oldukça etkili (Frick, 1992) olmasına rağmen, BBST uygulamalarında genellikle MTK yaklaşımının kullanımı tercih edilmektedir. Bunun nedeni, MTK yaklaşımında; hem maddelerin, yetenek düzeylerinin ve kesme puan(lar)ının aynı ölçekte yer alması (Thompson, 2007b) hem de madde parametrelerinin örneklemeden, yetenek parametrelerinin ise maddelerden bağımsız olarak kestirilebiliyor olmasıdır (Hambleton ve Swaminathan, 1985). Bu yönüyle MTK, BBST'nin diğer bileşenleri için de önemli bir avantaj sağlamaktadır (Thompson, 2007b). Alanyazındaki BBST çalışmalarının

büyük çoğunluğunda, ikili puanlanan maddeler için 1 Parametrelili Lojistik Model (1 PLM), 2 PLM, 3 PLM gibi MTK modelleri (Eggen, 1999; Reckase, 1983; Spray ve Reckase, 1996) kullanılmış olsa da Genelleştirilmiş Kısmi Kredi Modeli, Aşamalı Tepki Modeli, Kısmi Puan Modeli gibi çoklu puanlanan maddeler için tercih edilen MTK modellerinin kullanıldığı çalışmalara da (örneğin, Lau ve Wang, 1998) rastlanmaktadır. Bu çalışmada ise kullanılan tek boyutlu ve ikili puanlanan maddeler için MTK yaklaşımı temel alınmıştır.

Madde Tepki Kuramı (MTK). MTK, bireye ait yetenek düzeyi aracılığıyla, bireyin testteki performansının kestirilebileceğini varsaymaktadır. Başka bir ifadeyle bireyler için kestirilen yetenek puanları sayesinde bireyin belirli bir madde veya testte göstereceği performans tahmin edilir (Lord ve Novick, 1968, aktaran, Hambleton ve Swaminathan, 1985). MTK, bireyin gözlenen test performansı ile gözlenemeyen ancak test performansının altında yattığı varsayılan baskın yetenek(ler) arasında ilişki kurar. Dolayısıyla, bireyin gözlenen test performansı üzerinden gözlenemeyen gizil özellikleriyle ilgili kestirimler yapılır. Bireyin maddeye ilişkin gözlenen performansı ile bu performansını etkilediği varsayılan yetenek düzeyi arasındaki ilişki monoton artan ve madde karakteristik eğrisi (MKE) olarak adlandırılan bir fonksiyonla tanımlanır. Yetenek ölçeğinin farklı noktalarındaki bireyler için maddenin doğru yanıtlanma olasılığını gösteren MKE, daha yüksek yetenek düzeyine sahip olan bireylerin, daha düşük yetenek düzeyine sahip olanlara kıyasla, maddeye doğru cevap verme olasılıklarının daha yüksek olduğunu belirtmektedir (Hambleton ve Swaminathan, 1985).

MTK yaklaşımı temelinde, MKE'nin matematiksel formu için çok sayıda mevcut seçenek olduğundan, birbirinden farklı pek çok model kullanılabilir. Testin boyutluluk durumuna göre tek boyutlu veya çok boyutlu MTK modelleri tercih edilebilirken, testteki madde yanıtlarının kategori sayısına göre ikili puanlanan, çoklu puanlanan veya süreklilik özelliği gösteren MTK modelleri tercih edilebilir. Ancak, kullanılacak modelin belirlenmesinde, modelin veriye uyumu yapılacak uyum iyiliği araştırmasıyla incelenmelidir. Çünkü yapılacak yetenek kestirimlerinin doğruluğu seçilen modelin veriye uyumuna bağlıdır. MTK'nın belki de en önemli özelliği, gruptan bağımsız madde parametreleri (madde parametrelerinin değişmezliği) ve uygulanan maddelerden bağımsız yetenek kestirimleri (yetenek parametrelerinin değişmezliği) sağlamasıdır. Bu

özellikler seçilen modelin veri setine uyumuyla doğrudan ilişkilidir (Hambleton ve Swaminathan, 1985). BBT ve BBST uygulamalarında bireylerin madde havuzundan seçilen farklı maddeleri yanıtlama durumları söz konusu olduğundan MTK'nın sağladığı değişmezlik özelliği bu tip bireyselleştirilmiş uygulamalarda MTK'yı KTK'ya kıyasla çok daha cazip hale getirmektedir. Mevcut araştırmada tek boyutlu ve ikili puanlanan maddeler kullanıldığından ilgili varsayımlar ve modellerin matematiksel altyapısı ayrıntılı olarak ele alınmıştır.

MTK varsayımları. Tek boyutluluk, yerel bağımsızlık ve testin hız testi olmaması MTK'nın temel varsayımları olarak ifade edilebilir (Hambleton ve Swaminathan, 1985). Tek boyutluluk, maddenin tek bir boyutu, bir değişkeni ölçmesi anlamında kullanılmaktadır. Bunun için maddenin ölçtüğü ve bireylerin yanıtlama performansları altında yatan baskın tek bir örtük özellik olması gereklidir. Bu demek oluyor ki, bireylerin yanıtlama performansları arasındaki varyans, tek bir örtük özellik tarafından açıklanabilir olmalıdır (Hambleton ve Swaminathan, 1985). Tek boyutluluk varsayımının ihlali, parametrelerin veya standart hatanın yanlış kestirilmesine yol açabilmektedir (Demars, 2010). Bu nedenle tek boyutluluk sağlanmadığında çok boyutlu MTK modellerinin tercih edilmesi daha uygundur.

Yerel bağımsızlık ise tek boyutluluk varsayımının bir sonucu olarak düşünülebilmektedir. Yerel bağımsızlık sağlanıyor ise bireylerin farklı test maddelerine verdiği yanıtların istatistiksel olarak bağımsız olduğu varsayılmaktadır. Bireyin bir maddeye verdiği yanıt diğer maddelerin yanıtlanmasını iyi ya da kötü yönde etkilememelidir. Başka bir deyişle, maddeler birbirinden bağımsız olarak cevaplandırılabilir, maddelerden birine verilen yanıt başka maddeye verilecek yanıtı etkilememeli ve her madde tek başına yanıtlanabilir olmalıdır. Aynı zamanda, belirli bir yetenek düzeyinde iki farklı maddeye verilen yanıtın birbirinden bağımsız oluşunu ifade eden yerel bağımsızlık varsayımı, yetenek düzeyi sabitlendiğinde maddeler arası kovaryansın sıfıra eşitliğinin incelenmesiyle test edilebilir. Bu, yerel bağımsızlık varsayımının zayıf formu olarak da ifade edilmektedir (Hambleton ve Swaminathan, 1985). McDonald'a (1967) göre tek boyutluluk aynı zamanda yerel bağımsızlık varsayımını da içermektedir ve aynı/benzer yetenek düzeyleri için maddeler arasındaki kovaryans sıfır ise test, hem yerel bağımsız hem de tek boyutludur

(aktaran, Hambleton ve Swaminathan, 1985). Buna göre, bir test tek boyutluluk varsayımını sağlıyorsa, testte yer alan maddelerin de yerel bağımsızlık varsayımını sağladığı söylenebilir. Yen'in (1993) yerel bağımsızlık testi ise ayrı ayrı madde çiftleri aralarındaki korelasyona göre yorumlanmaktadır. Yen (1993), 0,20'den yüksek korelasyonlara problemlilik olarak yaklaşılmasını tavsiye etmiştir (aktaran, Demars, 2010). Ancak madde çiftleri arasında yeterince yüksek korelasyon olmasa bile maddelerin bir alt kümesi ikinci bir boyut oluşturabilmektedir. Bu nedenle çoğunlukla tek boyutluluk ve yerel bağımsızlık ayrı ayrı test edilmektedir (Demars, 2010).

Testin hız testi olup olmadığı, yanıtlanmamış madde sayısı ve testin sonuna erişememiş birey sayısı dikkate alınarak kontrol edilebilir. Bunun yanı sıra, hızsızlık testi veya paralel test uygulaması gibi yöntemlerle de kullanılabilir (Hambleton ve Swaminathan, 1985). Hız testlerinde, bireylerin test performansı üzerinde süre ya da hız gibi başka bir boyutun etkisi kaçınılmaz bir durumdur. Dolayısıyla testin hız testi olması durumunda, tek boyutluluk ve yerel bağımsızlık varsayımları da ihlal edilmiş olur. Testin hız testi olmama varsayımını sağlayabilmesi için bireylere maddeleri yanıtlamaları için yeterli zamanın verilmesi alınabilecek bir önlem olarak düşünülebilir.

MTK modelleri. Bir yetenek ölçülürken ilk olarak bu yeteneğin/becerinin ölçülmesine yönelik belli sayıda madde içeren bir test geliştirilmesi gereklidir. Bu testteki maddelerin her biri ele alınan yeteneğin bazı yönlerini ölçmeye yöneliktir. Geliştirilen test, çoktan seçmeli, doğru-yanlış tipinde veya kısa cevaplı maddeler içeriyorsa, puanlama maddeye verilen doğru yanıtlar için 1, yanlış yanıtlar için ise 0 şeklinde iki kategorili olarak yapılabilmektedir. Bununla birlikte, her madde için çeşitli düzeylerin yer aldığı Likert tipi ölçeklerde ise maddeler için çoklu puanlama yapılabilmektedir (Hambleton ve Swaminathan, 1985). Bu çalışmada ikili puanlanan maddeler tercih edildiğinden, ikili puanlanan maddeler için sıklıkla kullanılan MTK modellerine ayrıntılı olarak değinilmiştir. Ayrı başlıklar altında yer verilen MTK modellerinden 1 PLM, 2 PLM ve 3 PLM birbirlerinden MKE'lerin formülasyonu bakımından farklılaşmaktadırlar.

1 parametrelilik lojistik model (1 PLM). Yetenek düzeyi θ_j ile verilen bir j bireyinin i maddesini doğru yanıtlama olasılığını ifade eden 1 PLM için MKE'nin formülü Eşitlik 1'deki gibidir (Hambleton ve Swaminathan, 1985).

$$P_i(X = 1|\theta_j) = \frac{e^{[D\bar{a}(\theta-b_i)]}}{1+e^{[D\bar{a}(\theta-b_i)]}} \quad (1)$$

Burada \bar{a} , a madde ayırıcılık parametresinin tüm maddeler için sabit olarak aldığı değeri ifade etmektedir. b_i madde güçlük parametresi veya konum parametresi ve D ise 1.702'ye eşit bir ölçekleme sabitidir. 1PLM'de a parametresi tüm maddelerde birbirine eşit sabit bir değer olarak alınırken, b parametreleri farklılaşmaktadır. 1PLM'nin özel bir hali olan Rasch Model'de ise ayırıcılık parametresi tüm maddeler için $a = 1$ olarak sabitlenmektedir. Modelde yer alan b parametresinin teorik değer aralığı $(-\infty, \infty)$ olsa da genel olarak uygulamadaki değer aralığı $[-3, 3]$ olarak alınmaktadır. b parametresinin güçlük parametresi olarak ifade edilmesinin yanı sıra konum parametresi olarak da anılmasının nedeni, 0,50 olasılıkla maddeyi doğru yanıtlayabilecek yetenek düzeyinin konumunu göstermesidir. 1PLM'de diğer modellerden farklı olarak MKE'ler çakışmaz. Çünkü tüm maddeler için eşit ve sabit olarak alınan a parametresi MKE'nin $\theta = b$ noktasındaki eğimiyle orantılıdır. Amaç yalnızca MTK yardımıyla birey yeteneklerini kestirmekse, çalışma ve anlaşılma kolaylığı açısından az sayıda parametre içeren 1 PLM'nin kullanımı önerilmektedir.

2 parametrelili lojistik model (2 PLM). Yetenek düzeyi θ_j ile verilen bir j bireyinin i maddesini doğru yanıtlama olasılığını ifade eden 2 PLM için MKE'nin formülü Eşitlik 2'deki gibidir (Hambleton ve Swaminathan, 1985).

$$P_i(X = 1|\theta_j) = \frac{e^{[Da_i(\theta-b_i)]}}{1+e^{[Da_i(\theta-b_i)]}} \quad (2)$$

Burada a_i madde ayırıcılık parametresi, b_i madde güçlük parametresi veya konum parametresi ve D ise ölçekleme sabitidir. 2 PLM'de maddelerin b parametrelerinin yanı sıra a parametreleri de her bir madde için değişkenlik göstermektedir. Dolayısıyla belirlenen test performansı üzerinde a parametresinin de etkisi vardır ve madde ayırıcılığı arttıkça MKE sivrileşir. Şöyle ki, bir testte aynı sayıda farklı maddeyi doğru yanıtlayan bireylerden ayırıcılığı yüksek maddeleri yanıtlayan bireyin yetenek düzeyi daha yüksek kestirilir. Modelde yer alan a parametresinin teorik değer aralığı $(-\infty, \infty)$ olsa da genel olarak uygulamadaki değer aralığı $[0, 2]$ dir.

3 parametrelili lojistik model (3 PLM). Yetenek düzeyi θ_j ile verilen bir j bireyinin i maddesini doğru yanıtlama olasılığını ifade eden 3 PLM için MKE'nin formülü Eşitlik 3'teki gibidir (Hambleton ve Swaminathan, 1985).

$$P_i(X = 1|\theta_j) = c_i + (1 - c_i) \frac{e^{[Da_i(\theta-b_i)]}}{1+e^{[Da_i(\theta-b_i)]}} \quad (3)$$

Burada a_i madde ayırıcılık parametresi, b_i madde güçlük parametresi veya konum parametresi, c_i düşük asimptot veya tahminle yanıtlama olasılığını gösteren şans parametresidir ve D ise ölçekleme sabitidir. Diğer iki modelden farklı olarak 3 PLM'de modele c şans parametresi eklenmiştir. 3 PLM'de MKE'nin düşük asimptotu olan c parametresi, şansla veya tahminle doğru yanıtlanma olasılığının bulunduğu bir durumda, en düşük yetenek düzeyinde maddenin doğru yanıtlanma olasılığına eşittir. Modelde yer alan c parametresinin teorik değer aralığı $[0, 1]$ iken uygulamada kullanılan genel bir değer aralığı tanımlanmamıştır. Ancak dört seçeneikli bir maddenin şansla veya tahminle doğru yanıtlanma olasılığının 0,25 oluşu dikkate alınarak c parametresiyle ilgili yorumlar yapılabilmektedir. Eğer bir sınırlama getirme ihtiyacı duyulursa, 0,35'in üzerindeki c değerlerinin kabul edilmediği düşünüldüğünden c parametresi $[0, 0,35]$ aralığında tanımlanabilir (Baker, 2001). Diğer modellerden farklı olarak 3 PLM'de b parametresi, 0,50 olasılıkla maddeyi doğru yanıtlayabilecek yetenek düzeyini değil, c parametresindeki artışa bağlı olarak bu düzeyin üstüne denk gelen, $\frac{1+c}{2}$ olasılıkla maddeyi doğru yanıtlayabilecek yetenek düzeyini göstermektedir. İçerdiği şans parametresinden dolayı 3 PLM çoktan seçmeli testlerde sıklıkla tercih edilmektedir.

BBST ile ilgili alanyazın incelendiğinde, Lau (1996) ile Reckase'nin (1983) çalışmalarının ortak bir sonuç olarak; OTU ve OSD bakımından 3 PLM'nin 1 PLM'ye kıyasla daha iyi performans gösterdiği görülmüştür. Ayrıca Jiao ve Lau (2003) tarafından yapılan çalışmada da MTK model seçiminin BBST uygulamaları için oldukça önemli olduğu, en büyük sınıflama hatasının gerçek model 3 PLM iken diğer modellerden birinin kullanılmasından kaynaklandığı ifade edilmiştir. Bu açıklamalar dikkate alınarak mevcut çalışmada madde havuzunun türetilmesinde MTK'ya dayalı 3 PLM temel alınmıştır.

Model uyumu. MTK'nın temel varsayımları test edildikten sonra model uyumunun test edilmesi ve veriye en uygun modelin seçilmesi gerekmektedir. Bu süreçte iki kategoride yapılan puanlamalarda kullanılan MTK modellerinden 1 PLM, 2 PLM ve 3PLM'nin özelliklerinin değerlendirilmesi; parametre kestiriminin her üç modelle de yapılması ve model uyumunun incelenmesi gerekmektedir. Kullanılacak modelin gerçekten de veriye en uygun model olması, sonrasında yapılacak analizler için büyük öneme sahiptir. Model veri uyumu, modellere ait log-olabilirlik (-2LL) değerleri farkının, ki-kare (X^2) dağılımının kritik tablo değeriyle karşılaştırılması yoluyla incelenebilmektedir (Demars, 2010). Model uyumunu test etmeye yönelik olarak; örneğin hesaplanan $X^2 = -2LL_{1PLM} - (-2LL_{2PLM})$ değeri, modeldeki parametre sayısı serbestlik derecesi olarak alınıp kritik X^2 tablo değeriyle karşılaştırılır. Hesaplanan değer kritik değerden büyük olması verinin 2 PLM'ye daha uygun olduğunun bir göstergesidir. Ardından aynı işlem 2 PLM ve 3 PLM için gerçekleştirilir ve $X^2 = -2LL_{2PLM} - (-2LL_{3PLM})$ hesaplanır. Bu değer yine modele eklenen parametre sayısını serbestlik derecesi kabul eden kritik X^2 tablo değeriyle karşılaştırılır. Hesaplanan değer, kritik değerden büyükse verinin 3 PLM'ye daha uygun olduğuna karar verilir.

Değişmezlik özelliği. MTK'nın belki de en büyük avantajı olduğunu söyleyebileceğimiz değişmezlik özelliği; madde parametrelerinin gruptan, birey parametrelerinin ise maddelerden bağımsızlığını ifade etmektedir (Hambleton ve Swaminathan, 1985). Bu temel ilke, iki şartın karşılanmasına dayanır. Bu şartlardan ilki, testteki tüm maddelerin aynı örtük özelliği ölçmeye yönelik olması, ikincisi ise tüm maddelerin parametrelerinin aynı ortak ölçek üzerinde yer almasıdır (Baker, 2001). MTK'nın değişmezlik özelliği sayesinde, BBT ve BBST uygulamalarının doğası gereği farklı maddelerle test edilen bireyleri karşılaştırabilmek mümkündür. Çünkü bireyin yeteneği uygulanan maddelerden bağımsız olarak kestirilmektedir (Hambleton ve Swaminathan, 1985). Madde parametrelerinin değişmezliğine göre madde parametreleri, maddeyi yanıtlayan gruba bağlı değil maddeye özgü özelliklerdir. Birey parametrelerinin değişmezliğine göre ise bireyin yetenek düzeyi, yeteneğin belirlenmesinde kullanılan maddelerden bağımsız olarak belirlenir (Baker, 2001).

Madde havuzu. Kalibre edilmiş madde havuzu BBST'nin bir diğer bileşenidir. Thompson'a (2007b) göre madde havuzunun özellikleri diğer

bileşenlere bağlı olup kalibrasyon sürecinde madde havuzunun yapısı ve tercih edilen psikometrik model oldukça önemlidir. Madde parametreleri için istenen ranj ise madde seçme algoritmasıyla ilişkilidir. Örneğin madde seçimi, güçlük düzeyi kesme noktasına yakın olan maddelerden yapılıyorsa, güçlük düzeyi bu bölgede olan çok sayıda maddeye ihtiyaç duyulacaktır. Eğer madde seçim algoritması, güçlük düzeyi bireyin kestirilen yeteneğine uygun maddenin seçimiye, bireylerin yetenekleri geniş bir aralıkta dağılabileceğinden güçlük parametresi ranjı yetenek ölçeği boyunca yayılmalıdır (Thompson, 2007b). Madde havuzu oluşturulurken belirlenmesi gereken temel özelliklerden biri madde havuzunun genişliğidir. Yüksek riskli ve sadece çok küçük bir sınıflama hatasını tolere edebilen testlerde, düşük riskli testlere kıyasla daha fazla maddeye ihtiyaç duyulmaktadır. Buna karşılık, psikometrik model olarak MTK kullanılacaksa ve maddeler nispeten yüksek ayrıcalığa sahip olup yüksek miktarda da bilgi sağlıyorsa, madde havuzu oluşturulurken daha az maddeye ihtiyaç duyulacaktır (Thompson, 2007b). Bu çalışmada yapılan sınıflamalarda, hem kesme noktasında (KN) hem de kestirilen yetenekte (KY) Maksimum Fisher Bilgisi'ne (MFB) dayanan madde seçme yöntemleri ele alındığından madde havuzu, kesme noktası ve etrafında yüksek bilgi veren ve belirlenen yetenek düzeyi aralığını kapsayan maddelerden oluşturulmuştur.

Başlama noktası. BBST uygulamalarında başlama noktası olarak bireyler hakkında mevcut olan ön bilgiler kullanılabilir (Weiss ve Kingsbury, 1984; Yang ve diğerleri, 2006). Örneğin test periyodik olarak tekrarlanıyorsa, testi daha önceden almış olan bireyler için başlama noktası, önceki testten kestirilen yetenek düzeyleri olarak belirlenebilir. Çok sık kullanılmamakla birlikte, popülasyon ortalaması da başlama noktası olarak belirlenebilir (Thompson, 2007b). Bu çalışmada tüm koşullar için başlama noktası $\theta = 0$ olarak belirlenmiştir.

Madde seçimi. Genel olarak, çok zor veya çok kolay olan veya az bilgi sağlayan maddeler, iyi çalışmayan testlerle sonuçlanacağından doğru madde seçimi oldukça önemlidir (Reckase, 1983). Alanyazın incelendiğinde BBST çalışmalarının büyük çoğunluğunun madde seçimine odaklandığı görülmektedir (Eggen, 1999; Lin ve Spray, 2000; Spray ve Reckase, 1994). BBST'nin en basit madde seçme algoritması, maddelerin havuzdan tesadüfi olarak seçildiği rastgele madde seçimidir (Kingsbury ve Weiss, 1983). Ancak ne maddeler ne de bireylerle

ilgili bilgileri hiçbir şekilde kullanmayan bu yöntem test etkililiğinin düşmesine neden olmaktadır. Bu nedenle, bilgisayarın havuzdaki kullanılmamış maddeleri değerlendirdiği ve sonraki madde olarak uygulanmasına karar verdiği "en iyi" maddenin seçildiği akıllı madde seçimi (intelligent item selection) daha uygun bir yaklaşım olarak görülmektedir. Akıllı madde seçme yöntemleri genel olarak; kesme puanı temelli ve kestirim temelli olmak üzere iki türde sınıflandırılmaktadır (Thompson, 2007b). Kesme puanı temelli yöntemlerde madde seçimi, kesme puanında maksimum bilgiyi veren maddeye göre yapılırken, kestirim temelli yöntemlerde ise kesme puanından bağımsız olarak, bireyin kestirilen geçici yetenek düzeyi için en uygun madde seçilmektedir.

Kullanılan psikometrik model KTK olduğunda, Maksimum Ayırıcılık, Minimum Beklenen Değer ve Bilgi Kazanımı olmak üzere üç farklı kesme puanı temelli yöntem önerilmiştir (Rudner, 2002). Buna karşılık kullanılan psikometrik model MTK olduğunda tercih edilebilecek kesme puanı temelli yöntemler ise Maksimum Fisher Bilgisi (MFB), Maksimum Kullback Liebler Bilgisi (KLB) ve Log-Odds Oranıdır (Lin ve Spray, 2000). Kavramsal olarak benzer olan bu üç yöntem için; MFB'de kesme puanında maksimum bilgiyi veren maddenin, KLB'de kesme puanı çevresindeki küçük bir bölgede maksimum bilgiyi veren madde ve Log-Odds Oranı yönteminde kesme puanının her iki tarafındaki olasılık değerleri arasında en büyük farklılıkları üreten madde olduğu söylenebilir. Dolayısıyla bu üç yöntem karşılaştırılabilir (Lin ve Spray, 2000). MFB yöntemi, doğru yanıt olasılığı P ve yanlış yanıt olasılığı θ olmak üzere tek bir noktadaki bilgiyi en üst düzeye çıkarmayı (maksimize etmeyi) amaçlar ve Eşitlik 4 ile ifade edilir (Embretson ve Reise, 2000).

$$I_i(\theta) = \left[\frac{\partial P_i(\theta)}{\partial \theta} \right]^2 / P_i(\theta) \theta_i(\theta) \quad (4)$$

Bu araştırmada madde seçimi için tercih edilen MFB, kesme puanı temelinde olduğu gibi kestirim temelli olarak da kullanılabilir (Eggen, 1999; Reckase, 1983; Spray ve Reckase, 1994). Hesaplamalarda kullanılan eşitlikler aynı kalır, ancak hesaplamalar testteki her bir noktada bireyin kestirilen geçici yetenek düzeyine göre yapılır (Thompson, 2007b). Kestirim temelli madde seçimi aynı zamanda bireyselleştirilmiş madde seçimi olarak da adlandırılabilir, çünkü testi bireyselleştirmek için bireysel olarak öğrenciye ait bilgiyi yani yanıt vektörünü

kullanır. Kesme puanı temelli madde seçimi ise bazen “ardışık seçim (sequential selection)” olarak adlandırılabilir (Thompson, 2007b).

Kesme puanı ve kestirim temelli madde seçim yöntemlerine ek olarak, mutual (ortak) madde seçimi, madde bilgisini θ yetenek düzeyi ranjında değerlendiren bir yöntemdir. Mutual (ortak) madde seçim yöntemi, daha geniş bir yetenek ranjında bilgiye ihtiyaç duyulduğunda, teste başlarken bireyin yeteneğiyle ilgili çok az bilgi sahibi olduğunda veya hiç bilgi sahibi olunmadığında ve birkaç kesme puanı olduğunda oldukça kullanışlıdır. Buna karşılık, istenen bilginin tek bir kesme puanı gibi tam bir komşuluğunun belirlendiği durumlarda uygulanabilir değildir (Weissman, 2006). Bu çalışmada iki, üç veya dört kategoride yapılan sınıflama karşılaştırmaları için bir, iki ve üç kesme noktası belirlendiğinden MFB-kestirilen yetenek (MFB-KY) ve MFB-kesme noktası (MFB-KN) madde seçme yöntemleri tercih edilmiştir.

Yetenek kestirimi. MTK’da bireylere uygulanan testlerin temel amacı, bireylerin yetenek ölçeği üzerindeki yerlerinin belirlenmesidir. Bu amaç gerçekleştirildiğinde: (i) testi alan bireyin yetenek düzeyiyle ilgili bir değerlendirme yapılabilir, (ii) Bireyle ilgili bir karar verilebilir (bir sınıfa atama, bursla ödüllendirme gibi) (Baker, 2001). Bireyleri iki veya daha fazla kategoriden birinde sınıflandırmak amacıyla yapılan BBST uygulamalarında ise kullanılan yetenek kestirim yöntemi, alınacak sınıflama kararlarının etkililiği ve uygunluğu bakımından oldukça önemlidir (Yang ve diğerleri, 2006). BBST’lerde yetenek kestirimi ne kadar doğru yapılırsa, bireye uygun madde seçiminin de o kadar doğru yapılabileceği söylenebilir. Böylece BBST’nin temel amacı gerçekleştirilip, olabildiğince az sayıda maddeyle yüksek oranda doğru sınıflamalar yapılabilir (Thompson, 2009).

Alanyazın incelendiğinde, ikili puanlanan (1-0) ve tek boyutlu MTK’ya dayanan farklı yetenek kestirim yöntemlerinin bulunduğu görülmektedir. Bu yöntemlerden en temel kabul edilen ve sıklıkla rastlananlar; Maksimum Olabilirlik Kestirim yöntemi (MOK; Birnbaum, 1968), Marjinal Maksimum Olabilirlik Kestirim yöntemi (MMOK; Bock ve Aitkin, 1981), Ağırlıklandırılmış Olabilirlik Kestirim yöntemi (AOK; Warm, 1989) ve Bayesci yetenek kestirim yöntemlerinden; Owen’ın Ardışık Bayesci kestirim yöntemi (Owen, 1975), Maksimum Sonsal Dağılım yöntemi (MSD; Samejima, 1969) ile Beklenen Sonsal Dağılım yöntemidir (BSD; Bock ve Aitkin, 1981). Bu yöntemlerin çeşitli dezavantajlarından da bahsedilebilir.

Örneğin Warm (1989), tüm bu yöntemlerin bir miktar yanlış kestirimler yapabildiğini ifade etmiştir. Yanlılık, sistematik bir şekilde sınıflama kararlarının doğruluğunu etkilemektedir (Wang ve Wang, 2001). Bunun dışında, yetenek kestiriminde en sık kullanılan yöntemlerden olan (Nydick ve diğerleri, 2012) ve olabilirlik fonksiyonunu temel alan MOK ile tamamen doğru veya tamamen yanlış cevap örüntüsü için hesaplama yapılamamaktadır. MOK bu yönüyle zayıf bir yetenek kestirim yöntemi olarak görülmektedir. Bu noktada MOK'a alternatif olarak Bayeşçi yetenek kestirim yöntemleri önerilmiştir (Embretson ve Reise, 2000). MC simülasyonuna dayanan bu çalışmada önsel bilgiye sahip olunmadığından BSD yöntemi tercih edilmemiştir. Bununla birlikte, Warm (1989) yaptığı çalışmada, özellikle sabit uzunluklu testlerde, AOK yöntemiyle yapılan kestirimlerin, MOK ve MSD yöntemleriyle yapılan kestirime kıyasla daha az yanlılığa sahip olduğu sonucuna ulaşmıştır. Değişken uzunluktaki bireyselleştirilmiş testlerde ise testin MSD ile benzer ama MOK'tan daha az sayıda maddeyle sonlandığını ifade ederek yetenek kestirimi için MOK'un modifiye edilmiş bir hali olan AOK yöntemini önermiştir. Madde kullanım sıklığını düşürebilecek ve test süresini azaltarak testin kullanılabilirliğini artıracak bu durum BBT ve BBST uygulamaları için AOK yönteminin kullanımına yönelik bir avantaj olarak değerlendirilebilir. Ayrıca Nydick ve diğerlerine (2012) göre, AOK yetenek kestirim yöntemi sınıflama çalışmaları için önemli bir yere sahiptir. Bu bilgiler ışığında mevcut çalışmada yetenek kestiriminde AOK yönteminin kullanılması tercih edilmiştir.

Ağırlıklandırılmış olabilirlik kestirim yöntemi (AOK). AOK yöntemi, MOK'un bir dezavantajı olarak düşünülebilecek yanlılığı azaltmak amacıyla Warm (1989) tarafından önerilmiştir. Olabilirlik fonksiyonunun moduna dayalı olarak çalışan MOK, yetenek ranjı üzerinden olabilirlik fonksiyonunu maksimize etmektedir. Buna karşılık AOK, yanlılığı azaltan ve madde parametreleri ile yetenek düzeyine özgü bir ağırlıklandırma fonksiyonu temelinde çalışmaktadır (Warm, 1989). Alanyazındaki BBST uygulamaları incelendiğinde, yetenek kestirim yöntemi olarak çoğunlukla AOK'un tercih edildiği görülmektedir (Eggen ve Straetmans, 2000; Gündeğer, 2017; Nydick ve diğerleri, 2012; Wouda ve Eggen, 2009; Yang ve diğerleri, 2006). Sağladığı avantajlar ve sınıflama literatüründeki yeri de göz önüne alınarak, bu çalışmada yetenek kestirim yöntemi olarak AOK ele alınmıştır.

Sonlandırma (sınıflama) kriteri. BBST'nin son bileşeni sonlandırma (sınıflama) kriteridir. BBST'de madde seçme yöntemleri ve sınıflama kriterleri, test uzunluğu ve sınıflama doğruluğu üzerinde etkili iki önemli bileşendir (Lin ve Spray, 2000). BBST'de MTK'ya dayalı; Ardışık Olasılık Oran Testi (AOOT; Reckase, 1983), Güven Aralığı (GA; Kingsbury ve Weiss, 1983) ve Bayeşçi Karar Kuramı (BKK; Lewis ve Sheenan, 1990; Sheenan ve Lewis, 1992) olmak üzere üç temel sınıflama kriterinden söz edilebilmektedir. Her üç sınıflama kriteri de sabit formulu geleneksel testlerden daha az madde gerektirmekle birlikte benzer düzeyde sınıflama doğruluğu sağlamaktadır (Kingsbury ve Weiss, 1983; Rudner, 2002). Bununla birlikte, seçilen kriterlerin uygunluğu ve etkili kullanımı, tercih edilen psikometrik model ve madde seçme algoritması gibi diğer birkaç faktöre bağlıdır (Thompson, 2007b). Örneğin MTK temelli GA kriteri, büyük bir örneklem ve MTK yaklaşımına göre kalibre edilmiş geniş bir madde havuzu gerektirmektedir.

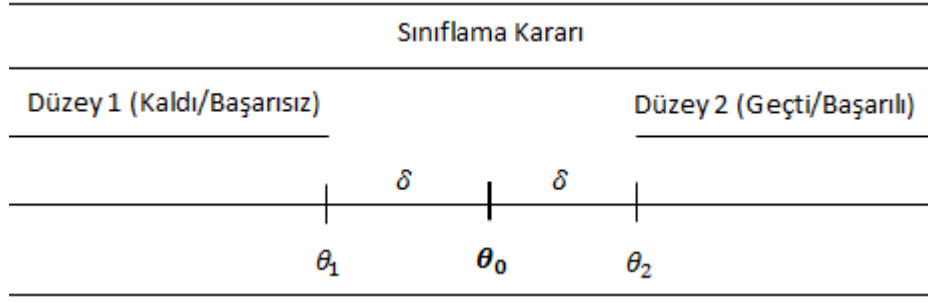
Yapılan araştırmalar, madde seçimi KY temelinde iken GA yönteminin daha etkili olduğunu, madde seçimi KN temelinde iken ise AOOT ve GOO yöntemlerinin daha etkili olduğunu göstermektedir (Eggen ve Straetmans, 2000; Spray ve Reckase, 1996; Thompson, 2009). Bunun yanı sıra alanyazındaki BBST uygulamalarında AOOT yönteminin özellikle sınıflama doğruluğu bakımından, yetenek kestirimleri etrafındaki GA yönteminden daha etkili olduğunu gösteren çalışmalar da bulunmaktadır (Eggen ve Straetmans, 2000). Buna karşılık Gündeğer (2017) yaptığı çalışmada, OTU bakımından GA yönteminin AOOT'den daha iyi performans gösterdiği ve OSD bakımından ise benzer performansa sahip oldukları sonucuna ulaşmıştır. Bartroff, Finkelman ve Lai (2008) ile Thompson (2009) tarafından yapılan çalışmalarda AOOT'nin genişletilmiş hali olduğunu söyleyebileceğimiz Genelleştirilmiş Olabilirlik Oranı (GOO) sınıflama kriterinin kullanımı da önerilmiştir. Ancak Thompson'ın (2009) da dikkat çektiği gibi alanyazındaki BBST uygulamalarında en sık kullanılan sınıflama yönteminin AOOT olduğu görülmektedir. Yapılan açıklamalar ışığında, bu çalışmanın odağındaki sınıflama kriterleri AOOT ve GA olarak belirlendiğinden aşağıda bu iki kritere ayrıntılı olarak yer verilmiştir.

Ardışık olasılık oran testi (AOOT) sonlandırma kriteri. AOOT'nin geçmişten günümüze kullanımına bakıldığında ilk kez Fergusson (1969) tarafından ve KTK temelindeki sınıflamalarda kullanıldığı, sonrasında ise Reckase

(1983) tarafından MTK temelinde kullanıldığı görülmektedir. Bununla birlikte çok aşamalı testlerde de (multi-stage testing) kullanıldığını gösteren çalışmalara (Mead, 2006; Zenisky, Hambleton ve Luecht, 2010) rastlanmaktadır (aktaran, van Groen, Eggen ve Veldkamp, 2014).

BBST uygulamaları MTK temelinde ele alındığında, sınıflama kriterleri; Olabilirlik Oranı (OO) ve Güven Aralıkları (GA) olmak üzere iki paradigmayla ele alınmaktadır (Thompson, 2011). Eggen (1999) tarafından “istatistiksel sınıflama” ve “istatistiksel kestirim” olarak adlandırılan bu paradigmaların her ikisi de bireyin yeteneğinin olasılık fonksiyonunu kullanmaktadır. OO, Reckase (1983) tarafından bir hipotez testi olarak Ardışık Olasılık Oran Testi (AOOT) adıyla formüle edilmiştir. Bir testin amacı, bireylerin yetenek veya beceri düzeylerini kestirmek yerine, bireyleri iki veya daha fazla kategoriden birine sınıflandırmak ise, BBST prosedürü, AOOT gibi istatistiksel hipotez testine dayanan algoritmalarla en uygun maddeleri seçip uygulamak suretiyle etkili sınıflama kararları vermek için uygulanabilmektedir (Lin ve Spray, 2000; Spray ve Reckase, 1994, 1996). AOOT, sınıflama kuralını hipotez testi olarak formüle etmekte (Eggen ve Straetmans, 2000) ve iki alternatif hipotezin (tek kesme puanı için) olasılık oranlarını karşılaştırmaktadır. Bireylerin gözlenen yanıt dağılımının olabilirliğine göre iki hipotezden birinin doğruluğuna karar verilmektedir. Eğer iki hipotezin olabilirlikleri benzerse bireye yeni bir madde uygulanır, aksi halde olabilirliği diğerinden belirgin olarak büyük olan hipotez kabul edilir ve süreç bu şekilde iki hipotezden biri kabul edilene kadar devam eder (Reckase, 1983). AOOT, dayanağı olan hipotez testleriyle, belirli bir i bireyinin θ_i yetenek düzeyinin, kesme noktasının (θ_0) altındaki (θ_1) veya üstündeki (θ_2) bir sabit değere eşitliğini test ederek sınıflama kararını vermeye çalışır. Bu iki sabit değer (θ_1 ve θ_2) arasındaki aralık farksızlık bölgesi (FB; indifference region) olarak adlandırılır. AOOT ile geleneksel olarak, kesme noktasında Fisher bilgisini en üst düzeye çıkaran madde seçme yöntemi kullanılmaktadır. Özellikle de farksızlık bölgesi genişledikçe, AOOT'nin daha iyi sonuçlar üretmesi beklenmektedir (Eggen, 1999).

Uygulanan test bireyleri geçti-kaldı gibi iki kategoriden birinde sınıflama amacı taşıyorsa, tek bir kesme noktası (θ_0) kullanılır ve birey hakkında verilecek sınıflama kararı için, δ farksızlık bölgesi sabiti olmak üzere, Şekil 1'de şematize edilmiş olan hipotezler Eşitlik 5 ve 6'daki gibi kurulabilir.



Şekil 1. İki kategoride yapılacak sınıflamalar için sınıflama kararı.

$$H_0: \theta_i \leq \theta_0 - \delta = \theta_1 \quad (\text{Düzyey 1}) \quad (5)$$

$$H_1: \theta_i \geq \theta_0 + \delta = \theta_2 \quad (\text{Düzyey 2}) \quad (6)$$

Eşitlik 5 ve 6'da görülen hipotez testlerine göre, FB'nin alt sınırı $\theta_1 = \theta_0 - \delta$ ve üst sınırı ise $\theta_2 = \theta_0 + \delta$ olarak elde edilmektedir. $\theta_2 - \theta_1 = 2\delta$ aralığının genişliği ise θ_0 ' a yakın sınıflama kararları için tolere edilebilen belirsizlik düzeyini belirten FB'ye eşittir. O halde, farksızlık bölgesi sabiti δ , FB değerinin yarısına eşittir.

AOOT'de kullanılan test istatistiğinin karşılaştırdığı, null hipotezi (H_0) ve alternatif hipotez (H_1) altındaki olasılık fonksiyonlarının oranı, ikili puanlanan maddeler için Eşitlik 7'deki gibi formüle edilmektedir.

$$LR(\underline{x}) = \frac{L(\theta_2; \underline{x})}{L(\theta_1; \underline{x})} = \frac{\prod_{i=1}^k L(\theta_2; x_i)}{\prod_{i=1}^k L(\theta_1; x_i)} = \frac{\prod_{i=1}^k P_i(\theta_2)^{x_i} [1 - P_i(\theta_2)]^{1-x_i}}{\prod_{i=1}^k P_i(\theta_1)^{x_i} [1 - P_i(\theta_1)]^{1-x_i}} \quad (7)$$

Eşitlik 7'deki formülde $L(\theta_2; \underline{x})$ ikili puanlanan bir madde için k tane gözlemin ($\underline{x} = x_1, x_1, \dots, x_i, \dots, x_k$) θ_2 ile verilen gözlenme olasılığını ifade ederken, $L(\theta_1; \underline{x})$ ise benzer şekilde θ_1 ile verilen gözlenme olasılığını ifade etmektedir. $P_i(\theta_1)$ ve $P_i(\theta_2)$, θ_1 ve θ_2 şartlarına bağlı olarak i maddesinin doğru yanıtlanma olasılığını belirtmektedir. Bir testin başlangıcında, $LR(x)$ 1'e eşittir (Lin, 2011).

α ve β ile verilen Tip I ve Tip II hata oranlarına göre bireylerin hangi kategoride sınıflanacağına karar verme kuralı aşağıdaki gibi uygulanmaktadır (Wald, 1947).

$$A = \beta / (1 - \alpha) \quad (8)$$

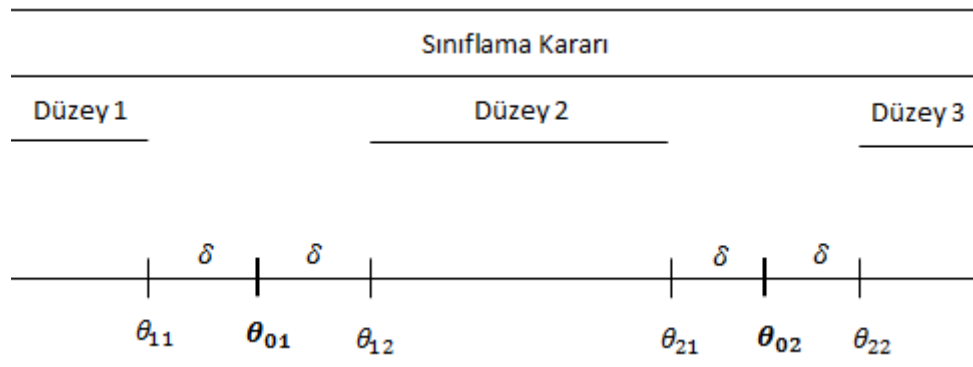
$$B = (1 - \beta) / \alpha \quad (9)$$

A ve B deęerleri Eşitlik 8 ve 9'daki gibi tanımlanmak üzere,

- $\log(A) < LR(x) < \log(B)$ ise başka bir madde seçilir, test devam eder.
- $LR(x) \leq \log(A)$ ise H_0 kabul edilir ve birey başarısız olarak sınıflandırılır, test sonlanır.
- $LR(x) \geq \log(B)$ ise H_0 reddedilir ve birey başarılı olarak sınıflandırılır, test sonlanır.

Uygulamada, genellikle minimum ve maksimum test uzunlukları belirlenmektedir. Maksimum test uzunluęuna ulaşıldığında, henüz bir karar verilememişse, $(\log(A), \log(B))$ aralığının orta noktasını ifade eden $(\log(A) + \log(B))/2$ kritik deęeri hesaplanır. $LR(x)$ deęeri, belirlenen kritik deęerden büyükse H_0 'ı reddetmek suretiyle zorunlu bir sınıflandırma yapılır ve birey başarılı kabul edilir; aksi takdirde H_0 kabul edilerek birey başarısız olarak sınıflandırılır (Lin, 2011). Araştırmalarda genellikle $\alpha = \beta$ şeklinde belirlendiğinden $(\log(A) + \log(B))/2 = 0$ olarak hesaplanır. Dolayısıyla bazen son hataya baęlı olarak yanlış sınıflamanın yapılması kaçınılmaz olabilir. Bu durum AOOT'nin dezavantajlarından biri olarak kabul edilebilir.

AOOT sınıflama kriteri iki kategoride yapılacak sınıflamaların yanı sıra üç veya daha fazla kategoride yapılacak sınıflamalar için de genişletilebilmektedir (Eggen ve Straetmans, 2000). Uygulanan test bireyleri üç kategoriden birinde sınıflama amacı taşıdığında, iki kesme noktası $(\theta_{01}, \theta_{02})$ kullanılır ve birey hakkında verilecek sınıflama kararı için, δ farksızlık bölgesi sabiti olmak üzere, Şekil 2'de şematize edilmiş olan hipotezler Eşitlik 10, 11, 12 ve 13'teki gibi kurulabilir (Eggen ve Straetmans, 2000).



Şekil 2. Üç kategoride yapılacak sınıflamalar için sınıflama kararı.

$$H_{0_1}: \theta_i \leq \theta_{01} - \delta = \theta_{11} \quad (\text{Düzey 1}) \quad (10)$$

$$H_{1_1}: \theta_i \geq \theta_{01} + \delta = \theta_{12} \quad (\text{Düzey 1'den yüksek}) \quad (11)$$

$$H_{0_2}: \theta_i \leq \theta_{02} - \delta = \theta_{21} \quad (\text{Düzey 3'ten düşük}) \quad (12)$$

$$H_{1_2}: \theta_i \geq \theta_{02} + \delta = \theta_{22} \quad (\text{Düzey 3}) \quad (13)$$

AOOT'de kullanılan test istatistiğinin karşılaştırdığı, null hipotezi (H_0) ve alternatif hipotez (H_1) altındaki olasılık fonksiyonlarının oranı, ikili puanlanan maddeler için Eşitlik 14 ve 15'teki gibi formüle edilmektedir.

$$LR(\theta_{12}, \theta_{11}; \underline{x}) = \frac{L(\theta_{12}; \underline{x})}{L(\theta_{11}; \underline{x})} = \frac{\prod_{i=1}^k L(\theta_{12}; x_i)}{\prod_{i=1}^k L(\theta_{11}; x_i)} = \frac{\prod_{i=1}^k P_i(\theta_{12})^{x_i} [1 - P_i(\theta_{12})]^{1-x_i}}{\prod_{i=1}^k P_i(\theta_{11})^{x_i} [1 - P_i(\theta_{11})]^{1-x_i}} \quad (14)$$

$$LR(\theta_{22}, \theta_{21}; \underline{x}) = \frac{L(\theta_{22}; \underline{x})}{L(\theta_{21}; \underline{x})} = \frac{\prod_{i=1}^k L(\theta_{22}; x_i)}{\prod_{i=1}^k L(\theta_{21}; x_i)} = \frac{\prod_{i=1}^k P_i(\theta_{22})^{x_i} [1 - P_i(\theta_{22})]^{1-x_i}}{\prod_{i=1}^k P_i(\theta_{21})^{x_i} [1 - P_i(\theta_{21})]^{1-x_i}} \quad (15)$$

H_{0_1} 'e karşı H_{1_1} testi için bireylerin hangi kategoride sınıflanacağına karar verme kuralı aşağıdaki gibi uygulanmaktadır:

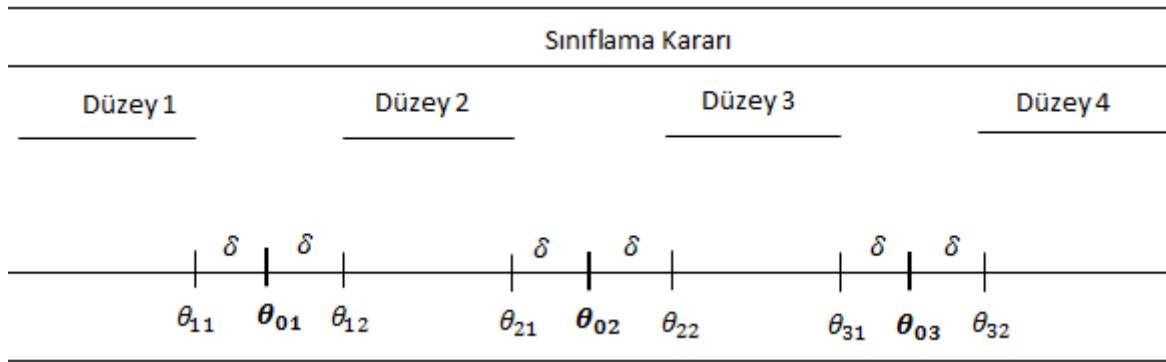
- $\log(A) < LR(\theta_{12}, \theta_{11}; \underline{x}) < \log(B)$ ise başka bir madde seçilir, test devam eder (testin kritik eşitsizliğidir).
- $LR(\theta_{12}, \theta_{11}; \underline{x}) \leq \log(A)$ ise H_{0_1} kabul edilir ve birey Düzey 1 olarak sınıflanır, test sonlanır.
- $LR(\theta_{12}, \theta_{11}; \underline{x}) \geq \log(B)$ ise H_{0_1} reddedilir ve bireyin Düzey 1'in üstünde olduğu belirlenir (Düzey 2 veya Düzey 3).

H_{0_2} 'ye karşı H_{1_2} testiyle, bireylerin hangi kategoride sınıflanacağına karar verme kuralına aşağıdaki gibi devam edilir:

- $\log(A) < LR(\theta_{22}, \theta_{21}; \underline{x}) < \log(B)$ ise başka bir madde seçilir, test devam eder (testin kritik eşitsizliğidir).
- $LR(\theta_{22}, \theta_{21}; \underline{x}) \leq \log(A)$ ise H_{0_2} kabul edilir ve birey Düzey 2 olarak sınıflanır, test sonlanır.
- $LR(\theta_{22}, \theta_{21}; \underline{x}) \geq \log(B)$ ise H_{0_2} reddedilir ve birey Düzey 3 olarak sınıflanır, test sonlanır.

Maksimum test uzunluğuna ulaşılmış ancak henüz bir karar verilememişse, bu durumda sınıflama kararı için dikkate alınacak kritik değer ise $(\log(A) + \log(B))/2$ olarak hesaplanır.

Uygulanan testle bireyleri dört kategoriden birinde sınıflamak amaçlandığında ise üç kesme noktası $(\theta_{01}, \theta_{02}, \theta_{03})$ belirlenir ve birey hakkında verilecek sınıflama kararı için, δ farksızlık bölgesi sabiti olmak üzere, Şekil 3'te şematize edilen hipotezler Eşitlik 16, 17, 18, 19, 20 ve 21'deki gibi kurulabilir.



Şekil 3. Dört kategoride yapılacak sınıflamalar için sınıflama kararı.

$$H_{0,1}: \theta_i \leq \theta_{01} - \delta = \theta_{11} \quad (\text{Düzey 1}) \quad (16)$$

$$H_{1,1}: \theta_i \geq \theta_{01} + \delta = \theta_{12} \quad (\text{Düzey 1'den yüksek}) \quad (17)$$

$$H_{0,2}: \theta_i \leq \theta_{02} - \delta = \theta_{21} \quad (\text{Düzey 3'ten düşük}) \quad (18)$$

$$H_{1,2}: \theta_i \geq \theta_{02} + \delta = \theta_{22} \quad (\text{Düzey 2'den yüksek}) \quad (19)$$

$$H_{0,3}: \theta_i \leq \theta_{03} - \delta = \theta_{31} \quad (\text{Düzey 4'ten düşük}) \quad (20)$$

$$H_{1,3}: \theta_i \geq \theta_{03} + \delta = \theta_{32} \quad (\text{Düzey 4}) \quad (21)$$

AOOT'de kullanılan test istatistiğinin karşılaştırdığı, null hipotezi (H_0) ve alternatif hipotez (H_1) altındaki olasılık fonksiyonlarının oranı, ikili puanlanan maddeler için Eşitlik 22, 23 ve 24'teki gibi formüle edilmektedir.

$$LR(\theta_{12}, \theta_{11}; \underline{x}) = \frac{L(\theta_{12}; \underline{x})}{L(\theta_{11}; \underline{x})} = \frac{\prod_{i=1}^k L(\theta_{12}; x_i)}{\prod_{i=1}^k L(\theta_{11}; x_i)} = \frac{\prod_{i=1}^k P_i(\theta_{12})^{x_i} [1 - P_i(\theta_{12})]^{1-x_i}}{\prod_{i=1}^k P_i(\theta_{11})^{x_i} [1 - P_i(\theta_{11})]^{1-x_i}} \quad (22)$$

$$LR(\theta_{22}, \theta_{21}; \underline{x}) = \frac{L(\theta_{22}; \underline{x})}{L(\theta_{21}; \underline{x})} = \frac{\prod_{i=1}^k L(\theta_{22}; x_i)}{\prod_{i=1}^k L(\theta_{21}; x_i)} = \frac{\prod_{i=1}^k P_i(\theta_{22})^{x_i} [1 - P_i(\theta_{22})]^{1-x_i}}{\prod_{i=1}^k P_i(\theta_{21})^{x_i} [1 - P_i(\theta_{21})]^{1-x_i}} \quad (23)$$

$$LR(\theta_{32}, \theta_{31}; \underline{x}) = \frac{L(\theta_{32}; \underline{x})}{L(\theta_{31}; \underline{x})} = \frac{\prod_{i=1}^k L(\theta_{32}; x_i)}{\prod_{i=1}^k L(\theta_{31}; x_i)} = \frac{\prod_{i=1}^k P_i(\theta_{32})^{x_i} [1 - P_i(\theta_{32})]^{1-x_i}}{\prod_{i=1}^k P_i(\theta_{31})^{x_i} [1 - P_i(\theta_{31})]^{1-x_i}} \quad (24)$$

H_{0_1} 'e karşı H_{1_1} testi için bireylerin hangi kategoride sınıflanacağına karar verme kuralı aşağıdaki gibi uygulanmaktadır:

- $\log(A) < LR(\theta_{12}, \theta_{11}; \underline{x}) < \log(B)$ ise başka bir madde seçilir, test devam eder (testin kritik eşitsizliğidir).
- $LR(\theta_{12}, \theta_{11}; \underline{x}) \leq \log(A)$ ise H_{0_1} kabul edilir ve birey Düzey 1 olarak sınıflanır, test sonlanır.
- $LR(\theta_{12}, \theta_{11}; \underline{x}) \geq \log(B)$ ise H_{0_1} reddedilir ve bireyin Düzey 1'in üstünde olduğu belirlenir (Düzey 2, Düzey 3 veya Düzey 4).

H_{0_2} 'ye karşı H_{1_2} testiyle, bireylerin hangi kategoride sınıflanacağına karar verme kuralına aşağıdaki gibi devam edilir:

- $\log(A) < LR(\theta_{22}, \theta_{21}; \underline{x}) < \log(B)$ ise başka bir madde seçilir, test devam eder (testin kritik eşitsizliğidir).
- $LR(\theta_{22}, \theta_{21}; \underline{x}) \leq \log(A)$ ise H_{0_2} kabul edilir ve birey Düzey 2 olarak sınıflanır, test sonlanır.
- $LR(\theta_{22}, \theta_{21}; \underline{x}) \geq \log(B)$ ise H_{0_2} reddedilir ve bireyin Düzey 2'in üstünde olduğu belirlenir (Düzey 3 veya Düzey 4).

H_{0_3} 'ye karşı H_{1_3} testiyle, bireylerin hangi kategoride sınıflanacağına karar verme kuralına aşağıdaki gibi devam edilir:

- $\log(A) < LR(\theta_{32}, \theta_{31}; \underline{x}) < \log(B)$ ise başka bir madde seçilir, test devam eder (testin kritik eşitsizliğidir).
- $LR(\theta_{32}, \theta_{31}; \underline{x}) \leq \log(A)$ ise H_{0_3} kabul edilir ve birey Düzey 3 olarak sınıflanır, test sonlanır.
- $LR(\theta_{32}, \theta_{31}; \underline{x}) \geq \log(B)$ ise H_{0_3} reddedilir ve birey Düzey 4 olarak sınıflanır, test sonlanır.

Maksimum test uzunluğuna ulaşılmış ancak henüz bir karar verilememişse, bu durumda sınıflama kararı için dikkate alınacak kritik değer ise $(\log(A) + \log(B))/2$ olarak hesaplanır.

Bu çalışmada AOOT sınıflama kriteri için; iki, üç veya dört kategoride yapılacak sınıflamalarda farksızlık bölgesi sabiti δ değerlerinin belirlenmesinde

Huebner (2012), Lin (2011) ve Thompson'ın (2011) çalışmaları göz önüne alınmıştır. Huebner (2012) çalışmasında farksızlık bölgesi sabiti δ değerini 0,20 olarak almışken, Lin (2011) 0,10, 0,20 ve 0,30 değerlerini, Thompson (2011) ise 0,20 ve 0,30 değerlerini tercih etmiştir. Bu çalışmada ise ortalama bir değer olduğu düşünülerek δ : 0,20 değeri belirlenmiştir. Bunun yanı sıra hem iki, hem üç hem de dört kategoride yapılan sınıflamalar için kesme noktalarının belirlenmesinde, bireyler için R'da üretilen yetenek parametrelerinden faydalanılmıştır. Bu amaç doğrultusunda, türetilen yetenek parametreleri, Eggen ve Straetmans'ın (2000) çalışmalarına paralel olarak, düşük yetenek düzeyinden yüksek yetenek düzeyine doğru sıralanmış ve iki kategoride yapılacak sınıflamalar için bir, üç kategoride yapılacak sınıflamalar için iki, dört kategoride yapılacak sınıflamalar içinse üç kesme noktası belirlenmiştir. İki kategoride yapılan sınıflamalarda, sıralanmış yetenek düzeyleri düzey 1 ve düzey 2 olarak kodlanmış, ardından düzey 1'deki en yüksek yetenek düzeyinin %70'i alınarak kesme noktası (KN= 0,0) belirlenmiştir. Üç kategoride yapılacak sınıflamalarda, sıralanmış yetenek düzeyleri düzey 1, düzey 2, düzey 3 şeklinde kodlanmış ve düşük kesme puanı, düzey 1'deki en yüksek yetenek düzeyinin %70'i alınarak (KN1= -0,29), yüksek kesme puanı da düzey 2'deki en yüksek yetenek düzeyinin %70'i alınarak (KN2= 0,31) belirlenmiştir. Benzer şekilde dört kategoride yapılacak sınıflamalar için de sıralanmış yetenek düzeyleri düzey 1, düzey 2, düzey 3, düzey 4 şeklinde kodlanmış ve düşük kesme puanı, düzey 1'deki en yüksek yetenek düzeyinin %70'i alınarak (KN1= -0,47), ortanca kesme puanı düzey 2'deki en yüksek yetenek düzeyinin %70'i alınarak (KN2= -0,01), ve yüksek kesme puanı da düzey 3'teki en yüksek yetenek düzeyinin %70'i alınarak (KN3= 0,48) belirlenmiştir.

Güven aralığı (GA) sonlandırma kriteri. Yetenek güven aralığı (GA) sonlandırma kriteri, bir sınıflama kararı vermek için olasılık (likelihood) fonksiyonunun kullanıldığı AOOT'ye alternatif bir yöntemidir. GA, sınıflama kuralını istatistiksel kestirim problemi olarak formüleştirmektedir (Eggen ve Straetmans, 2000). GA yöntemi başlangıçta Bayesçi (Owen, 1975) kestirim prosedürü için (Kingsbury ve Weiss, 1983; Spray ve Reckase, 1996) önerilmiş olsa da maksimum olasılık kestirimiyle de kullanılabilir (Eggen ve Straetmans, 2000; Yang ve diğerleri, 2006). Benzer şekilde, başlangıçta iki yönlü aralık olarak önerilmiş olsa da tek yönlü bir aralık olarak da tasarlanabilir (Chang, 2005). GA ile yürütülen

süreçte amaç, j bireyi için kestirilen $\hat{\theta}_j$ yetenek düzeyi için belirlenen güven aralığını kesme puanı ile karşılaştırarak bireyi uygun kategoride sınıflamaktır.

GA yönteminde, ölçmenin koşullu standart hatası (Conditional Standard Error of Measurement: CSEM) kullanılarak maksimum olabilirlik kestirimi (veya Bayesçi kestirimler) etrafında yeteneğin bir güven aralığı oluşturulur. Bu aralık Eşitlik 25'teki gibi ifade edilebilir (Hambleton ve Swaminathan, 1985; Thompson, 2007b).

$$\hat{\theta}_j - z_\alpha(CSEM) \leq \theta_j \leq \hat{\theta}_j + z_\alpha(CSEM) \quad (25)$$

Eşitlik 25'te görülen z_α , $1-\alpha$ güven aralığına karşılık gelen normal sapma değeridir. Bu çalışmada, güven aralığını belirlemede kullanılan CSEM, "catlrr" paketinde de uygulanan şekliyle, ilgili cevap deseninden bağımsız olarak değerlendirilen test bilgi fonksiyonu kullanılarak Eşitlik 26'daki gibi hesaplanmıştır (Embretson ve Reise, 2000).

$$CSEM = 1/\sqrt{TI(\theta)} \quad (26)$$

Bireyleri iki kategoride sınıflamayı amaçlayan uzmanlık (mastery) testleri bu sınıflama kriteri için basit bir örnek olarak verilebilir. Uzmanlık testlerinde, uygulanan her bir maddeden sonra, güven aralığının kesme puanının hangi tarafına düştüğü (altında veya üstünde) incelenmektedir. Belirlenen güven aralığı kesme puanının tam olarak üstündeyse birey "geçti" kategorisinde sınıflanır. Buna karşılık güven aralığı kesme puanının tam olarak altındaysa birey "kaldı" kategorisinde sınıflanır. Belirlenen güven aralığının kesme puanını içermesi durumunda ise uygulanan yeni bir maddeyle test devam eder.

Eggen ve Straetmans'ın (2000) da ifade ettiği gibi GA sınıflama kriteri iki kategoride yapılacak sınıflamaların yanı sıra üç veya daha fazla kategoride yapılacak sınıflamalar için de genişletilebilmektedir. Ayrıca Eggen ve Straetmans'ın (2000) güven aralığı değerlerini %70 ve %90 olarak belirledikleri çalışma sonucunda, güven aralığı değeri büyüdükçe test bir miktar uzasa da sınıflama doğruluğunun arttığı görülmüştür. Bu çalışmada da Eggen ve Straetmans'ın (2000) çalışmaları dikkate alınarak; GA yöntemiyle iki, üç veya dört kategoride yapılan sınıflamalar için güven aralığı değeri %90 olarak belirlenmiştir.

Pratik Kısıtlamalar

BBST'nin temel bileşenlerine ek olarak, zorunlu olmamakla birlikte temel bir unsur, bilgisayar algoritmalarında pratik kısıtlamalar oluşturmaktadır. Çoğu durumda, en bilgilendirici maddelerin kullanım sıklığının fazla olmaması amacıyla madde kullanım sıklığı üzerinde kısıtlamalar yapılır (Kalohn ve Spray, 1998). Bunun yanı sıra, sıklıkla karşılaşılan bir durum olarak test maddelerinin belirli içerik alanlarına dağıtıldığı içerik dengeleme kontrolünün de BBST'lerde dikkate alınması oldukça önemlidir (Eggen ve Straetmans, 2000). BBST'nin yalnızca birkaç madde uygulamasıyla karar verebileceği ya da belirli yetenek düzeylerindeki bireyler için süresiz devam edebileceği durumlarda testlerin uzunluğuna getirilen kısıtlamalar çok kısa ve çok uzun testleri engellemek amacıyla kullanılabilir (Parshall ve diğerleri, 2006). Bu sayede minimum test uzunluğu ve maksimum test uzunluğu tüm bireyler için ortalama test uzunluğuna yaklaşacaktır. Eggen ve Straetmans'a (2000) göre bu pratik kısıtlamalar önemli amaçlara hizmet etmekle birlikte, çok büyük psikometrik etkileri yoktur. Thompson'a (2007) göre ise madde kullanım sıklığı ve içerik dengeleme kısıtlamaları genel olarak yalnızca ortalama test uzunluğunda artışa neden olmaktadır.

Son yıllarda yapılan BBT çalışmalarında, içerik dengeleme ve madde kullanım sıklığı kontrolünün istatistiksel olmayan kısıtlamalar olarak eşzamanlı ele alındığı (örneğin, Cheng, Chang, Douglas ve Guo, 2009) görülmektedir (Lin, 2011). Ancak BBST uygulamalarındaki bu eksiklik, içerik dengeleme ve madde kullanım sıklığı kontrolüyle ilgili konuları ele alma ihtiyacı doğurmaktadır. Bu amaçla, mevcut çalışmada, çeşitli içerik dengeleme ve madde kullanım sıklığı kontrol yöntemlerinin araştırma koşullarına dahil edilmesi durumunda iki, üç veya dört kategoride yapılan sınıflamalar için farklı sınıflama kriterleri ve madde seçme yöntemlerinin etkililiği, gelecekteki araştırmalar için bir temel oluşturmak üzere değerlendirilmiştir.

İçerik Dengeleme

BBT literatüründe genel olarak "içerik dengeleme" olarak bilinen çeşitli yaklaşımlar, istatistiksel olmayan kısıtlamalarla etkili ve doğru bir BBT uygulaması gerçekleştirebilmek için önerilmiştir. Belirlenen sabit bir içerik dengesini (fixed content balancing) sağlamaya yönelik yöntemler arasında, Kısıtlanmış BBT

(KBBT; Constrained CAT: CCAT; Kingsbury ve Zara, 1989) yöntemi, Modifiye Edilmiş KBBT (MKBBT; Modified CCAT: MCCAT; Leung, Chang ve Hau, 2003) yöntemi ve Modifiye Edilmiş Multinomial Model (MMM; Modified Multinomial Model: MMM; Chen ve Ankenmann, 2004) bulunmaktadır (Lin, 2011). Bunların yanı sıra, büyük ölçekli BBT uygulamaları için önemli bir alternatif olarak (Stocking ve Swanson, 1993) geliştirilmiş olan bazı etkili yöntemlerle esnek içerik dengeleme (flexible content balancing) de yapılabilmektedir (Lin, 2011). Bu yöntemler arasında, Ağırlıklandırılmış Sapmalar Modeli (ASM; Weighted Deviations Model: WDM; Stocking ve Swanson, 1993), Gölge Testi Yaklaşımı (GTY; Shadow Test Approach: STA; van der Linden ve Reese, 1998), iki aşamalı madde seçme prosedürü (two-phase item selection procedure; Cheng, Chang, ve Yi, 2007), Ağırlıklandırılmış Ceza Modeli (ACM; Weighted Penalty Model: WPM; Shin, Chien, Way, ve Swanson, 2009) ve Maksimum Önsel İndeks yöntemi (MÖİ; Maximum Priority Index: MPI; Cheng ve Chang, 2009) sayılabilir (aktaran, He, Diao ve Hauser, 2014).

Bütün bu içerik dengeleme yöntemlerinin içinde, KBBT, MKBBT ve MMM ortak bir metodolojik yaklaşıma dayanmaktadır. Bu yaklaşımda, bir madde havuzu, madde öznitelikleri doğrultusunda birkaç alt havuza bölünmüştür ve maddeler önceden belirlenmiş içerik alanlarını karşılamak üzere alt havuzlardan seçilir. Genel olarak, bu yöntemlerde, seçilen bir madde tek bir kısıtlamanın sağlanmasına katkıda bulunur ve bu nedenle madde seçiminde tek kısıtlamanın göz önüne alındığı BBT'lerle birlikte kullanılmaları daha uygun ve etkilidir. Buna karşılık, GTY, ASM, ACM ve MÖİ yöntemleri, karmaşık bir kısıtlama kümesini karşılamaya çalışarak madde seçimi yapan, farklı bir yaklaşıma dayanır ve bir madde aynı anda birden fazla kısıtlamanın sağlanmasına katkıda bulunabilir (He ve diğerleri, 2014). He ve diğerleri (2014) tarafından yapılan çalışmada, karmaşık bir kısıtlama kümesi ele alındığından bu dört yöntem üzerine odaklanılmış ve hem ölçme kesinliği bakımından hem de istenen kısıtlamaların sağlanması bakımından en iyi performansı GTY yönteminin gösterdiği, bu yöntemi ACM yönteminin izlediği ve en düşük performansı birbirine benzer değerler veren ASM ve MÖİ yöntemlerinin gösterdiği ifade edilmiştir. Ayrıca, Robin, van der Linden, Eignor, Steffen ve Stocking (2005) ile van der Linden'in (2005) ASM ve GTY yöntemlerini, Cheng ve Chang'ın (2009) MÖİ ve ASM yöntemlerini, Shin, Chien ve Way'in

(2012) KBBT, ACM ve ASM yöntemlerini karşılaştırdıkları çalışmaların da BBT literatüründe yerini aldığı görülmektedir. Bunların yanı sıra alanyazında, içerik dengelemesi yapılmış BBST çalışmalarında en sık kullanılan içerik dengeleme yöntemlerinin “spiral yöntem” (spiraling method; Kingsbury ve Zara, 1989) (örneğin, Finkelman, 2008; Huebner, 2012) ve KBBT yöntemi (örneğin, Eggen ve Straetmans, 2000; Huebner ve Li, 2012) olduğu görülmektedir. Lin (2011) ise yaptığı çalışmada içerik dengeleme yöntemi olarak MMM’i kullanmıştır. Ayrıca Lin (2011) BBT uygulamalarında, MMM yönteminin KBBT yöntemindeki içerik sıralamasının yüksek öngörülebilirliğini engellediğini ifade etmiştir. Bununla birlikte alanyazında, KBBT ve MMM yöntemlerini karşılaştıran bir BBST çalışmasına rastlanmamıştır. Bu nedenle bu çalışmada alanyazındaki çalışmalardan farklı olarak; KBBT ve MMM olmak üzere iki farklı içerik dengeleme yöntemi ele alınmıştır. Ele alınan yöntemlerin araştırma koşullarına dahil edildiği durumlarda, 500 maddeden oluşan madde havuzu rastgele madde atamasıyla dört içerik alanına ayrılmıştır. Belirlenen içerik alanlarının istenen oranları ise sırasıyla %40, %30, %20 ve %10 olacak şekilde ayarlanmıştır. Aşağıda KBBT ve MMM yöntemlerinin mantıksal altyapılarına ayrıntılı olarak değinilmiştir.

Kısıtlanmış bireyselleştirilmiş bilgisayarlı test yöntemi (KBBT).

Kingsbury ve Zara (1989) uygulanacak bireyselleştirilmiş testte içerik dengelemesi yapıldığında, bireylerin, içerik alanlarının her birini mümkün olduğu kadar uygun şekilde temsil eden bir testle karşılaştıklarını ve bu sayede yanıtlarında boyutluluk bakımından gözlenebilecek bireysel farklılıkların etkisinin azaltılabileceğini ifade etmişlerdir. Bir testin müfredata veya test kapsamına uyumunu sağlamak için belirli içerik alanlarının çok fazla test edilmesi önlenmelidir. Bu amaç doğrultusunda KBBT, madde havuzundaki her bir içerik alanının önceden belirlenmiş bir madde yüzdesine sahip olmasını gerektirmektedir (Kingsbury ve Zara 1989). Kingsbury ve Zara’nın (1989) önerdiği şekliyle KBBT, uygulanan içerik oranı ile önceden belirlenmiş istenen içerik oranı arasında en büyük farklılığın gözlemlendiği alt içerik alanından en uygun maddenin seçilip bireye uygulanmasıyla gerçekleştirilir. Bir testin istenen içerik oranları, testteki alt hedef alanların her birinden gelmesi gereken test maddelerinin yüzdesi olarak belirlenebilir. Örneğin, bir aritmetik testi, %30 toplama, %20 çıkarma, %30 çarpma ve %20 bölme alt

içerik alanlarından oluşabilir. İstenen içerik oranları belirlendikten sonra, bu oranları sağlayan testi elde etmek için içerik dengelemeden faydalanılır.

Bir test uygulanırken, KBBT yöntemiyle içerik dengeleme süreci sırasıyla aşağıdaki adımlar izlenerek gerçekleştirilmektedir (Kingsbury ve Zara 1989).

1. Adım: Bir maddenin uygulamasının ardından bireyin geçici yetenek düzeyi kestirilir.

2. Adım: Mevcut testte her bir alt içerik alanında halihazırda uygulanmış maddelerin yüzdesi hesaplanır.

3. Adım: Hesaplanan yüzdeler, önceden belirlenmiş istenen içerik oranlarıyla karşılaştırılır ve en büyük farklılığa sahip olan alt içerik alanı belirlenir.

4. Adım: En büyük farklılığa sahip alt içerik alanından, bireyin geçici yetenek düzeyinde en yüksek bilgiyi sağlayan madde seçilir ve bireye uygulanır.

Bu adımlar test sonlanana kadar uygulanan her bir madde için sürekli olarak tekrar eder ve test sonlandığında nihai alt içerik oranları elde edilmiş olur.

Modifiye Edilmiş Multinomial Model (MMM). MMM'de, multinomial bir dağıtımı takiben her bir alt içerik alanı için önceden belirlenmiş istenen içerik oranlarına dayalı olarak kümülatif bir dağıtım oluşturulur. Daha sonra, 0 ile 1 arasında değişen tek biçimli bir dağılımdan rastgele bir sayı seçilir ve bu sayının karşılık geldiği alt içerik alanından seçilen en uygun madde bireye uygulanır. Örneğin A, B ve C alt içerik alanları için istenen içerik oranlarının sırasıyla 0,35, 0,40 ve 0,25 olduğu düşünülüründe, A, B ve C içerikleri için kümülatif yüzdeler sırasıyla 0,35, 0,75 (0,35 + 0,40) ve 1 (0,75 + 0,25) olacaktır. 0 ile 1 arasında değişen tek biçimli dağılımdan rastgele bir sayı olarak 0,3 seçilirse, içerik A seçilmiş olur; rastgele bir sayı olarak 0,6 seçilirse, içerik B seçilmiş olur ve seçilen içerik alanındaki en uygun maddenin bireye uygulanmasıyla süreç tekrar eder. İçerik alanının rastgele seçimi, test sürecindeki içerik sıralamasının öngörülebilirliğini önlemektedir. İstenen içerik oranlarına ulaşıldığında rastgele içerik seçimi durdurulur. İstenen içerik oranlarına ulaşılamamış ise içerik alanlarının istenen oranları ayarlanarak yeni bir multinomial dağılım oluşturulur. Multinomial dağılımın yenilenmesi, tüm alt içerik alanlarında istenen içerik oranlarına tam olarak ulaşıncaya kadar tekrar eder (Lin, 2011).

Madde Kullanım Sıklığı

Alanyazın incelendiğinde, geçmişten günümüze kadar BBT için birçok madde kullanım sıklığı kontrol yönteminin geliştirildiği görülmektedir. Georgiadou, Triantafillou ve Economides'e (2007) göre bu yöntemleri beş kategoride sınıflamak mümkündür. Bunlar; (1) rastgelelik stratejileri (randomization strategies), (2) koşullu seçim (conditional selection), (3) tabakalandırma stratejileri (stratified strategies), (4) birleşik stratejiler (combined strategies) ve (5) çok aşamalı bireyselleştirilmiş test desenleridir (multiple-stage adaptive test designs) (Lin, 2011).

Rastgelelik stratejisi, en bilgilendirici maddelerin olduğu bir gruptan rastgele bir maddenin seçilip uygulanmasıyla madde kullanım sıklığını kontrol etmeye çalışmaktadır. Ancak, gerçekçi test koşulları altında rastgelelik stratejilerinin etkililiğini tahminin etmenin zorluğundan dolayı araştırmacılar, BBT'lerde rastgelelik stratejilerinden ziyade daha ayrıntılı hazırlanmış olan madde kullanım sıklığı kontrol yöntemlerini tercih etmektedirler (Stocking ve Lewis, 1995). Rastgelelik stratejileri, maddenin sağladığı bilginin sadece bir veya birkaç kesme noktasında değerlendirildiği BBST'lerde, BBT'den daha basit olarak uygulanabilmektedir (örneğin, Lin ve Spray, 2000). Öyle ki, rastgelelik stratejilerinin bahsedilen dezavantajına rağmen, kavramsal açıdan basit ve uygulanması kolay olduğu için BBST'lerde en yaygın seçenek olarak tercih edildiği görülmektedir (Lin, 2011). *Koşullu seçim stratejisinde*, madde kullanım sıklığı, kontrol parametresi yardımıyla kontrol edilir. Sympson-Hetter yöntemi (SH; Sympson ve Hetter, 1985) koşullu seçim stratejisini temel alan yöntemlerin ilk versiyonudur (prototype version) (Lin, 2011). Alanyazındaki BBT ve BBST çalışmalarının çoğunluğunda tercih edildiği görülen SH yönteminin, istenen maksimum madde kullanım sıklığı oranının (r^{max}) belirlenerek havuzdaki her maddenin kullanım sıklığı oranını bu değere eşitlemeyi ya da bu değerden daha az olması için sınırlamayı amaçladığı görülmektedir. Bununla birlikte, SH yönteminin olasılıklı doğasından dolayı, özellikle ardışık madde seçimi esnasında maddelerin büyük bir bölümünde r^{max} oranının aşıldığının gözlenebilmesi mümkündür (Huebner, 2012). SH Yöntemi dışında Kısıtlayıcı Yöntem (KS; Restricted Method: RT; Revuelta ve Ponsoda, 1998) ile Madde Uygunluk Yöntemi (MU; Item Eligibility Method:IE; van der Linden ve Veldkamp, 2004) de koşullu

seçim stratejilerine dayanan ve maksimum madde kullanım sıklığı oranı r^{max} ile madde kullanım sıklığını kontrol etmeye çalışan diğer yöntemlerdendir. Koşullu seçim stratejilerinin en büyük dezavantajı, zaman alan simülasyonlar ve madde kullanım sıklığı kontrol parametrelerinin yetenek dağılımı ve madde havuzu özellikleri gibi BBT koşullarına bağımlı olmasıdır (Lin, 2011). *Tabakalandırma stratejileri* incelendiğinde ise Chang ve Ying'in (1999), testin ilk aşamasında düşük a (ayırıcılık) parametresine sahip maddeleri uygulayarak madde havuzu kullanımını dengeleyen a -tabakalama yöntemini önerdikleri görülmektedir. *Birleşik stratejiler*, tasarımında çeşitli stratejileri birleştirerek madde kullanım sıklığını kontrol etmeyi amaçlamaktadır. Örneğin, Leung, Chang ve Hau (2002) çalışmalarında a -tabakalamayı SH yöntemiyle birleştirmişlerdir. *Çok aşamalı bireyselleştirilmiş test tasarımlarında* ise bireyselleştirilmiş test formları önceden bir araya getirilerek madde kullanım sıklığı önsel (a priori) olarak kontrol edilmektedir (Lin, 2011).

Alanyazın incelendiğinde, BBST uygulamalarında en sık kullanılan madde kullanım sıklığı kontrol yöntemlerinin, tesadüfi madde seçiminin yapıldığı rastgelelik stratejileri ve SH yöntemi olduğu görülmektedir. Rastgelelik stratejilerinin gerçekçi test koşullarında çok fazla etkili olmayacağı düşünüldüğünden, bu çalışmada öncelikle SH yöntemi ve SH yöntemiyle aynı yaklaşımı temel alan KS ve MU yöntemlerine odaklanılmıştır. Barrada, Abad ve Veldkamp (2009), bu üç yöntemin performansını sabit uzunluklu bir BBT uygulamasıyla karşılaştırmışlardır. Huebner (2012) ise, Barrada ve diğerlerinin (2009) bu üç yöntemi karşılaştırdığı BBT uygulamasından elde edilen sonuçların ardışık madde seçiminin yapıldığı BBST ile tam olarak uyuşmayacağı düşüncesiyle değişken uzunluktaki bir BBST uygulamasında SH, KS, MU madde kullanım sıklığı kontrol yöntemlerini karşılaştırmış ve her bir yöntemin artı ve eksileriyle ilgili somut bilgiler sunmaya çalışmıştır. Barrada ve diğerleri (2009) ile Chen, Lei ve Lao'nun (2008) yaptıkları çalışmalar sonucunda, öngörülebilir madde kullanım sıklığı örüntülerinin KS yönteminin dikkat çekilen bir dezavantajı olduğunu ifade etmişlerdir. Ulaşılan bu sonuç ve yapılan diğer açıklamalar ışığında bu çalışmada ele alınan madde kullanım sıklığı kontrol yöntemleri, SH ve MU olarak belirlenmiştir. Bu yöntemler için tercih edilen maksimum madde kullanım sıklığı oranı ise Huebner (2012) ile Huebner ve Li'nin (2012) çalışmaları dikkate alınarak

yaygın ve sıklıkla seçilen bir değer olan $r^{max} = 0,20$ (Leung ve diğerleri, 2002) olarak alınmıştır. Ele alınan yöntemlerin matematiksel alt yapılarına aşağıda ayrıntılı olarak yer verilmiştir.

SH, KS ve MU yöntemleri, maksimum madde kullanım sıklığı oranlarını, havuzdaki her maddeye bir kullanım sıklığı kontrol parametresi atayarak kontrol etmektedir. Yöntemler arasındaki fark, bu kullanım sıklığı kontrol parametrelerinin nasıl ve ne zaman hesaplandığıdır. $m = 1, 2, \dots, M$, havuzdaki maddeler için bir indeks olmak üzere m . madde için kullanım sıklığı kontrol parametresinin değeri K_m ile gösterilmektedir. Bu yöntemler, madde seçimini madde uygulamasından ayırt etmektedir ve seçilen bir madde mutlaka uygulanmak zorunda değildir. Dolayısıyla $P(S_m)$ ve $P(A_m)$, sırasıyla, m maddesinin seçimi ve uygulanması için olasılıklar olmak üzere $P(A_m) \leq P(S_m)$ olması beklenen bir durumdur. K_m , her bir yöntemde $P(S_m)$ veya $P(A_m)$ niceliklerini kullanarak hesaplanmakta ve daima $[0,1]$ aralığında değerler almaktadır. Sınava giren i bireyi için bir madde seçildiğinde, 0 ile 1 arasında rastgele tek biçimli bir U değeri belirlenir ve madde yalnızca $U < K_m$ olduğunda uygulanır. Bir madde seçildikten sonra, uygulanıp uygulanmadığına bakılmaksızın, o madde artık o birey için bir daha seçilmez (Huebner, 2012). K_m 'nin her bir yöntemde nasıl elde edildiğine değinmek gerekirse;

Sympson-Hetter yöntemi (SH). SH yönteminde K_m değeri, test bireylere uygulanmadan önce tekrarlı (iteratif) simülasyonlar aracılığıyla hesaplanır. Bu ön simülasyonların her bir tekrarında $P(S_m)$ yeniden hesaplanır ve madde kullanım sıklığı kontrol parametreleri Eşitlik 27'deki kurala göre güncellenir (Huebner, 2012).

$$K_m = \begin{cases} 1 & \text{eğer } P(S_m) \leq r^{max} \\ \frac{r^{max}}{P(S_m)} & \text{eğer } P(S_m) > r^{max} \end{cases} \quad (27)$$

Gerçek testlerde son simülasyon tekrarından elde edilen K_m parametreleri kullanılmaktadır. Bir madde havuzu için belirlenen kullanım sıklığı parametre değerleri (K_m) değişmez ve hesaplanmasında, havuzdaki maddeler ile bireylerin simülasyonlarda belirtilen θ yetenek dağılımları dikkate alınır. Özel olarak belirtmek gerekirse, SH yönteminde K_m 'nin alabileceği değerler $[r^{max}, 1]$ aralığında değişmektedir. Bununla birlikte r^{max} için belirlenebilecek en büyük değer ise 1'dir ve bu değer madde kullanım sıklığı kontrolü yapılmayacağı anlamına gelmektedir.

SH yöntemiyle madde kullanım sıklığı kontrolüne bir örnek olarak; 500 maddelik bir madde havuzu ve 1000 kişilik bir örneklem düşünelim. Maddeler MFB aracılığıyla en çok bilgi verenden en az bilgi verene doğru sıralanmış ve K_m parametrelerini belirlemek için yapılan ön simülasyonlarda, istenen maksimum madde kullanım sıklığı oranı $r^{max} = 0,20$ olarak belirlenmiş olsun. Şunu belirtmek gerekir ki havuzdaki en bilgilendirici maddeler en katı kullanım sıklığı kontrol parametresine ($K_m = 0,20$) sahip olacaktır ve maddenin sağladığı bilgi azaldıkça parametre artacaktır (alabileceği en yüksek değer 1'dir). Test sürecine geçildiğinde, öncelikle MFB'ye göre en çok bilgiyi sağlayan ve bu sebeple ilk sırada olan madde 1, bireye uygulanmak üzere seçilir. Eğer $U < K_1 = 0,20$ ise, madde 1 bireye uygulanır. Aksi takdirde, madde 1 uygulanmaz ve bireye uygulanması uygun olmadığı için ileride de seçilmez ve işlem madde 2 için tekrarlanır. Bu süreç test sonlanana kadar devam eder. Bunların yanı sıra $N = 1000$ bireyin, en az $J_{min} = 50$ maddelik bir test aldığı varsayılırsa, havuzdaki en bilgilendirici ilk 50 maddenin tüm bireyler için seçilmesi kesin bir durum olur ancak $K_m = 0,20$ 'den dolayı, bu maddelerin her birinin gerçekten uygulanma şansı % 20 olur. Barrada ve diğerlerine (2009) göre, bu ve benzeri durumlarda testin sonunda elde edilecek madde kullanım sıklığı oranları belirlenen r^{max} oranından daha yüksek olabilir.

Kısıtlayıcı Yöntem (KS). KS yöntemiyle maksimum madde kullanım sıklığı kontrolü testin uygulanma anında (on-the-fly) yapılmaktadır. SH yönteminin aksine KS yönteminde, gerçek test başlamadan önce, madde kullanım sıklığı kontrol parametreleri belirlenmez. Bunun yerine, madde kullanım sıklığı kontrol parametreleri, testi alan birey sayısı arttıkça gözlemlenen madde kullanım sıklığı oranlarına göre sürekli olarak güncellenir. Kısacası, $K_m^{(i)}$ değerini belirlemek için ön simülasyonlar gerektirmeyen KS yönteminde, $K_m^{(i)}$ değeri havuzdaki maddelerden bağımsız olarak hesaplanır. KS yönteminin matematiksel ifadesinde, SH yöntemine benzer şekilde i . birey testi aldıktan sonraki m maddesi için kullanım sıklığı kontrol parametresi $K_m^{(i)}$ olarak gösterilir ve bir maddenin uygulanma olasılığı 1'den i 'ye kadar olan bireylerden hesaplanan $p^{(1...i)} (A_m)$ ile ifade edilir. Ardından, $(i + 1)$. birey için hesaplanan madde kullanım sıklığı kontrol parametreleri Eşitlik 28'deki kurala göre güncellenir (Huebner, 2012).

$$K_m^{(i+1)} = \begin{cases} 1 & \text{eğer } p^{(1\dots i)}(A_m) \leq r^{max} \\ 0 & \text{eğer } p^{(1\dots i)}(A_m) > r^{max} \end{cases} \quad (28)$$

Eşitlik 28'de görüldüğü gibi madde kullanım sıklığı kontrol parametreleri sadece 0 ile 1 değerlerini alabilir ve herhangi bir maddenin kullanım sıklığı oranının r^{max} değerini aşması mümkün değildir. SH yöntemiyle karşılaştırıldığında KS yönteminin, kavramsal olarak SH yönteminden daha basit olduğu söylenebilir (Huebner, 2012). Örnek olarak; $r^{max} = 0,20$ olan bir BBST uygulaması düşünecek olursak, madde m 'nin, $(i + 1)$. bireye uygulanıp uygulanmayacağı, bu maddenin $1, 2, \dots, i$ bireylerine ne sıklıkta uygulandığına bağlıdır. Eğer madde daha önceki i bireyin % 20'sine veya daha azına uygulanmışsa, yani $p^{(1\dots i)}(A_m) \leq 0,20$ ise, o zaman $K_m^{(i+1)} = 1$ olur ve seçilen bu madde $(i + 1)$. bireye uygulanır. Aksi takdirde, daha önceki bireylerin %20'sinden fazlası bu maddeyi görmüş olacağından $K_m^{(i+1)} = 0$ olur ve bu madde bireye uygulanmaz. KS yönteminin Chen ve diğerleri (2008) tarafından dikkat çekilen ve Barrada ve diğerleri (2009) tarafından tartışılan bir dezavantajı, bazı maddeler için öngörülebilir kullanım sıklığı örüntüleri üretmesidir. Örneğin, madde havuzu ve r^{max} değerine bağlı olarak bir maddenin veya bir grup maddenin her n . bireyde kullanımı gözlenebilir. Dolayısıyla, bu bireyler arasındaki madde örtüşme oranı (overlap rate) genel örtüşme oranından daha yüksek olur. Buna rağmen Huebner (2012), KS yöntemini, r^{max} 'a daha katı şekilde uyulması istenen BBST uygulamaları için özellikle önermiştir.

Madde uygunluk yöntemi (MU). MU yöntemi, KS'ye benzer şekilde maksimum madde kullanım sıklığı kontrolünün testin uygulanma anında yapıldığı bir başka yöntemdir. Madde kullanım sıklığı kontrol parametreleri ise Eşitlik 29'daki kurala göre güncellenir (Huebner, 2012).

$$K_m^{(i+1)} = \begin{cases} 1 & \text{eğer } \frac{p^{(1\dots i)}(A_m)}{K_m^{(i)}} \leq r^{max} \\ \frac{r^{max} K_m^{(i)}}{p^{(1\dots i)}(A_m)} & \text{eğer } \frac{p^{(1\dots i)}(A_m)}{K_m^{(i)}} > r^{max} \end{cases} \quad (29)$$

MU yöntemi kullanıldığında, maddeler r^{max} 'dan daha yüksek bir kullanım sıklığı oranına sahip olsalar bile, küçük de olsa, uygulanma şansları vardır. Bu sayede MU yöntemi, KS yönteminin öngörülebilir kullanım sıklığı örüntüleri üretebilmesi probleminin önüne geçer. MU yöntemi de KS yöntemi gibi ön simülasyonlara gerek duymaz ve $K_m^{(i)}$ değeri, havuzdaki maddelerden ve testi alan

bireylerin yetenek dağılımından bağımsız olarak belirlenir. MU yönteminin daha iyi anlaşılması amacıyla bir örnek verilecek olursa; bir BBST uygulamasında, $r^{max} = 0,20$ olan ve $1,2, \dots, i$ bireylerine dayalı kullanım sıklığı kontrol parametresi $K_m^{(i)} = 0,30$ olan madde m 'yi düşünelim. Bu maddenin kullanım sıklığı kontrol parametresi, mevcut uygulanma olasılığı $p^{(1\dots i)}(A_m)$ 'ye dayalı olarak, Eşitlik 29'da verilen kurala göre $(i + 1)$. birey için güncellenecektir. Eğer $p^{(1\dots i)}(A_m)$, r^{max} 'ın çok altındaysa örneğin $p^{(1\dots i)}(A_m) = 0,05$ ise $K_m^{(i+1)} = 1$ olur ve seçilen bu madde kesin olarak uygulanır. Barrada ve diğerlerine (2009) göre $p^{(1\dots i)}(A_m)$ değeri arttıkça, $K_m^{(i+1)}$ değeri azalır, diğer bir deyişle, kullanım sıklığı kontrol parametresi daha katı hale gelir ve maddenin uygulanma olasılığı düşer. Özetle, MU yönteminin $K_m^{(i)}$ parametresini KS yönteminden çok daha hassas bir biçimde güncellediği söylenebilir (Huebner, 2012).

İlgili Araştırmalar

Kingsbury ve Weiss (1980) tarafından yapılan simülasyon çalışmasında, sabit uzunluklu geleneksel kağıt-kalem testiyle AOOT ve BUT sınıflama kriterleri; OTU, OSD ve hata oranlarına göre karşılaştırılmıştır. Çalışmanın amacı doğrultusunda 100'er maddelik 4 madde havuzu oluşturulmuştur. Bunlar; i) tek biçimli (uniform) dağılımdan oluşturulan ($a= 1,0$, $b= 0,0$ ve $c= 0,2$ olacak şekilde), ii) Yalnızca b parametresinin değişkenlik gösterdiği, iii) a ve b parametrelerinin değişkenlik gösterdiği, iv) a, b ve c parametrelerinin değişkenlik gösterdiği madde havuzlarıdır. Tüm madde havuzları için 500'er bireye ait yetenek parametreleri $N(0,1)$ dağılımından türetilmiştir. Sabit uzunluklu geleneksel test için üç farklı uzunluktaki (10, 25 ve 50 madde) form, madde havuzundan rastgele seçilen maddelerle oluşturulmuştur. Araştırma sonucunda, tek biçimli madde havuzunda üç farklı test uzunluğu için de AOOT sınıflama kriterinin kullanıldığı durumda testin en az sayıda maddeyle sonlandığı; bunu sırasıyla BUT ve geleneksel kağıt-kalem testinin takip ettiği görülmüştür. Diğer madde havuzları içinse AOOT ile BUT'un benzer sonuçlar verdiği ifade edilmiştir. Bunların yanı sıra, OTU, OSD ve hata oranları bakımından en iyi performansın genel olarak AOOT sınıflama kriteriyle elde edildiği, buna karşılık geleneksel testin tüm madde havuzlarında ve tüm koşullarda AOOT ve BUT sınıflama kriterlerine kıyasla daha fazla madde gerektirdiği görülmüştür.

Reckase (1983) çalışmasında, iki kategoride yapılan sınıflamalar için AOOT'nin performansını iki farklı MTK modeli (1 PLM ve 3 PLM) temelinde incelemiştir. Bu doğrultuda, üç farksızlık bölgesi değeri ($\pm 0,3$, $\pm 0,8$ ve $\pm 1,0$) için elde edilen BBST sonuçları OTU ve OSD bakımından karşılaştırılmıştır. Çalışmada, 72 maddeden oluşan bir kelime testi 1 PLM ve 3 PLM ile kalibre edilerek iki farklı madde havuzu oluşturulmuştur. Bireylerin yetenek düzeyleri ise -3 ile 3 arasında 0,25 aralıkla türetilmiştir. Çalışmada MOK yetenek kestirim yöntemi ve kestirilen yetenekte maksimum bilgiyi veren madde seçme yöntemi kullanılmıştır. Çalışma sonucunda hem 1 PLM hem de 3 PLM için farksızlık bölgesinin genişliğinin OSD'yi fazla etkilemediği ancak farksızlık bölgesi daraldıkça sınıflama kararında gerekli OTU'nun arttığı görülmüştür. Buna ek olarak, farksızlık bölgesi genişlediğinde daha iyi hata olasılıkları elde edildiği ifade edilmiştir. Ayrıca c parametresinin OTU ve OSD'ye dikkate değer bir etkisi olduğu OTU bakımından 3 PLM'nin 1 PLM'den daha iyi performans gösterdiği görülmüştür.

Spray ve Reckase (1994) tarafından yapılan çalışmada; kesme noktasında, gerçek yetenek düzeyinde ve kestirilen geçici yetenek düzeyinde maksimum bilgi sağlayan madde seçme yöntemlerinin etkililiği OTU bakımından karşılaştırılmıştır. 200 maddelik gerçek veri setinden oluşturulan madde havuzunun kullanıldığı çalışmada, BBST simülasyonu (-3, 3) aralığında türetilmiş olan her bir yetenek düzeyi için 1000 tekrarla gerçekleştirilmiştir. Çalışmanın gerçek uygulamaya yakın olması amacıyla maksimum madde sayısı 50 olarak belirlenmiştir. Çalışmada daha genellenebilir sonuçlara ulaşmak amacıyla, üç ayrı kesme noktası (-0,5, 0,0, 1,0) kullanılmıştır. Araştırma sonucunda, kesme noktasında en yüksek bilgiyi veren madde seçme yönteminin diğer yöntemlere kıyasla daha iyi performans sağlandığı görülmüştür. Başka bir ifadeyle, madde seçimi kesme noktasında en yüksek bilgiyi veren maddeye göre yapıldığında bireyler daha az maddeyle sınıflanmıştır.

Spray ve Reckase (1996) çalışmalarında, iki kategoride yapılan sınıflamalar için AOOT ve BUT sınıflama kriterlerini, OTU ve sınıflama hata oranları bakımından karşılaştırmışlardır. Çalışmada kullanılan madde havuzu 200 maddelik olup BBST simülasyonu (-3, 3) aralığında türetilmiş olan her bir yetenek düzeyi için 1000 tekrarla gerçekleştirilmiştir. Çalışmada madde seçme yöntemi MFB-KN olarak, minimum madde sayısı 10 ve maksimum madde sayısı ise 50

olarak belirlenmiştir. Araştırma sonucunda, OTU bakımından AOOT'nin BUT'tan daha iyi performans gösterdiği ve genel olarak daha az maddeyle sınıflama kararı verildiği görülmüştür. Sınıflama hata oranlarının AOOT ve BUT için benzer olduğu, her iki yöntemin benzer sayıda maddeyle sınıflama yapabildiği durumlarda ise AOOT için hesaplanan sınıflama hatasının daha düşük olduğu ifade edilmiştir.

Lau (1996) çalışmasında, tek boyutluluk varsayımının ihlal edildiği durumlarda, iki kategoride yapılan sınıflamalar için AOOT sınıflama kriterinin performansını, OTU ve OSD bakımından incelemiştir. Çalışmada Monte Carlo simülasyonu ile iki boyutlu madde havuzu türetilmiş ve bu madde havuzu tek boyutlu MTK'ya dayalı olarak 1 PLM ve 3 PLM ile kalibre edilmiştir. Ayrıca çalışmada maksimum madde sayısı 50 olarak belirlenmiştir. Çalışma sonucunda, tek boyutluluk varsayımının ihlal edildiği durumda bile AOOT'nin başarılı bir performans gösterdiği, ayrıca OTU ve OSD bakımından 3 PLM'nin 1 PLM'ye kıyasla daha başarılı sonuçlar verdiği görülmüştür. Buna ek olarak çalışmada, iki boyutlu madde havuzu üzerinden yapılan yetenek kestirimlerinde tek boyutlu MTK modellerinin kullanılmasının OSD'yi düşürdüğü ifade edilmiştir.

Spray, Abdel-Fattah, Huang ve Lau (1997) tarafından yapılan çalışmada, çok boyutlu madde havuzu üzerinden tek boyutlu MTK'ya dayalı olarak gerçekleştirilen simülasyonlarda AOOT'nin performansı incelenmiştir. Yapılan Post-Hoc simülasyon çalışmasında kullanılan iki boyutlu madde havuzu, 30'ar maddelik altı paralel matematik testinden oluşturulmuştur. Araştırmada AOOT'nin performansı; herhangi bir sabitlemenin olmadığı, maksimum madde sayısının 60 ve 120 olarak sabitlendiği, sabitlenen maksimum madde sayısı ile madde kullanım sıklığı kontrolünün birlikte ele alındığı koşullar altında değerlendirilmiştir. Çalışma sonucunda, OTU'daki artışın OSD'yi de artırdığı ve düşük yetenek düzeylerinin madde kullanım sıklığı kontrolünden daha fazla etkilediği görülmüştür. Ayrıca AOOT'nin OSD bakımından çok boyutlu madde havuzunda bile oldukça iyi performans göstermiş olduğu ifade edilmiştir.

Lau ve Wang (1998) tarafından yapılan çalışmada, iki kategorili, çok kategorili ve karma maddelerden oluşan testlerde AOOT'nin performansı incelenmiştir. Yapılan Post-Hoc simülasyonunda kullanılan madde havuzu, 246 iki kategorili puanlanan ve 266 çok kategorili puanlanan maddelerden oluşmaktadır.

Çalışmada, rastgele madde seçimi, kesme noktasında en yüksek bilgiyi veren maddeler arasından rastgele madde seçimi ve kesme noktasında en yüksek bilgiyi veren madde seçimi ele alınmıştır. Yapılan rastgele madde seçimiyle, madde kullanım sıklığının kontrolü amaçlanmıştır. Çalışmada ayrıca minimum madde sayısı 10 ve maksimum madde sayısı 50 olarak belirlenmiştir. Araştırma sonuçlarına göre; iki kategorili, çok kategorili veya karma maddeler içeren testlerin tümünde, madde sayısı ve sınıflama hatası bakımından en başarılı performansı kesme noktasında en yüksek bilgiyi veren madde seçme yönteminin gösterdiği görülmüştür. Ayrıca AOOT'nin çok kategorili veya karma maddelerden oluşan testlerle de oldukça iyi bir performans gösterdiği ifade edilmiştir.

Eggen (1999) çalışmasında, iki ve üç kategoride yapılan sınıflamalarda MFB ve KLB madde seçme yöntemlerinin performansını, OTU ve OSD bakımından karşılaştırmıştır. 250 maddelik gerçek veri setinden oluşturulan madde havuzunun kullanıldığı çalışmada yetenek düzeyleri de aynı gerçek veriden tesadüfi olarak türetilmiştir. MFB-KY, MFB-KN ve KLB madde seçme yöntemlerinin incelendiği çalışmada gerçekleştirilen simülasyonlarda 5000 tekrar yapılmıştır. Çalışma sonucunda, iki ve üç kategoride yapılan sınıflamalarda KLB, MFB-KY ve MFB-KN'nin OTU ve OSD bakımından genel olarak benzer ve iyi performans gösterdiği, bazı koşullarda ise KLB'nin MFB'den daha başarılı sonuçlar verdiği görülmüştür. Buna ek olarak MFB-KN madde seçme yönteminin MFB-KY'ye kıyasla daha iyi performans gösterdiği ifade edilmiştir.

Lau ve Wang (1999) tarafından yapılan çalışmada, AOOT'nin performansı çok kategorili maddelerden oluşan madde havuzu üzerinden OTU, sınıflama hatası ve madde kullanım sıklığı bakımından incelenmiştir. Çalışmada; iki madde seçme yöntemi (KLB ve MFB), madde kullanım sıklığı kontrolü (SH yöntemi, rastgelelik yöntemi, kontrolün yapılmadığı koşullar), iki kesme noktası (-0,8 ve 0,8), iki madde havuzu genişliği (266 madde ve ilk havuzdan rastgele seçilen 90 madde) ve iki farksızlık bölgesi değeri (0,5 ve 1,0) dikkate alınmıştır. Gerçekleştirilen Post-Hoc simülasyonlarında bireylere uygulanacak minimum madde sayısı 3 maksimum madde sayısı 30 olarak belirlenmiştir. Çalışma sonucunda, OTU ve sınıflama hatası bakımından MFB ve KLB'nin benzer sonuçlar verdiği, madde kullanım sıklığı kontrolü yapılmadığında testin daha düşük OTU ve daha yüksek sınıflama doğruluğuyla sonlandığı görülmüştür. Buna ek olarak,

hesaplanan deęerlerin kesme noktalarına gre farklılıklar gsterdięi, daha geniř madde havuzu kullanıldıęında daha dřk OTU ve daha yksek sınıflama doęruluęunun elde edildięi ve farksızlık blgesi bydke OTU'nun azaldıęı ancak sınıflama hatasının arttıęı ifade edilmiřtir.

Eggen ve Straetmans (2000) tarafından yapılan alıřmada, AOOT ve GA sınıflama kriterlerinin performansları  kategorili sınıflamalar zerinden OTU, OSD ve madde kullanım sıklıęı bakımından incelenmiřtir. Gerek veri setiyle yrtlen alıřmada,  farklı madde seme ynteminin yanı sıra madde kullanım sıklıęı ve ierik dengeleme de dikkate alınmıřtır. Arařtırma kořulları; (i) tesadfi madde seme yntemi, (ii) MFB-KY, (iii) MFB-KN, (iv) MFB-KY ve ierik dengeleme, (v) MFB-KN ve ierik dengeleme, (vi) MFB-KY ve madde kullanım sıklıęı kontrol, (vii) MFB-KN ve madde kullanım sıklıęı kontrol, (viii) MFB-KY, ierik dengeleme ve madde kullanım sıklıęı kontrol, (ix) MFB-KN, ierik dengeleme ve madde kullanım sıklıęı kontrol bir arada ele alınacak řekilde belirlenmiřtir. 1198 bireye uygulanan 250 maddeden oluřan madde havuzunun kalibrasyonunda 2 PLM kullanılmıřtır. Gerekleřtirilen Post-Hoc simlasyonunda maksimum madde sayısı 25 olarak belirlenmiř ve AOK yetenek kestirim yntemi kullanılmıřtır. alıřma sonucunda, OTU ve OSD bakımından AOOT'nin GA'dan daha iyi performans gsterdięi, madde seme yntemlerinin genel olarak benzer sonular verdięi, ancak MFB-KY madde seme yntemi kullanıldıęında MFB-KN'ye kıyasla daha az sayıda maddeyle sınıflama kararı verildięi, buna karřılık MFB-KN yntemi kullanıldıęında sınıflama hatasının daha dřk olduęu grlmřtr. Buna ek olarak, sınıflama hatası bakımından ierik dengelemenin madde kullanım sıklıęı kontrolne gre daha iyi sonular verdięi ifade edilmiřtir. alıřmada ayrıca sabit uzunluklu 25 maddeyle yapılan simlasyon alıřmasına kıyasla testin daha az maddeyle sonlandıęı ve bu sabit uzunluklu testte OSD bakımından AOOT'nin GA'ya gre daha iyi performans gsterdięi grlmřtr.

Lin ve Spray (2000) tarafından yapılan alıřmada, AOOT sınıflama kriterinin kullanıldıęı durumda MFB, KLB ve Aęırlıklandırılmıř Log-Odds Oranı madde seme yntemlerinin performansları OTU ve OSD bakımından karřılařtırılmıřtır. Arařtırma kořullarında,  farklı kesme puanı, 11 farksızlık blgesi, iki farklı bykteki gerek madde havuzu (360 ve 180 madde),  farklı madde kullanım sıklıęı oranı yer almaktadır. alıřmada, madde kullanım sıklıęı kontrolnde

rastgelelik stratejileri kullanılmıştır. Gerçekleştirilen simülasyon çalışması için normal dağılımdan 10000 yetenek düzeyi türetilmiştir. Çalışma sonucunda, karşılaştırılan üç madde seçme yönteminin de tüm koşullarda birbirine benzer sonuçlar verdiği görülmüştür.

Jiao ve Lau (2003) tarafından yapılan çalışmada, AOOT sınıflama kriteri kullanıldığında 1 PLM, 2 PLM ve 3 PLM için model uyumsuzluklarının sınıflama hatalarına etkisi incelenmiştir. Çalışma sonucunda, gerçek model 1 PLM ve 2 PLM iken başka bir modelin kullanılmasının sınıflama hatalarını değiştirmedeği, buna karşılık gerçek model 3 PLM iken diğer modellerden birinin kullanılmasının sınıflama hatalarını artırdığı görülmüştür. Öyle ki, 1 PLM kullanıldığında bireyler olduklarından üst kategoriye; 2 PLM kullanıldığında ise olduklarından alt kategoriye hatalı olarak sınıflanmıştır. Bu bulgu doğrultusunda, BBST uygulamaları için MTK model seçiminin oldukça önemli olduğu ifade edilmiştir.

Yang ve diğerleri (2006) tarafından yapılan çalışmada, MTK'ya dayalı bir BBST uygulamasında, beş farklı yetenek kestirim yönteminin (MOK, AOK, MSD, BSD ve Owen'in Bayesçi Kestirimi) performansı, bireyleri beş kategoride sınıflamadaki etkililikleri bakımından karşılaştırılmıştır. Gerçekleştirilen simülasyon çalışmasında sonlandırma kuralı olarak sabit test uzunluğu ve sabit standart hata belirlenmiştir. Yapılan karşılaştırmalarda, kestirilen yetenek için hesaplanan standart hata, yanlışlık ve RMSE değerleri dikkate alınmıştır. Çalışmada, simülatif olarak türetilmiş 1000 maddeden oluşan bir madde havuzu ve yedinci sınıf öğrencilerinin matematik derslerini değerlendirmek amacıyla kullanılan dört paralel formun maddelerinden (toplam 236 madde) oluşan gerçek bir madde havuzu kullanılmıştır. Çalışma sonucunda; (i) Her birey için $\theta = 0$ yetenek düzeyi sabit başlangıç noktası olarak belirlendiğinde, Bayesçi kestirimlerin (MSD, BSD ve Owen'in Bayesçi Kestirimi) bireyleri daha içteki kategoriye yanlış sınıfladıkları, (ii) Orta derecedeki yetenek düzeyi başlangıç noktası olarak belirlendiğinde, Bayes kestirimlerinin, MOK ve AOK'a göre biraz daha doğru sınıflamalar yaptığı, buna ek olarak BSD'nin, üç Bayesçi yöntem arasında en doğru sonuçları üreten yetenek kestirim yöntemi olduğu, (iii) Karşılaştırılan tüm yetenek kestirim yöntemlerinin, testi sonlandırmada gerekli madde sayısı bakımından benzer etkililiğe sahip olduğu görülmüştür.

Thompson (2007a) tarafından yapılan çalışmada, çok kategorili maddelerin kullanıldığı BBST simülasyon sonuçları, OTU ve OSD bakımından karşılaştırılmıştır. Çalışmada, iki sınıflama kriteri (AOOT ve BUT), iki madde seçme yöntemi (MFB-KY ve MFB-KN), iki psikometrik model (3 PLM ve GKKM), iki madde havuzu (tek biçimli ve normal dağılımdan türetilen) ve iki farklı kesme puanı sayısı (bir ve iki kesme puanı) dikkate alınmıştır. Gerçekleştirilen simülasyon çalışmasında 16 farksızlık bölgesi belirlenmiş ve her bir araştırma koşulu için 25 tekrar yapılmıştır. Yetenek parametreleri, belirlenen 13 yetenek düzeyinin her birinde 1000 birey olacak şekilde türetilmiştir. Çalışma sonucunda, en düşük OTU'nun, hem AOOT hem de BUT sınıflama kriterleri için GKKM yönteminin tek kesme puanıyla ve MFB-KN madde seçme yöntemiyle kullanıldığı durumda elde edildiği görülmüştür. Buna karşılık, en yüksek OTU ise BUT sınıflama kriteri için 3 PLM'nin iki kesme puanıyla ve MFB-KN madde seçme yöntemiyle kullanıldığı durumda elde edilmiştir. ANOVA ile özetlenen araştırma sonuçlarına göre OTU ve OSD'ye ait hata varyansının oldukça düşük olduğu görülmüştür. Dolayısıyla gerçekleştirilen simülasyon çalışmalarında örnekleme hatasının düşük olduğu ve elde edilen sonuçların tekrarlar arasında değişkenlik göstermediği ifade edilebilir. Çalışmada ayrıca, OTU'daki varyansın %54'ünün kesme puanı sayısından; OSD'deki varyansın ise %79,3'ünün MTK modelinden kaynaklandığı sonucuna ulaşılmıştır.

Thompson ve Ro (2007) tarafından yapılan çalışmada, üç sınıflama kriteri (AOOT, Bileşik Olabilirlik Oranı (BOO) ve GA) OTU ve OSD bakımından karşılaştırılmıştır. Çalışmada, AOOT için çeşitli farksızlık bölgesi değerleri, BOO için AOOT'dekine benzer farklı güvenilirlik düzeyleri ve GA için farklı güven aralığı değerleri ele alınmıştır. Çalışmada ayrıca, madde seçme yöntemi olarak; AOOT için MFB-KN, GA için MFB-KY ve BOO için hem MFB-KN hem de MFB-KY tercih edilmiştir. Gerçekleştirilen simülasyon çalışmasında 750 maddelik madde havuzu kullanılmış ve yetenek parametreleri 10000 birey için normal dağılımdan türetilmiştir. Kesme noktasının 0,5 olarak belirlendiği çalışmada maksimum madde sayısı 200 alınmıştır. Çalışma sonucunda; AOOT sınıflama kriteri için farksızlık bölgesi daraldıkça, BOO sınıflama kriteri için güvenilirlik düzeyi küçüldükçe ve GA sınıflama kriteri için güven aralığı genişledikçe OSD'nin yükseldiği buna karşılık OTU'nun arttığı görülmüştür. Araştırmanın tüm koşulları için %93'ün üstünde OSD

elde edilmiştir. Bunun yanı sıra BOO sınıflandırma kriterinin MFB-KN madde seçme yöntemiyle birlikte kullanıldığı durumda MFB-KY yöntemiyle birlikte kullanıldığı duruma kıyasla daha düşük OTU ve benzer OSD'ler elde edildiği görülmüştür.

Thompson (2009) tarafından yapılan çalışmada, AOOT ve GA sınıflama kriterlerinin performansı OTU ve OSD bakımından karşılaştırılmıştır. Çalışmada iki madde seçme yöntemi (MFB-KY ve MFB-KN) ve dört madde havuzu (350 veya 700 maddelik sivri veya basık dağılıma sahip) dikkate alınmıştır. Gerçekleştirilen simülasyon çalışmasında yetenek düzeyleri 10000 birey için normal dağılımdan türetilmiştir. Çalışma sonucunda, tüm madde havuzlarında en iyi performansın genel olarak AOOT sınıflama kriterinin MFB-KN madde seçme yöntemiyle birlikte kullanıldığı durumda elde edildiği görülmüştür. Ayrıca en düşük OTU ve en yüksek OSD, sivri dağılımlı 750 maddelik havuzdan elde edilmiştir. Bunun yanı sıra, GA sınıflama kriterlerinin MFB-KY madde seçme yöntemiyle birlikte kullanıldığında, OTU bakımından daha iyi performans gösterdiği görülmüştür. Ayrıca OSD bakımından, basık madde havuzlarında GA ile MFB-KY birlikte daha iyi performans gösterirken sivri dağılımlı madde havuzlarında GA ile MFB-KN birlikte daha iyi performans göstermiştir.

Thompson (2011) tarafından yapılan çalışmada, BBST uygulamalarında GOO sınıflama kriterinin kullanımı, iki farklı Monte Carlo simülasyon çalışmasıyla incelenmiştir. İlk simülasyon çalışmasında, iki farklı δ farksızlık bölgesi sabit değeri (0,20 ve 0,30) kullanılmış, üç madde havuzu (50, 100 ve 200 madde) oluşturulmuş ve GOO sınıflama kriteri AOOT ve GA sınıflama kriterleri ile karşılaştırılmıştır. İkinci simülasyon çalışmasında ise tek madde havuzu (200 madde) kullanılmış ve GOO ile AOOT'nin performansları tek bir δ değeri (0,20) için karşılaştırılmıştır. Yapılan karşılaştırmalarda OTU ve OSD kullanılmıştır. Her iki çalışmada da yetenek parametreleri 10000 birey için normal dağılımdan türetilmiştir. Gerçekleştirilen simülasyon çalışmaları sonucunda, GOO'nun en az AOOT ve GA kadar etkili bir sınıflama kriteri olduğu ifade edilmiştir. Buna ek olarak, her iki çalışmada da GOO sınıflama kriterinin OTU ve OSD bakımından AOOT'den daha iyi performans gösterdiği görülmüştür. Ayrıca ilk simülasyon çalışmasında, GOO için $\delta = 0,30$ olduğunda test, en düşük OTU ve oldukça yüksek OSD (%95,73) ile

sonlanmıştır. GA sınıflama kriteri kullanıldığında ise test görelisi olarak daha yüksek OTU ve benzer OSD'lerle sonlanmıştır.

Huebner (2012) tarafından yapılan çalışmada, AOOT sınıflama kriterinin kullanıldığı durumda üç farklı madde kullanım sıklığı kontrol yönteminin (SH, KS, MU) performansı karşılaştırılmıştır. Yapılan karşılaştırmalar; OTU, OSD, maksimum madde kullanım sıklığı oranı, madde kullanım sıklığını aşan maddelerin yüzdesi ve madde kullanım sıklığını aşan maddelerin ortalama kullanım sıklığı üzerinden yapılmıştır. Bu amaçla farklı güçlükte iki madde havuzu (500 ve 800 madde) simüle edilmiştir. Yetenek düzeyleri 1000 birey için normal dağılımdan türetilmiştir. Çalışmada üç farklı maksimum madde kullanım sıklığı oranı ($r^{max} = 0,15, 0,20$ ve $0,25$) dikkate alınmıştır. Çalışma sonucunda SH, KS, MU madde kullanım sıklığı kontrol yöntemlerinin, OTU ve OSD bakımından benzer sonuçlar verdiği görülmüştür. Ayrıca testin gücü arttıkça OTU'nun arttığı ve OSD'nin yükseldiği görülmüştür. Buna ek olarak, madde kullanım sıklığı kontrolü bakımından en iyi performansı KS yönteminin gösterdiği, bunu sırasıyla MU ve SH yöntemlerinin izlediği görülmüştür.

Huebner ve Li (2012) çalışmalarında, BBST uygulamaları için önerdikleri ve SH yöntemiyle birlikte çalışan Stokastik Madde Kullanım Sıklığı Dengeleme Yönteminin (SMKSD; Stochastic Item Exposure Balance: SIEB) madde kullanım sıklığı kontrolünü iyileştirmedeki etkililiğini incelemişlerdir. Bu amaç doğrultusunda gerçekleştirilen iki simülasyon çalışmasında SH ve SMKSD yöntemleri: (i) OSD; (ii) OTU; (iii) değişken uzunluktaki BBST için havuzdaki madde kullanım sıklığı oranı (iv) düşük kullanım sıklığına sahip maddelerin oranı; (v) yüksek kullanım sıklığına sahip maddelerin oranı bakımından AOOT sınıflama kriteri altında karşılaştırılmıştır. Her bir çalışmada sekiz farklı test koşulu simüle edilmiş ve her koşul için 50 tekrar gerçekleştirilmiştir. Simülasyon çalışmalarında MTK'ya dayalı modellerden 3 PLM tercih edilmiş ve a, b, c parametreleri normal dağılımdan sırasıyla $N(0,7, 0,2)$, $N(-0,75, 2,0)$ ve $N(0,25, 0,03)$ olacak şekilde türetilerek ilk çalışma için 750 maddelik, ikinci çalışma içinse 500 maddelik madde havuzu oluşturulmuştur. Yetenek parametreleri iki çalışma için de her koşulun her bir tekrarında 1000 birey için normal dağılımdan türetilmiştir. Çalışmada, maksimum madde kullanım sıklığı oranı 0,20 olarak belirlenmiştir. Ayrıca içerik dengelemenin de dikkate alındığı çalışmada, içerik dengesini sağlamada, spiral yöntem

kullanılmıştır. Çalışma sonucunda, SMKSD'nin genel olarak, OTU ve OSD'den ödün vermeden, maddelerin kullanım sıklıklarının önemli ölçüde iyileştirdiği ve bu sayede havuzlardaki kullanım sıklığı fazla olan maddelerin oranını düşürdüğü görülmüştür.

Nydick ve diğerleri (2012) tarafından yapılan çalışmada; AOOT, GOO, GA ve Stokastik Azaltmalı AOOT (SA-AOOT; SPRT with Stochastic Curtailment: SCSVRT) sınıflama kriterleri, üç ve beş kategorili sınıflamalar üzerinden OTU ve OSD bakımından karşılaştırılmıştır. Çalışmada, 600 maddelik ve sekiz alt boyuttan oluşan gerçek bir veri seti kullanılmıştır. Madde havuzu 3 PLM ile kalibre edilmiş ve kesme noktaları keyfi olarak belirlenmiştir. Gerçekleştirilen simülasyonlarda, test başlangıcında sekiz alt boyutun her birinden birer madde uygulanacak şekilde minimum madde sayısı sekiz, maksimum madde sayısı 21 olarak belirlenmiştir. Yetenek kestiriminde, ilk maddeler için MOK, testin geri kalanı için AOK kullanılmıştır. Çalışma sonucunda, SA-AOOT ve GOO sınıflama kriterlerinin OTU ve OSD bakımından birbirine benzer ve diğer sınıflama kriterlerine kıyasla daha iyi performans gösterdiği görülmüştür. Ancak SA-AOOT'nin gerektirdiği karmaşık hesaplamalardan dolayı uygulamalarda GOO sınıflama kriterinin daha kullanışlı olabileceği ifade edilmiştir.

Sie ve diğerleri (2015) çalışmalarında, Fan, Wang, Chang ve Douglas'ın (2012) çalışmalarını genişleterek, BBST uygulamalarında test için bir zaman sınırının tanımlandığı ve hem madde seçiminde hem de testi sonlandırmada katılımcıların yanıt/tepki sürelerinin (TS; Response Time-RT) dikkate alındığı yeni bir yaklaşım önermeyi amaçlamışlardır. Test maddelerini seçerken TS'lerin dikkate alınmasının amacı, yüksek sınıflama doğruluğunu koruyarak toplam test süresini kısaltmaktır. Araştırmanın amacı doğrultusunda, BUT ve AOOT sınıflama kriterleri, MFB madde seçme yöntemiyle birlikte kullanılmıştır. Madde seçimi, Fisher bilgisinin kendisi yerine Fisher bilgisini zaman birimi başına (per time unit) maksimize ederek gerçekleştirilmiştir. BBST uygulaması, (i) AOOT bir sınıflandırma kararı verdikten sonra, (ii) zaman sınırı aşıldığında (iii) zaman sınırı aşılmadığı halde geriye kalan yeterince yüksek olasılıkla cevaplandırılacak bir madde olmadığında sonlandırılacak şekilde simüle edilmiştir. Yapılan simülasyon çalışmasında, 500 maddeden oluşturulan tek biçimli madde havuzu 3PLM ile kalibre edilmiştir. Yetenek parametrelerinin 500 birey için türetildiği çalışmada, TS

ayırıcılık parametreleri ise, $U(1, 3)$ tek biçimli dağılımından türetilmiştir. Çalışma sonucunda, önerilen yaklaşımın sınıflama doğruluğunu az miktarda geliştirdiği, test için harcanan ortalama sürenin ise önemli ölçüde azaldığı görülmüştür. Buna ek olarak, önerilen yaklaşımın zaman sınırını aşan bireylerin yüzdesini de azalttığı ifade edilmiştir.

van Groen ve diğerleri (2016) tarafından yapılan çalışmada, tek boyutlu BBST'lerle çok boyutlu BBST'lerin performansları, çok boyutlu ve tek boyutlu MTK ile yapılan sınıflamalar aracılığıyla OTU ve OSD bakımından karşılaştırılmıştır. Araştırmanın temel amacı, tek boyutlu BBST'lerde sıklıkla kullanılan AOOT sınıflama kriterinin çok boyutlu BBST'ler için de uygulanıp uygulanamayacağını araştırmaktır. Çalışmada kullanılan iki boyutlu madde havuzu, daha önce NOHARM II kullanılarak telafi edici bir MTK modeliyle kalibre edilmiş olup 180 maddeden oluşmaktadır. Yetenek kestiriminde AOK yöntemi, madde seçiminde ise rastgele madde seçimi, kesme noktasında en yüksek bilgiyi veren madde ve kestirilen yetenekte en yüksek bilgiyi veren madde tercih edilmiştir. Çalışma sonucunda, çok boyutlu BBST için AOOT'nin farksızlık bölgesi genişledikçe OTU'nun azaldığı ve OSD'nin değişmediği görülmüştür. Buna ek olarak, aralarında oldukça az fark olmasına rağmen, kesme noktasında en yüksek bilgiyi veren madde seçme yönteminin, kestirilen yetenekte en yüksek bilgiyi veren madde seçme yöntemi ve rastgele madde seçimine göre daha etkili olduğu, ancak bireyin yeteneği, kesme noktasına yaklaştıkça sınıflama hatasının arttığı ifade edilmiştir. Çalışmada genel olarak, tek boyutlu MTK ile karşılaştırıldığında, çok boyutlu MTK ile uygulanan AOOT'nin daha uzun testlerle sonlandığı ve daha doğru sınıflama kararları verdiği görülmüştür. OSD'nin önemi göz önüne alındığında, veri setinin model uyumu çok boyutlu MTK ile sağlandığında, AOOT'nin çok boyutlu MTK ile birlikte kullanılması gerektiği ifade edilmiştir.

Gündeğer (2017) doktora tez çalışmasında, BBST'de kullanılan farklı sınıflama kriterlerini (AOOT, GO ve GOO), yetenek kestirim yöntemlerini (BSD ve AOK) ve madde seçme yöntemlerini (MFB-KY, MFB-KN, KLB-KY ve MFB-KN) hem Monte Carlo (MC) hem de Post Hoc (PH) simülasyonları altında, OTU, OSD ve ölçme kesinliği bakımından karşılaştırmıştır. Sınıflama kriterlerinden AOOT ve GOO için farksızlık bölgeleri 0,05 ve 0,10 olarak GA için güven aralıkları ise %70 ve %90 düzeylerinde belirlenmiştir. R ortamında yürütülen çalışmada, yapılan

çaprazlamalar sonucunda toplam 96 koşul incelenmiş ve sonuçlar 25 tekrarın ortalaması alınarak hesaplanmıştır. MC simülasyonu için MTK'ya dayalı 3 PLM temelinde 500 maddelik bir havuz oluşturulmuş; PH simülasyonu için ise Kezer'in (2013) çalışmasındaki 80 maddelik gerçek veri seti kullanılmıştır. Yetenek parametreleri, MC simülasyonu için normal dağılımdan 3000 birey için türetilmiş olup PH simülasyonu için ise veri setindeki 994 birey üzerinden 3 PLM temelinde BSD ile kestirilmiştir. Çalışma sonucunda hem MC hem de PH simülasyonlarında OTU ve OSD bakımından GOO ve GA yöntemlerinin daha iyi performansla sahip olduğu, buna karşılık, ölçme kesinliği bakımından AOOT'nin daha etkili olduğu görülmüştür. Ayrıca, sınıflama kriterlerden tüm koşullarda oldukça yüksek OSD elde edildiği, güven aralığı daraldıkça veya farksızlık bölgesi genişledikçe OTU'nun azaldığı ifade edilmiştir. Gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyonlar incelendiğinde AOK ve BSD yöntemlerinin oldukça iyi kestirimlerde bulunduğu, buna karşılık, yanlılık, RMSE ve OMH değerleri bakımından BSD'nin daha başarılı olduğu sonucuna ulaşılmıştır. Ele alınan madde seçme yöntemlerinin ise birbirine benzer performanslar gösterdiği ancak MFB-KY'nin performansının tüm koşullarda daha iyi olduğu görülmüştür.

Gündeğer (2018) tarafından yapılan çalışmada, BBST'lerde kullanılan farklı yetenek kestirim yöntemleri (MOK, AOK, BSD ve MSD), çok kategorili maddelerden oluşan farklı büyüklükteki madde havuzları (50, 100 ve 150 madde) kullanılarak karşılaştırılmıştır. Madde havuzları R ortamında gerçekleştirilen MC simülasyonu ile Aşamalı Tepki Modeli'ne dayalı olarak beşli Likert ölçeği tipinde oluşturulmuştur. Yetenek parametreleri ise normal dağılımdan 1000 birey için türetilmiştir. Yapılan karşılaştırmalarda; OTU, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE, madde kullanım sıklığı ve madde çakışması dikkate alınmış olup bu değerler, yapılan 25 tekrarın ortalaması alınarak hesaplanmıştır. Çalışma sonucunda, OTU bakımından tüm yetenek kestirim yöntemlerinin benzer sonuçlar verdiği, bununla birlikte MOK ve AOK yöntemleri kullanıldığında OTU'nun azaldığı görülmüştür. Çalışma sonuçlarına göre madde havuzu genişledikçe, BSD ve MSD'nin yanı sıra, MOK yönteminin de madde kullanım sıklığı ve madde çakışması bakımından iyi performans gösterdiği belirlenmiştir. Ancak korelasyon, yanlılık ve RMSE değerleri de dikkate

alındığında, çok kategorili puanlanan madde havuzlarında, BSD ve MSD'nin kullanımının daha uygun olacağı ifade edilmiştir.

Gündeğer ve Doğan (2018) tarafından yapılan çalışmada, madde seçme yöntemi (rastgele madde seçmi, MFB ve KLB), madde havuzu dağılımı (sivri ve basık) ve madde havuzu büyüklüğünün (50, 100, 200 ve 300 madde) OTU ve OSD'ye etkisi incelenmiştir. Sivri dağılıma sahip havuzdaki maddeler için b parametresi $N(1, 0,4)$ olarak, basık dağılıma sahip havuzdaki maddeler için ise b parametresi $N(1, 1,5)$ olarak türetilmiştir. Yetenek parametreleri ise normal dağılımdan 1000 birey için türetilmiştir. R yazılımında gerçekleştirilen simülasyon çalışması sonucunda, tüm madde havuzlarında OTU bakımından en kötü performansı rastgele madde seçme yönteminin gösterdiği, MFB ve KLB yöntemlerinin ise oldukça benzer sonuçlar verdiği görülmüştür. Madde havuzu büyüdükçe OTU'nun azaldığı ve OSD'nin düştüğü ancak yine de %90'ın üstünde sınıflama doğruluğu elde edildiği ifade edilmiştir. Bununla birlikte, sivri dağılıma sahip madde havuzlarında OTU'nun azaldığı, OSD'nin ise değişmediği görülmüştür. Elde edilen sonuçlara göre, BBST uygulamalarında, sivri dağılıma sahip geniş madde havuzlarıyla yüksek OSD ve daha düşük OTU'ya sahip testlerin oluşturulabileceği ifade edilmiştir.

İlgili araştırmalar özet. BBST ile ilgili araştırmalar incelendiğinde, 1980'lerden günümüze kadar yapılmış araştırmaların büyük çoğunluğunun yurtdışında yapıldığı, ancak son yıllarda yurtiçinde de birkaç çalışmaya rastlandığı görülmektedir. BBST çalışmalarında genel olarak bir veya birden fazla sınıflama kriterinin kullanıldığı, sınıflama kriterlerinin farklı madde seçme yöntemleri, yetenek kestirim yöntemleri ve farklı büyüklükteki madde havuzlarıyla çaprazlanarak genellikle OTU ve OSD bakımından, bazen de bunlara ek olarak bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri bakımından karşılaştırıldığı görülmektedir. Bazı çalışmalarda ise BBST uygulamalarının en temel sınıflama kriterlerinden olduğu görülen AOOT'nin performansı; farklı MTK modelleri, madde havuzu dağılımları (sivri veya basık), tek boyutluluk varsayımının ihlali veya çok boyutlu madde havuzu kullanımı, madde havuzunun ikili puanlanan, çoklu puanlanan veya karma maddelerden oluşması gibi durumlarda incelenmiştir. Bununla birlikte alanyazında oldukça az sayıdaki

çalışmada ise test süreci için belirlenen diğer koşulların yanında madde kullanım sıklığı, içerik dengelemesi ve tepki/yanıt süresi sınırlandırması gibi ek kısıtlamalar getirildiği dikkat çekmektedir. Ayrıca çalışmaların çoğunluğunda sınıflamanın iki kategoride yapıldığı, bazı çalışmalarda ise kategori sayısındaki değişimin sonuçlara etkisini inceleyebilmek amacıyla farklı sayıdaki kategorilerde yapılan sınıflamaların karşılaştırıldığı dikkat çekmektedir. Bunların yanı sıra çalışmaların genelinde simülatif veri kullanıldığı ancak gerçek veri kullanılan çalışmaların da mevcut olduğu görülmektedir. BBST ile ilgili çalışmalar genel bir çerçevede değerlendirildiğinde; bu araştırmada amaçlandığı şekliyle, farklı sınıflama kriterlerinin, farklı madde seçme yöntemleri, içerik dengeleme yöntemleri ve madde kullanım sıklığı kontrol yöntemlerinin yanı sıra farklı sınıflama kategori sayılarıyla çaprazlandığı bir çalışmaya rastlanmamıştır. Bu durumun, özellikle içerik dengeleme ve madde kullanım sıklığı kontrolünün, OTU ve OSD üzerinde oluşturabileceği düşünülen olumsuz etkiden kaynaklanmış olabileceği söylenebilir. Ancak sınav güvenliği ve yüksek riskli testler düşünüldüğünde yapılacak BBST uygulamaları için madde kullanım sıklığı ve içerik dengelemenin önemi anlaşılmaktadır. Dolayısıyla bu koşulların da dahil edildiği BBST çalışmalarının alanyazında yer almasının oldukça önemli olduğu düşünülmektedir.

Bölüm 3

Yöntem

Bu bölümde araştırmanın yöntemine, Monte Carlo simülasyonunda kullanılan verilerin türetilmesine, BBST simülasyon koşullarına ve verilerin analizine yer verilmiştir.

Araştırma Yöntemi

Bu çalışmada, BBST'lerde kullanılan farklı sınıflama kriterleri, madde seçme yöntemleri ve AOK yetenek kestirim yönteminin yanı sıra BBT ve BBST uygulamalarında kullanılan farklı içerik dengeleme ve madde kullanım sıklığı kontrol yöntemleri gibi pratik kısıtlamalar altında yapılacak iki, üç veya dört kategorili sınıflamaların; OTU, OSD, gerçek ve kestirilen son yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH bakımından karşılaştırmalı olarak incelenmesi amaçlanmıştır. Dolayısıyla araştırmanın betimsel olduğu söylenebilir. Betimsel araştırmalar, bir durumu tam ve dikkatli bir şekilde tanımlayan, bireylerin, grupların, olayların ya da fiziksel ortamların özelliklerini özetleyen veya açıklamaya çalışan araştırmalardır (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz ve Demirel, 2008, s.23). Bu araştırma aynı zamanda bir simülasyon çalışmasıdır. Simülasyon çalışmaları sayesinde çeşitli istatistiksel yöntemlerin performansları değerlendirilebilmekte ve "... olsa ne olurdu?" sorusuna yanıt aranmaktadır (Dooley, 2002). Gerçek veri kullanan çalışmalarla sadece mevcut veri üzerinden yöntemlerin performansları arasında fark olup olmadığını belirleyebilmek mümkünken, simülasyon çalışmalarında yöntemlerin performanslarını farklı koşullar altında karşılaştırmak mümkündür.

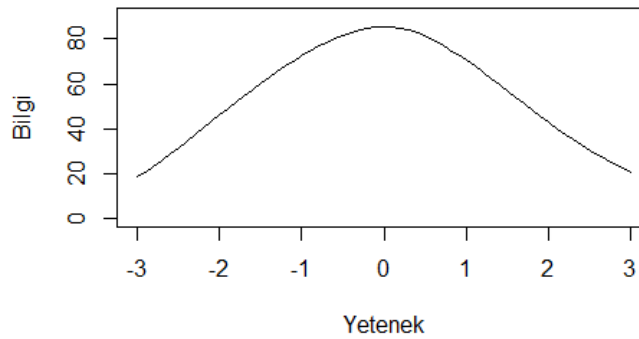
Verilerin Türetilmesi

Bu çalışma, Monte Carlo (MC) simülasyonu ile gerçekleştirilmiş ve BBST uygulama sonuçları simülatif veri setleri üzerinden karşılaştırılmıştır. Kullanılan veri setleri, çalışmada incelenen koşullara uygun şekilde R ortamında türetilmiştir (R Core Team, 2013).

Monte Carlo (MC) simülasyonu için madde ve yetenek parametrelerinin türetilmesi. MC simülasyonu olarak gerçekleştirilen bu çalışmada, Thompson'ın (2011) çalışması dikkate alınmış ve madde havuzu, MTK modellerinden 3 PLM temel alınarak 30 tekrarın her biri için 500 maddeden oluşturulmuştur. Çalışmada hem KN hem de KY temelinde MFB'ye dayanan madde seçme yöntemleri kullanılarak iki, üç veya dört kategoride sınıflamalar yapıldığından madde havuzu, $\theta=0$ kesme noktası ve etrafında yüksek bilgi veren ve yetenek düzeyi (-3, 3) aralığını kapsayan maddelerden oluşturulmuştur. Havuzdaki maddeler için:

- a parametresi, Kingsbury ve Weiss'in (1980) çalışmaları dikkate alınarak, orta ve yüksek ayırıcılık düzeylerini temsil edebilmesi bakımından tek biçimli dağılımdan $U[0,5, 2,0]$ olacak şekilde;
- b parametresi, Thompson'ın (2009) ve Warm'ın (1989) çalışmalarında değinildiği gibi gerçek uygulamadaki değerlere yakın olabilmesi bakımından, normal dağılımdan $N(-0,5, 1,5)$ olacak şekilde;
- c parametresi ise Thompson'ın (2009) çalışması dikkate alınarak ve yine gerçek bir uygulamaya yakın olması düşünülerek, normal dağılımdan $N(0,20, 0,05)$ olacak şekilde türetilmiştir.

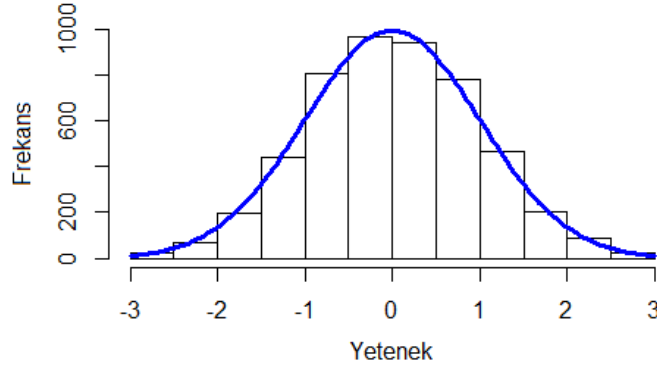
MC simülasyonu ile 500 madde için türetilen madde parametrelerinden rastgele seçilmiş bir tanesine ait betimsel istatistikler EK-A'da yer almaktadır. Bu maddelerin oluşturduğu madde havuzuna ait test bilgi fonksiyonu ise Şekil 4'te sunulmuştur.



Şekil 4. Test bilgi fonksiyonu.

Şekil 4'te görüldüğü gibi madde havuzu amaca uygun olarak; 0,0 kesme noktası ve etrafında yüksek bilgi veren ve yetenek düzeyi -3,+3 aralığını kapsayan maddelerden oluşmaktadır.

Bunun yanı sıra bireylerin yetenek parametreleri (-3,+3) aralığında, ortalaması 0 ve standart sapması 1 olan normal dağılımdan (N(0,1)) 30 tekrarın her biri için toplam 5000 birey olarak R yazılımında rastgele türetilmiştir. Türetilen yetenek parametrelerinden rastgele seçilmiş bir tanesine ait betimsel istatistikler EK-B'de yer almaktadır. Ayrıca rastgele seçilmiş olan bu yetenek parametrelerinin dağılımı Şekil 5'te sunulmuştur.



Şekil 5. Yetenek parametrelerinin dağılım grafiği.

Şekil 5'e göre 5000 birey için türetilen yetenek düzeylerinin amaca uygun olarak -3, +3 aralığında normal dağıldığı söylenebilir.

MC simülasyonu için madde ve birey parametrelerinin türetilmesinin ardından 5000 bireyin 500 maddeye ilişkin madde cevap örüntüsü, R ortamında Falissard (2012) tarafından yazılan "psy" paketi kullanılarak oluşturulan kodlar yardımıyla üretilmiştir. Bunun yanı sıra madde bilgilerini elde etmek için Nydick (2014) tarafından yazılan "catIrt" paketindeki "FI" fonksiyonundan, geçici ve son yetenek kestirimleri için ise yine "catIrt" paketindeki "wleEst" fonksiyonundan yararlanılmıştır.

MC Veri setinin tek boyutluluk varsayımının incelenmesi. Türetilen MC veri setleri üzerinden BBST simülasyonu gerçekleştirilmeden önce rastgele seçilmiş bir veri setinin boyutluluğu R'da açıklayıcı faktör analizi ve AMOS programında doğrulayıcı faktör analiziyle incelenmiştir. Veri seti için öncelikle R ortamında Revelle

(2015) tarafından yazılan “psych” paketindeki “KMO” fonksiyonuyla KMO değeri hesaplanmış ve “cortest.barlett” fonksiyonuyla da Barlett küresellik testi uygulanmıştır. Ardından aynı paketteki “irt.fa” fonksiyonu kullanılarak tetrakorik korelasyonlara dayanan temel bileşenler faktör analizi gerçekleştirilmiştir. Manidar olduğu görülen Barlett testi sonuçlarına (Ki-Kare= 449853,18; sd= 124750; p= 0,00) göre değişkenler arasında yüksek korelasyonlar olduğu yorumu yapılabilir. Bunun yanı sıra hesaplanan KMO değeri (KMO= 0,98 > 0,50) de dikkate alınarak veri setinin faktör analizine uygun olduğu söylenebilir. Temel bileşenler faktör analizi sonucu elde edilen açıklanan varyans ve özdeğerler Tablo 3’te sunulmuştur.

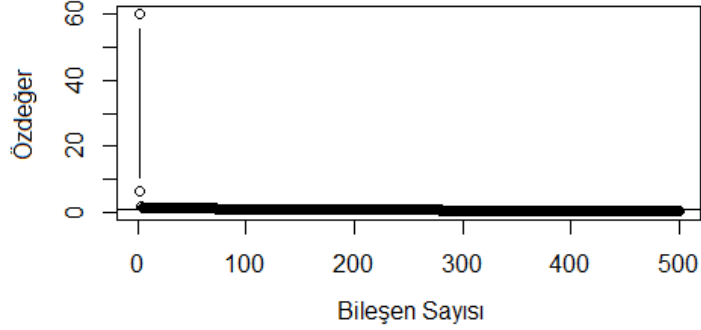
Tablo 3

MC Veri Setine Ait Açıklanan Varyans ve Özdeğerler

Değişken	Özdeğer	Varyans Oranı	Kümülatif Varyans Oranı
1	59,84	11,97	11,97
2	6,41	1,28	13,25
3	1,81	0,36	13,61
4	1,58	0,32	13,93
5	1,56	0,31	14,24
6	1,55	0,31	14,55
7	1,54	0,31	14,86

Faktör analizi sonucunda özdeğeri 1’in üzerinde olan 166 bileşen olduğu görülmüştür. Ancak bu bileşenlere ait özdeğerlerin birbirine oldukça yakın olmasından dolayı Tablo 3’te sadece ilk yedi bileşen sunulmuştur. İlk bileşene ait özdeğer 59,84 iken ikinci bileşene ait özdeğer 6,41’dir. Buna göre, birinci bileşenin özdeğerinin ikinci bileşenin özdeğerine oranı 9,34’tür. Bu oranın üçten büyük olması veri setinin tek faktörlü bir yapıya sahip olduğuna işaret etmektedir (Lord, 1980). Ayrıca Tablo 3’e göre birinci bileşen tek başına toplam varyansın %11,97’sini açıklamaktadır. Kümülatif varyans oranları incelendiğinde ise ikinci bileşen ve diğer bileşenlerin açıklanan varyansı çok az arttırdığı görülmektedir. Bunun yanı sıra, “psych” paketindeki “VSS.scree” fonksiyonu yardımıyla R’da çizdirilen ve Şekil 6’da sunulan yamaç-birikinti grafiği incelendiğinde, birinci faktörden sonra keskin bir düşüşün olduğu ve ikinci faktörden itibaren eğimin önemli ölçüde azaldığı görülmektedir.

Dolayısıyla, Tablo 3 ve Şekil 6'daki bulgular birlikte değerlendirildiğinde, veri setinin tek boyutluluk varsayımını sağladığı söylenebilir.



Şekil 6. Yamaç-birikinti grafiği.

Veri setine uygulanan doğrulayıcı faktör analizi sonucunda Ki-kare değeri 143865,692 ve serbestlik derecesi (sd) 124750 olarak hesaplanmıştır. Hesaplanan ki-kare değerinin sd'ye oranının ($ki-kare/sd = 1,15$) ikiden küçük olmasından dolayı veri setinin tek boyutluluğa mükemmel uyum sağladığı çıkarımı yapılabilir (Tabachnick ve Fidell, 2007). Bunun yanı sıra uyum indekslerinden CFI = 0,87; NFI = 0,87; GFI = 0,96; AGFI = 0,96 olarak, Alfa güvenilirlik katsayısı ise 0,98 olarak hesaplanmıştır. Dolayısıyla, uyum indekslerine göre de verinin tek boyutluluk varsayımını karşıladığı ve güvenirliliğin yüksek olduğu söylenebilir. Ancak Alfa güvenilirlik katsayısının madde sayısındaki artıştan etkilenerek yüksek değerler verme eğiliminde olabileceği unutulmamalıdır (Cortina, 1993).

MC Veri setinin yerel bağımsızlık varsayımının incelenmesi. Yerel bağımsızlık varsayımını test etmek amacıyla 5000 bireyin toplam puanlarına göre yüksekten düşüğe doğru oluşturulan dört yetenek grubu için yetenek sabitlendiğinde, tüm yetenek grupları için hesaplanan maddeler arasındaki Pearson korelasyon katsayılarının çoğunluğunun manidar olmadığı; manidar olanların ise oldukça düşük düzeyde ilişkiye sahip olduğu görülmüştür. Buna göre, yerel bağımsızlık varsayımının büyük oranda sağlandığı yorumu yapılabilir. Bununla birlikte McDonald'a (1980) göre tek boyutluluğun anlamlı bir açıklığı, yerel bağımsızlığı içermektedir ve aynı/benzer yetenek düzeylerinde maddeler arasındaki kovaryans sıfıra eşitse test tek boyutludur

(aktaran, Hambleton ve Swaminathan, 1985). Dolayısıyla, yapılan faktör analizi sonucunda, tek boyutluluk özelliği gösteren veri setindeki maddelerin aynı zamanda yerel bağımsızlık varsayımını da sağladığı söylenebilir.

Testin hız testi olmaması. Araştırmada kullanılan MC veri setleri arasından rastgele seçilmiş olan veri seti incelendiğinde, verinin simülasyonla üretilmiş olmasından dolayı, yanıtlanmamış madde veya testi bitirememiş birey bulunmadığı görülmüştür. Bu durum, testin hız testi olmadığı şeklinde yorumlanabilir. Dolayısıyla MTK'nın temel varsayımlarından olan tek boyutluluk ve yerel bağımsızlık varsayımlarını etkileyecek bir durum olmadığı, veri setinin MTK varsayımlarını sağladığı söylenebilir.

BBST Simülasyon Koşulları

Bu çalışmada gerçekleştirilen BBST simülasyonunda, Falissard (2012) tarafından yazılan “psy” paketinden ve Nydick (2014) tarafından yazılan “catIrt” paketinden yararlanılmıştır. Simülasyon koşullarının tamamında $\theta=0$ yetenek düzeyi başlama noktası olarak belirlenmiş ve oluşturulan 48 koşulun tümü için 30 tekrar yapılarak ortalaması alınmıştır. Gerçekleştirilen MC simülasyon çalışmasında R ortamında türetilmiş veri seti kullanılmış olup sınıflama kriterleri, madde seçme yöntemleri, içerik dengeleme yöntemleri, madde kullanım sıklığı kontrol yöntemleri ve sınıflama kategori sayıları çalışmanın amacına uygun şekilde manipüle edilmiştir. Çalışmada ele alınan sınıflama kriterleri; AOOT ($\delta: 0,20$) ve GA (%90) olarak, madde seçme yöntemleri ise MFB-KY ve MFB-KN olarak belirlenmiştir. Ayrıca çalışmada tercih edilen AOK yetenek kestirim yöntemi, simülasyonlarda sabit kalan bir koşuldur. Bununla birlikte, içerik dengelemede KBBT ve MMM yöntemleri ele alınmış olup içerik dengeleme koşullarını sağlamak için testi sonlandırmadan önce kullanılacak minimum madde sayısı 10, maksimum madde sayısı ise 70 olarak belirlenmiştir. KBBT ve MMM yöntemlerinin araştırma koşullarına dahil edildiği durumlarda, R ortamında türetilen 500 maddeyle oluşturulmuş madde havuzu, rastgele madde atamasıyla dört içerik alanına ayrılmıştır. Ardından araştırmacı tarafından yazılan fonksiyon ve döngüler yardımıyla madde seçiminin bu içerik alanları dikkate alınarak yapılması sağlanmıştır. Belirlenen içerik alanlarının istenen oranları ise sırasıyla %40, %30,

%20 ve %10 olacak şekilde ayarlanmıştır. Madde kullanım sıklığı kontrolünde kullanılacak SH ve MU yöntemleri için maksimum madde kullanım sıklığı oranı ise Huebner (2012) ile Huebner ve Li'nin (2012) çalışmaları dikkate alınarak yaygın bir değer olan $r^{max} = 0,20$ (Leung ve diğerleri, 2002) olarak alınmıştır. Yapılan sınıflamalar için ele alınan kategori sayıları ise iki, üç veya dört olarak manipüle edilmiştir.

Verilerin Analizi

Araştırma kapsamında oluşturulan 48 simülasyon koşulunun her biri için 30 tekrar yapılmıştır. Araştırmanın bağımlı değişkenleri olan OTU, OSD, bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon (r), yanlılık, RMSE, OMH değerleri yapılan tekrarların ortalaması alınarak hesaplanmıştır. Gerçek ve kestirilen yetenek düzeyleri arasındaki ilişki R ortamında "cor" komutu yardımıyla Pearson Momentler Çarpımı Korelasyon Katsayısı olarak, yanlılık, RMSE (Root Mean Squared Error) ve OMH (Ortalama Mutlak Hata) değerleri ise R ortamında yazılan fonksiyonlar yardımıyla elde edilmiştir.

Yanlılık, RMSE ve OMH değerleri; bireyin gerçek yetenek düzeyi (θ_i), kestirilen son yetenek düzeyi ($\hat{\theta}_i$), ve birey sayısına (n) bağlı olarak hesaplanabilmektedir (Miller ve Miller, 2004).

Yanlılık, bireylerin kestirilen son yetenek düzeylerinin ($\hat{\theta}_i$) gerçek yetenek düzeylerinden (θ_i) farkları toplamının birey sayısına (n) oranı alınarak hesaplanır ve formülü Eşitlik 30'daki gibidir.

$$Yanlılık = \frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)}{n} \quad (30)$$

RMSE değeri, bireylerin kestirilen son yetenek düzeylerinin ($\hat{\theta}_i$) gerçek yetenek düzeylerinden (θ_i) farklarının kareleri toplamının birey sayısına (n) oranının karekökü alınarak Eşitlik 31'deki gibi hesaplanmaktadır.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2}{n}} \quad (31)$$

OMH deęeri ise bireylerin kestirilen son yetenek düzeylerinin ($\hat{\theta}_i$) gerek yetenek düzeylerinden (θ_i) farklarının mutlak deęerleri toplamının birey sayısına (n) oranı alınarak Eşitlik 32'deki formül yardımıyla hesaplanmaktadır.

$$OMH = \frac{\sum_{i=1}^n |\hat{\theta}_i - \theta_i|}{n} \quad (32)$$

Bununla birlikte ierik dengeleme ve madde kullanım sıklığı kontrolü gibi pratik kısıtlamalar iin de arařtırmacı tarafından R ortamında madde seme yöntemine ek fonksiyon ve dōngüler yazılmıřtır.

Bölüm 4

Bulgular ve Yorumlar

Bu bölümde arařtırmanın alt problemlerine iliřkin bulgulara ve alanyazın desteęiyle yapılan yorumlara yer verilmiřtir.

Alt Problem 1'e İliřkin Bulgular ve Yorumlar

AOOT sınıflama kriterinin FB sabiti δ : 0,20 düzeyi ve GA sınıflama kriterinin GA: %90 güven düzeyi için yetenek kestirim yöntemi AOK iken iki, üç veya dört kategorili sınıflamalarda;

1.1. MFB-KY, MFB-KN madde seçme yöntemlerinin bir arada ele alındığı kořullarda OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH deęerleri nasıl deęiřmektedir?

Arařtırmada, madde seçme yöntemlerinden MFB-KY ve MFB-KN'ye göre yapılan karřılařtırmalarda, arařtırmanın baęımlı deęiřkenleri olarak belirlenen OTU, OSD, bireylerin kestirilen son yetenek düzeyleri ile gerçek yetenek düzeyleri arasındaki korelasyon (r), yanlılık, RMSE, OMH deęerleri kullanılmıř ve elde edilen deęerler Tablo 4'te sunulmuřtur. Sunulan deęerler, her bir kořul için yapılan 30 tekrarın ortalaması alınarak hesaplanmıřtır.

Tablo 4'e göre bireyler iki kategoride; GA (%90 güven düzeyi) sınıflama kriteri ile MFB-KY madde seçme yöntemi birlikte kullanıldıęında ortalama 11 madde ve 0,89 sınıflama doęruluęuyla, GA (%90 güven düzeyi) ile MFB-KN yöntemleri birlikte kullanıldıęında ortalama 12 madde ve 0,90 sınıflama doęruluęuyla, AOOT (δ : 0,20) ile MFB-KN yöntemleri birlikte kullanıldıęında ortalama 23 madde ve 0,94 sınıflama doęruluęuyla, AOOT (δ : 0,2) ile MFB-KY yöntemleri birlikte kullanıldıęında ortalama 25 madde ve 0,94 sınıflama doęruluęuyla sınıflanmıřtır. Buna göre, hem MFB-KY hem MFB-KN madde seçme yöntemlerinin kullanıldıęı durumlar için; en düşük OTU, bařka bir deyiřle bireyi sınıflamak ve testi sonlandırmak için gereken en az madde sayısının GA sınıflama kriterinin kullanıldıęı kořullardan elde edildięi görölmektedir. Bu bulguya göre KY veya KN temelli madde seçme yöntemi fark etmeksizin GA sınıflama kriteri, AOOT'ye kıyasla OTU bakımından daha iyi performans göstermiřtir.

Tablo 4

Yetenek Kestirim Yöntemi AOK Olduğunda Alt Problem 1.1. Koşullarına Ait OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri

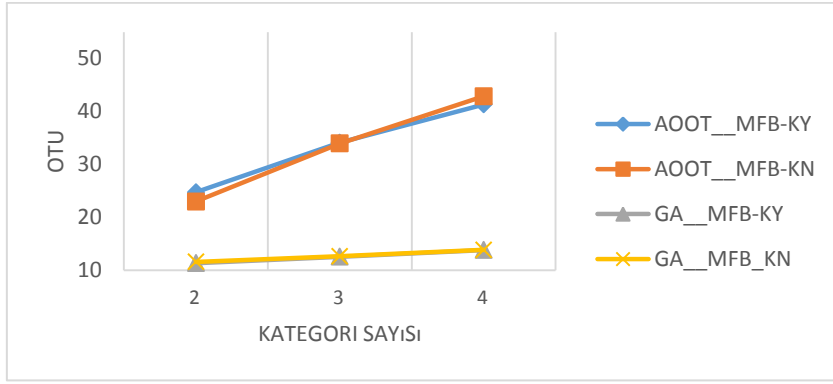
Sınıflama Kriteri	Koşullar		Bağımlı Değişkenler					
	Madde Seçme Yöntemi	Sınıflama Kategori Sayısı	OTU	OSD	r	Yanlılık	RMSE	OMH
AOOT δ: 0,20	MFB-KY	İki	24,72	0,94	0,94	-0,011	0,354	0,269
		Üç	34,08	0,88	0,96	-0,012	0,316	0,236
		Dört	41,34	0,82	0,96	-0,014	0,293	0,218
	MFB-KN	İki	22,95	0,94	0,90	0,019	0,435	0,325
		Üç	33,93	0,89	0,92	0,015	0,385	0,282
		Dört	42,88	0,82	0,93	0,012	0,353	0,257
GA %90 güven düzeyi	MFB-KY	İki	11,33	0,89	0,90	0,016	0,456	0,353
		Üç	12,52	0,79	0,91	0,015	0,450	0,346
		Dört	13,81	0,71	0,91	0,016	0,443	0,339
	MFB-KN	İki	11,55	0,90	0,87	0,019	0,488	0,377
		Üç	12,62	0,80	0,87	0,017	0,481	0,371
		Dört	13,82	0,71	0,88	0,020	0,473	0,363

Not: OTU= Ortalama Test Uzunluğu, OSD= Ortalama Sınıflama Doğruluğu, r= Gerçek ve Kestirilen Yetenek Düzeyleri Arasındaki Korelasyon.

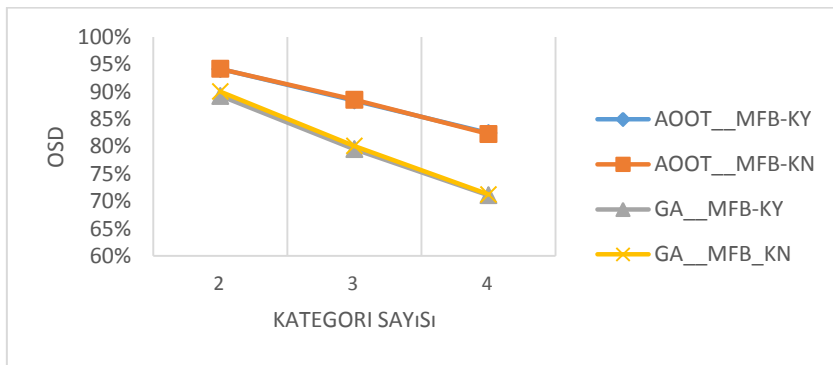
Tablo 4 incelendiğinde, madde seçme yöntemi fark etmeksizin iki, üç veya dört kategoride yapılan sınıflamalarda, AOOT sınıflama kriteri kullanıldığında elde edilen OSD'nin 0,82 ile 0,94 aralığında ve oldukça yüksek olduğu, GA sınıflama kriteri kullanıldığında elde edilen OSD'nin ise 0,71 ile 0,90 aralığında nispeten daha düşük olduğu görülmektedir. Buna göre, kullanılan sınıflama kriteri AOOT olduğunda daha yüksek OSD'lerin elde edilebileceği söylenebilir. Kısacası, Tablo 4'teki, OTU ve OSD değişkenleri için hesaplanan değerler birlikte incelendiğinde, OTU bakımından GA yönteminin, OSD bakımından ise AOOT yönteminin performansının daha yüksek olduğu söylenebilir. Araştırmanın bu bulgusu Nydick ve diğerleri (2012), Thompson (2009) ile Thompson ve Ro'nun (2007) araştırma sonuçlarıyla da benzerlik göstermektedir. Yapılan bu araştırmalarda genel olarak, GA ile yapılan sınıflamaların AOOT ile yapılan sınıflamalara kıyasla daha düşük OTU ve OSD ile sonlandığı ifade edilmiştir. Buna karşılık Gündeğer (2017) yaptığı çalışmada, GA yönteminin AOOT'ye kıyasla OTU bakımından daha iyi OSD bakımından ise benzer performans gösterdiği sonucuna ulaşmıştır. Bunun yanı sıra Tablo 4'te GA sınıflama kriteri ile MFB-KY madde seçme yönteminin, AOOT sınıflama kriteri ile de MFB-KN madde seçme

yönteminin birlikte kullanıldığı koşullarda test etkililiğinin (OTU ve OSD) az da olsa kısmen daha yüksek olduğu söylenebilir. Araştırmanın bu bulgu Thompson'ın (2009) sonuçlarıyla da örtüşmektedir. Thompson (2009) yaptığı çalışmada, GA sınıflama kriteri için KY temelindeki madde seçiminin, AOOT sınıflama kriteri için ise KN temelindeki madde seçiminin daha uygun olduğu sonucuna ulaşmıştır.

Tablo 4'e göre, OTU ve OSD için farklı koşullardan ayrı ayrı hesaplanan en iyi değerler, iki kategoride yapılan sınıflamalardan (OTU için 11,33; OSD için 0,94), en kötü değerler ise dört kategoride yapılan sınıflamalardan (OTU için 42,88; OSD için 0,71) elde edilmiştir.



Şekil 7. Sınıflama kriterleri ve madde seçme yöntemlerinin çaprazlandığı OTU değişim grafiği.



Şekil 8. Sınıflama kriterleri ve madde seçme yöntemlerinin çaprazlandığı OSD değişim grafiği.

Tablo 4'te sunulan değerlerin yanı sıra Şekil 7 ve Şekil 8'deki grafikler incelendiğinde, sınıflama kriteri ve madde seçme yöntemi fark etmeksizin, sınıflama

kategori sayısı arttıkça bireyleri sınıflamada gerekli olan madde sayısının (OTU) arttığı ve OSD'nin ise azaldığı görülmektedir. Araştırmanın bu bulgusu Eggen (1999) ile Nydick ve diğerlerinin (2012) araştırma sonuçlarıyla da desteklenmektedir. Eggen (1999) çalışmasında iki ve üç kategoride yapılan sınıflamaları, Nydick ve diğerleri (2012) ise üç ve beş kategoride yapılan sınıflamaları karşılaştırmışlar ve bireyler daha fazla kategoride sınıflandığında OTU artarken OSD'nin azaldığı, diğer bir ifadeyle test etkililiğinin düştüğü sonucuna ulaşmışlardır.

Tablo 4'te bireylerin gerçek yetenek düzeyleri ile kestirilen yetenek düzeyleri arasındaki en yüksek korelasyonların (r) AOOT sınıflama kriterinin kullanıldığı durumlarda gözlemlendiği, GA'nın kullanıldığı durumlarda ise nispeten daha düşük korelasyonların elde edildiği görülmektedir. Öyle ki, madde seçme yöntemi fark etmeksizin korelasyon (r) değerleri; AOOT sınıflama kriteri için 0,90 ile 0,96 aralığında, GA sınıflama kriteri için ise 0,87 ile 0,91 aralığında hesaplanmıştır. Bu bulguya göre korelasyon (r) değerleri bakımından AOOT yönteminin GA yöntemine kıyasla daha iyi performans gösterdiği söylenebilir. Tablo 4'teki sınıflama kriterlerinin madde seçme yöntemleriyle çaprazlandığı koşullar incelendiğinde ise her iki sınıflama kriteri için de MFB-KN'ye kıyasla MFB-KY madde seçme yönteminin dahil olduğu koşullarda daha yüksek korelasyonların belirlendiği görülmektedir. Araştırmanın bu bulgusu Gündeğer'in (2017) araştırma sonuçlarıyla da örtüşmektedir. Bunun yanı sıra Tablo 4'e göre sınıflama kategori sayısı arttıkça korelasyon değerleri de nispeten yükselmiştir. Tablo 4 incelendiğinde, AOOT ile MFB-KY yöntemleri birlikte kullanıldığında hesaplanan yanlılığın (-0,014, -0,011 aralığında), AOOT ile MFB-KN'nin birlikte kullanıldığı (0,012, 0,019 aralığında) durumlara kıyasla daha düşük olduğu görülmektedir. Benzer şekilde GA ile birlikte MFB-KY kullanıldığında hesaplanan yanlılık (0,015, 0,016 aralığında) GA ile birlikte MFB-KN'nin kullanıldığı (0,017, 0,020 aralığında) duruma kıyasla daha düşüktür. O halde, oldukça düşük oldukları gözlenen yanlılık değerlerinin; kullanılan sınıflama kriteri ve madde seçme yöntemine göre değiştiği, AOOT sınıflama kriterinin veya MFB-KY madde seçme yönteminin kullanıldığı durumlarda yanlılığın daha düşük olduğu söylenebilir. Buna ek olarak, her iki sınıflama kriteri için de MFB-KY madde seçme yönteminin kullanıldığı koşullarda hesaplanan yanlılıkların daha düşük, olması araştırmanın dikkat çeken bir

bulgusudur. Tablo 4'te, yanlılıkla birlikte kestirimin standart hatasını da dikkate alan RMSE ve OMH değerleri incelendiğinde ise sınıflama kriteri ve madde seçme yöntemi fark etmeksizin kategori sayısı arttıkça hem RMSE hem de OMH değerlerinin düştüğü görülmektedir. Elde edilen bulgulara göre en yüksek RMSE değeri (0,488) MFB-KN madde seçme yöntemi ile GA sınıflama kriterinin birlikte kullanıldığı ve sınıflamanın iki kategoride yapıldığı koşulda gözlenirken en düşük RMSE değeri (0,293) ise MFB-KY yönteminin AOOT ile birlikte kullanıldığı ve sınıflamanın dört kategoride yapıldığı koşulda gözlenmiştir. Benzer şekilde en yüksek OMH değeri (0,377) MFB-KN ile GA'nın birlikte kullanıldığı ve sınıflamanın iki kategoride yapıldığı koşulda gözlenirken en düşük OMH değeri (0,218) ise MFB-KY ile AOOT'nin birlikte kullanıldığı ve sınıflamanın dört kategoride yapıldığı koşulda gözlenmiştir. Kısacası, AOOT sınıflama kriteri GA sınıflama kriterine göre daha düşük RMSE ve OMH değerleri vermiştir. Buna ek olarak MFB-KY madde seçme yönteminin kullanıldığı koşullarda MFB-KN'nin kullanıldığı koşullara kıyasla daha düşük RMSE ve OMH değerleri elde edilmiştir. Araştırmanın bu bulgusu Gündeğer'in (2017) araştırma sonuçlarıyla da desteklenmektedir.

Özetle, Tablo 4'teki değerler dikkate alınarak, BBST simülasyonunda yetenek kestirim yöntemi AOK iken, madde seçme yöntemi fark etmeksizin, OTU bakımından GA sınıflama kriteri, OSD bakımından ise AOOT sınıflama kriteri daha başarılı bir performans göstermiştir. Ancak elde edilen OSD'lerin her iki sınıflama kriteri için de yeterince yüksek olduğu düşünülecek olursa, test etkililiği (OTU ve OSD) bakımından GA'nın performansının daha iyi olduğu söylenebilir. Bununla birlikte elde edilen bulgular, sınıflama kategori sayısındaki artışın test etkililiğini düşürdüğünü, başka bir ifadeyle OTU'yu artırırken OSD'yi azalttığını göstermektedir. Buna göre, test etkililiği bakımından en iyi sonuçların (OTU= 11,33; OSD= 0,89) gözlemlendiği MFB-KY madde seçme yöntemi ile GA sınıflama kriterinin birlikte kullanıldığı ve sınıflamanın iki kategoride yapıldığı koşulda bireyler ortalama 11 madde ve 0,89 sınıflama doğruluğuyla sınıflanabilirken, en kötü sonuçların (OTU= 42,88 ve OSD= 0,82) gözlemlendiği MFB-KN madde seçme yöntemi ile AOOT sınıflama kriterinin birlikte kullanıldığı ve sınıflamanın dört kategoride yapıldığı koşulda bireyler ortalama 43 madde ve 0,82 sınıflama doğruluğuyla sınıflanabilmiştir. Dolayısıyla bireylerin

sınıflanıp testin sonlanabilmesi için yaklaşık 4 kat daha fazla madde kullanılması gerekmiştir. Bununla birlikte, bireylerin kestirilen ve gerçek yetenek düzeyleri arasındaki korelasyon (r), yanlılık, RMSE ve OMH değerleri bakımından; AOOT sınıflama kriterinin GA'dan, MFB-KY madde seçme yönteminin ise MFB-KN'den daha başarılı bir performans sağlandığı görülmüştür. Öyle ki, AOOT sınıflama kriterinin veya MFB-KY madde seçme yönteminin kullanıldığı koşullarda; daha yüksek korelasyon değerleri, daha düşük yanlılık, RMSE ve OMH değerleri hesaplanmıştır. Buna göre, sınıflama kriteri AOOT olduğunda veya madde seçme yöntemi MFB-KY olduğunda bireylerin son yetenek düzeylerinin daha hassas ve gerçek yetenek düzeylerine daha yakın kestirildiği yorumu yapılabilir. Bu durumun kaynağı olarak, madde havuzunun $\theta = 0$ kesme noktası ve etrafında yüksek bilgi veren maddelerden oluşturulmuş olması görülebilir. Ayrıca tüm koşullarda, sınıflama kategori sayısındaki artışa rağmen korelasyon değerleri benzer sonuçlar verirken yanlılık, RMSE ve OMH değerlerinin genel olarak nispeten düşmüş olması araştırmanın önemli bulguları arasındadır. Buna göre, sınıflama kategori sayısı arttıkça testin sonlanması için gereken madde sayısının da artması sebebiyle, bireylerin son yetenek düzeylerinin daha hassas kestirildiği yorumu yapılabilir.

1.2. MFB-KY, MFB-KN madde seçme yöntemleri ile KBBT ve MMM içerik dengeleme yöntemlerinin bir arada ele alındığı koşullarda OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

Araştırma koşullarına içerik dengelemesi için KBBT veya MMM dahil edildiğinde OTU, OSD, kestirilen ve gerçek yetenekler arasındaki korelasyon (r), yanlılık, RMSE ve OMH değerlerindeki değişim, madde seçme yöntemlerinden MFB-KY ve MFB-KN'ye göre karşılaştırmalı olarak incelenmiştir. KBBT ve MMM içerik dengeleme yöntemlerinin etkililiği, belirlenen dört içerik alanı için istenen içerik oranlarını (birinci içerik alanı için %40, ikinci içerik alanı için %30, üçüncü içerik alanı için %20, dördüncü içerik alanı için %10) sağlama durumları üzerinden değerlendirilmiştir. Dolayısıyla elde edilen değerler dört içerik alanı için istenen içerik oranları ve uygulanan içerik oranlarıyla birlikte Tablo 5'te sunulmuştur. Sunulan değerler, her bir koşul için yapılan 30 tekrarın ortalaması alınarak hesaplanmıştır.

Tablo 5

Yetenek Kestirim Yöntemi AOK Olduğunda Alt Problem 1.2. Koşullarına Ait Uygulanan İçerik Oranı, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri

Sınıflama Kriteri	Koşullar			Bağımlı Değişkenler									
	Madde Seçme Yöntemi	İçerik Dengeleme Yöntemi	Sınıflama Kategori Sayısı	Uygulanan İçerik Oranı				OTU	OSD	r	Yanlılık	RMSE	OMH
				İçerik Alanları İstenen Oranları									
				1 %40	2 %30	3 %20	4 %10						
AOOT δ: 0,20	MFB-KY	KBBT	İki	36,77	29,73	21,54	11,96	25,25	0,94	0,94	-0,011	0,363	0,275
			Üç	34,80	29,13	22,34	13,72	34,90	0,88	0,95	-0,015	0,322	0,241
			Dört	33,63	28,74	22,83	14,79	42,10	0,82	0,96	-0,016	0,297	0,220
		MMM	İki	39,96	29,99	20,04	10,02	25,75	0,94	0,94	-0,011	0,364	0,275
			Üç	39,91	30,03	20,03	10,02	35,00	0,88	0,95	-0,012	0,325	0,242
			Dört	39,97	30,01	19,99	10,03	42,39	0,82	0,96	-0,014	0,299	0,222
	MFB-KN	KBBT	İki	37,77	29,86	21,11	11,27	23,49	0,94	0,90	0,012	0,433	0,324
			Üç	34,98	29,20	22,29	13,53	34,30	0,88	0,92	0,014	0,387	0,283
			Dört	33,43	28,72	22,89	14,96	43,16	0,82	0,93	0,011	0,355	0,258
		MMM	İki	39,85	29,99	20,09	10,07	23,70	0,94	0,90	0,012	0,435	0,326
			Üç	39,96	29,98	20,03	10,02	35,08	0,88	0,92	0,016	0,387	0,284
			Dört	39,96	30,00	20,02	10,02	43,77	0,82	0,93	0,011	0,355	0,259
GA %90 güven düzeyi	MFB-KY	KBBT	İki	42,06	30,90	19,16	7,87	11,23	0,89	0,90	0,012	0,472	0,364
			Üç	41,52	30,80	19,39	8,28	12,32	0,79	0,90	0,014	0,464	0,358
			Dört	41,03	30,69	19,64	8,64	13,42	0,71	0,91	0,010	0,457	0,350
		MMM	İki	40,01	29,93	20,04	10,02	11,25	0,89	0,90	0,010	0,471	0,364
			Üç	40,03	29,96	20,03	9,98	12,30	0,79	0,90	0,013	0,467	0,360
			Dört	39,95	30,02	20,05	9,98	13,42	0,71	0,91	0,012	0,454	0,348
	MFB-KN	KBBT	İki	42,06	30,89	19,12	7,93	11,36	0,90	0,87	0,018	0,491	0,381
			Üç	41,46	30,75	19,44	8,35	12,46	0,80	0,87	0,019	0,487	0,377
			Dört	40,89	30,69	19,69	8,73	13,46	0,70	0,87	0,015	0,479	0,370
		MMM	İki	39,99	30,07	19,94	10,00	11,36	0,89	0,87	0,016	0,494	0,384
			Üç	39,98	30,03	20,01	9,99	12,49	0,80	0,87	0,014	0,483	0,374
			Dört	39,97	29,98	20,07	9,98	13,48	0,71	0,88	0,014	0,477	0,368

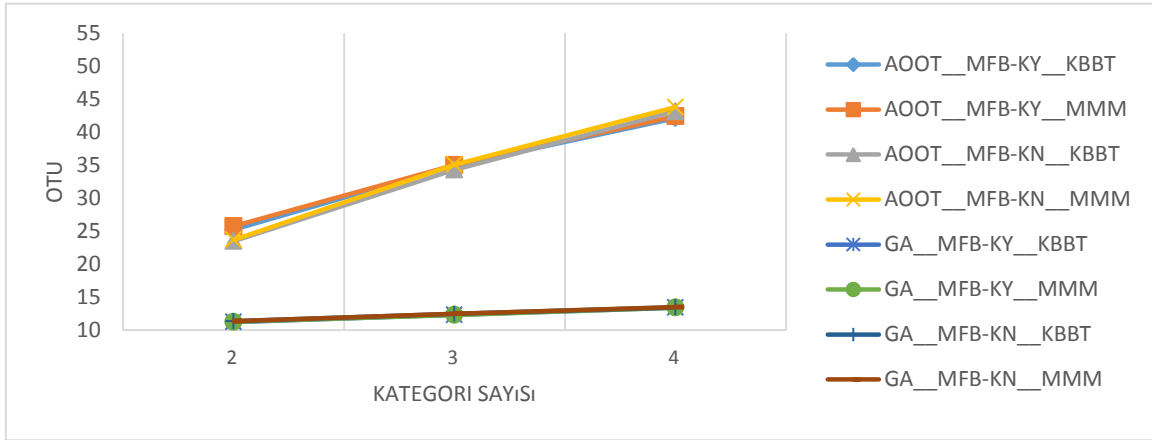
Not: OTU= Ortalama Test Uzunluğu, OSD= Ortalama Sınıflama Doğruluğu, r= Gerçek ve Kestirilen Yetenek Düzeyleri Arasındaki Korelasyon.

Tablo 5'e göre, tüm koşullarda MMM yöntemi KBBT yöntemine kıyasla içerik dengeleme bakımından daha iyi bir performans göstermiştir. Araştırmanın bu bulgusu Lin'in (2011) çalışmasındaki ifadelerle de desteklenmektedir. Lin (2011) BBT çalışmalarında, MMM yönteminin KBBT yöntemindeki içerik sıralamasının yüksek öngörülebilirliğini engellediğini ve istenen içerik dengesini sağlamada daha başarılı olduğunu ifade etmiştir. Tablo 5'e göre MMM içerik dengeleme yöntemi kullanıldığında; sınıflama kriteri, madde seçme yöntemi ve sınıflama kategori sayısı fark etmeksizin uygulanan içerik oranlarının istenen içerik oranlarını (sırasıyla %40, %30, %20 ve %10) sağladığı söylenebilir. KBBT içerik dengeleme yönteminin kullanıldığı koşulların ise genel olarak istenen içerik oranlarının kısmen üstünde veya altında kaldığı görülmektedir. Örneğin Tablo 5'te AOOT ile MFB-KN yöntemlerinin birlikte kullanıldığı ve sınıflamanın dört kategoride yapıldığı koşulda içerik dengelemesi için KBBT tercih edildiğinde, uygulanan içerik oranları sırasıyla yaklaşık %33, %29, %23 ve %15 olarak hesaplanmıştır.

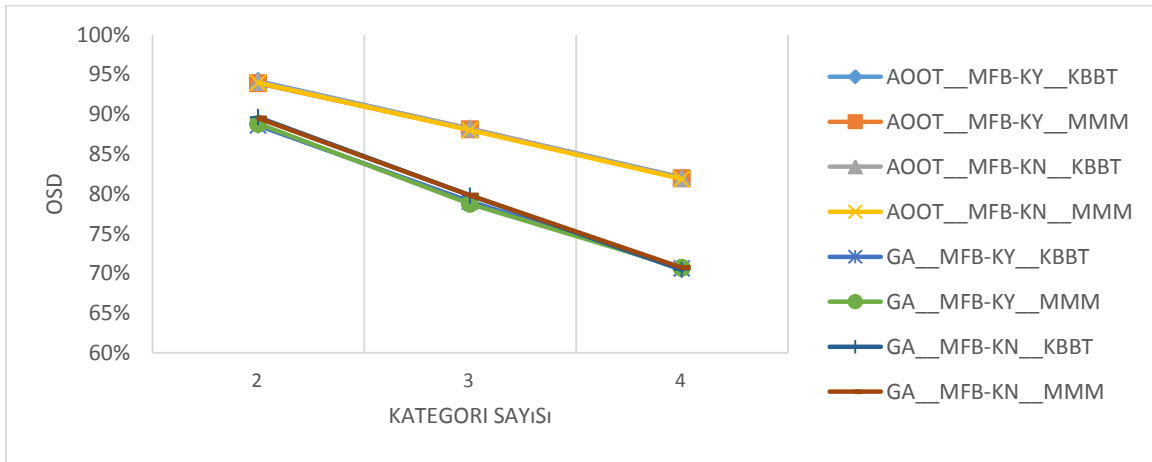
Tablo 5'te, aynı sınıflama kriteri ve aynı madde seçme yöntemi üzerinden bir karşılaştırma yapıldığında, test etkililiği bakımından KBBT'nin çok küçük farklarla daha iyi bir performans gösterdiği ancak genel olarak MMM ile benzer sonuçlar verdikleri görülmektedir. Örneğin, üç kategoride yapılan sınıflamada, AOOT, MFB-KN ve KBBT birlikte kullanıldığında OTU= 34,30 ve OSD= 0,88 olarak, AOOT, MFB-KN ve MMM birlikte kullanıldığında ise OTU= 35,08 ve OSD= 0,88 olarak hesaplanmıştır. Tablo 5 incelendiğinde, içerik dengeleme yöntemi ve madde seçme yöntemi fark etmeksizin en düşük OTU'nun (yaklaşık 11 madde) GA sınıflama kriterinin kullanıldığı koşuldan, en yüksek OSD'nin (0,94) ise AOOT'nin kullanıldığı koşuldan elde edildiği görülmektedir. O halde, Tablo 4'ten elde edilen bulguya paralel olarak, OTU bakımından GA, OSD bakımından ise AOOT daha iyi bir performans göstermiştir. Bunun yanı sıra hesaplanan OSD'lerin her iki sınıflama kriteri için de yeterince yüksek olduğu düşünülecek olursa test etkililiği bakımından GA sınıflama kriterinin daha avantajlı olabileceği söylenebilir. Öyle ki, içerik dengeleme yöntemi fark etmeksizin; GA sınıflama kriteri ile MFB-KY madde seçme yöntemi birlikte kullanıldığında, bireyler iki kategoride ortalama 11 madde ve 0,89 sınıflama doğruluğuyla, üç kategoride ortalama 12 madde ve 0,79 sınıflama doğruluğuyla, dört kategoride ortalama 13

madde ve 0,71 sınıflama doğruluğuyla sınıflanmışken GA ile MFB-KN birlikte kullanıldığında, iki kategoride ortalama 11 madde ve 0,89; 0,90 sınıflama doğruluğuyla, üç kategoride ortalama 12 madde ve 0,80 sınıflama doğruluğuyla, dört kategoride ortalama 13 madde ve 0,70; 0,71 sınıflama doğruluğuyla sınıflanmıştır.

Tablo 5'e göre, OTU ve OSD için farklı koşullardan ayrı ayrı hesaplanan en iyi değerlerin, iki kategoride yapılan sınıflamalardan (OTU için 11,23; OSD için 0,94), en kötü değerlerin ise dört kategoride yapılan sınıflamalardan (OTU için 43,77 ve OSD için 0,70) elde edildiği görülmektedir.



Şekil 9. Sınıflama kriterleri, madde seçme yöntemleri ve içerik dengeleme yöntemlerinin çaprazlandığı OTU değişim grafiği.



Şekil 10. Sınıflama kriterleri, madde seçme yöntemleri ve içerik dengeleme yöntemlerinin çaprazlandığı OSD değişim grafiği.

Tablo 5'te sunulan deęerlerin yanı sıra Őekil 9 ve Őekil 10'daki grafiklerden de grldęi gibi tm koŐullarda, sınıflama kategori sayısı arttıķça OTU artmıŐ OSD ise azalmıŐtır, baŐka bir ifadeyle test etkililięi dŐmŐtr. AraŐtırmanın bu bulgusu Eggen (1999) ile Nydick ve dięerlerinin (2012) araŐtırma sonuēlarıyla da desteklenmektedir. Eggen (1999) ile Nydick ve dięerleri (2012) tarafından yapılan ēalıŐmalarda, bireyler daha fazla kategoride sınıflandıęında OTU artarken OSD'nin azaldıęı sonucuna ulaŐmıŐlardır.

Tablo 5 incelendięinde, test etkililięi bakımından elde edilen en iyi sonuēların (OTU= 11,23 ve OSD= 0,89) GA'nın MFB-KY ve KBBT ile birlikte kullanıldıęı ve sınıflamanın iki kategoride yapıldıęı koŐulda, en kt sonuēların (OTU= 43,77 ve OSD= 0,82) ise AOOT'nin MFB-KN ve MMM ile birlikte kullanıldıęı ve sınıflamanın drt kategoride yapıldıęı koŐulda gzlendięi grlmektedir.

Tablo 5'te sunulan korelasyon (r), yanlılık, RMSE ve OMH deęerleri incelendięinde; Tablo 4'ten elde edilen bulgulara paralel olarak, AOOT'nin GA'dan, MFB-KY'nin ise MFB-KN'den daha baŐarılı bir performans gsterdięi grlmektedir. yle ki, AOOT'nin kullanıldıęı koŐullar iēin hesaplanan korelasyon (r) deęerleri 0,90 ile 0,96 aralıęında iken GA'nın kullanıldıęı koŐullar iēin 0,87 ile 0,91 aralıęında hesaplanmıŐtır. Bununla birlikte, AOOT iēin hesaplanan yanlılık deęerleri (-0,016, 0,011) aralıęında deęiŐirken GA iēin hesaplanan yanlılık deęerleri (0,010, 0,019) aralıęında deęiŐmektedir. Tablo 5'e gre en yksek RMSE deęeri (0,494) GA sınıflama kriterinin MFB-KN madde seēme yntemi ve MMM iēerik dengeleme yntemi ile birlikte kullanıldıęı ve sınıflamanın iki kategoride yapıldıęı koŐulda gzlenirken en dŐk RMSE deęeri (0,297) ise AOOT, MFB-KY, KBBT yntemlerinin birlikte kullanıldıęı ve sınıflamanın drt kategoride yapıldıęı koŐulda gzlenmiŐtir. Benzer Őekilde en yksek OMH deęeri (0,384) GA, MFB-KN, MMM yntemlerinin birlikte kullanıldıęı ve sınıflamanın iki kategoride yapıldıęı koŐulda gzlenirken en dŐk OMH deęeri (0,220) ise AOOT, MFB-KY, KBBT yntemlerinin birlikte kullanıldıęı ve sınıflamanın drt kategoride yapıldıęı koŐulda gzlenmiŐtir. Kısacası, AOOT sınıflama kriterinin veya MFB-KY madde seēme ynteminin kullanıldıęı koŐullarda, daha yksek korelasyon deęerleri, daha dŐk yanlılık, RMSE ve OMH deęerleri elde edilmiŐtir. Buna gre, sınıflama kriteri AOOT olduęunda veya madde

seçme yöntemi MFB-KY olduğunda bireylerin son yetenek düzeylerinin daha hassas ve gerçek yetenek düzeylerine daha yakın kestirildiği yorumu yapılabilir. Bu durumun kaynağı olarak, madde havuzunun $\theta = 0$ kesme noktası ve etrafında yüksek bilgi veren maddelerden oluşturulmuş olması görülebilir. Ayrıca tüm koşullar incelendiğinde, sınıflama kategori sayısındaki artışa rağmen korelasyon ve yanlılık değerleri benzer sonuçlar vermişken RMSE ve OMH değerleri nispeten düşmüştür. Buna göre, sınıflama kategori sayısı arttıkça testin sonlanması için gereken madde sayısının da artması sebebiyle, bireylerin son yetenek düzeylerinin daha hassas kestirildiği yorumu yapılabilir.

Tablo 4 ile Tablo 5 karşılaştırıldığında, araştırma koşullarına içerik dengeleme yöntemlerinin dahil edilmesine rağmen, sınıflama kriterleri ve madde seçme yöntemlerinin performansları bakımından paralel sonuçlar elde edildiği söylenebilir. Ayrıca araştırma koşullarına dahil edilen KBBT ve MMM içerik dengeleme yöntemlerinin beklentilere uygun şekilde OTU'yu bir miktar artırdığı (tüm koşullar için ortalama 1 madde), OSD'yi ise etkilemediği görülmüştür. Thompson'ın (2007b) çalışması da elde edilen bu araştırma bulgusunu destekler niteliktedir. Thompson'a (2007b) göre içerik dengeleme ve madde kullanım sıklığı kısıtlamaları genellikle yalnızca OTU'da artışa neden olmaktadır. BBST uygulamasında içerik dengeleme yapıldığında OTU'da gözlenen artışın, bireyin kestirilen geçici yetenek düzeyinde en yüksek bilgiyi sağlayan maddenin uygulanan içerik alanında olmamasından ya da uygulanan içerik alanında birey hakkında yeterince bilgi sağlayan bir maddenin olmamasından kaynaklanmış olabileceği yorumu yapılabilir. Bu sorunun çözümü için her bir içerik alanında madde havuzu yetenek ölçeği boyunca genişletilebilir. Ayrıca Tablo 4 ile Tablo 5 karşılaştırıldığında elde edilen bir diğer önemli bulgu olarak; araştırma koşullarına dahil edilen KBBT ve MMM içerik dengeleme yöntemleri, bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyonu (r) etkilemezken, yanlılığın düşmesine, RMSE ve OMH değerlerinin ise bir miktar artmasına neden olmuştur.

1.3. *MFB-KY, MFB-KN madde seçme yöntemleri ile SH ve MU madde kullanım sıklığı kontrolü yöntemlerinin bir arada ele alındığı koşullarda OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?*

Araştırma koşullarına madde kullanım sıklığı kontrolü için SH veya MU dahil edildiğinde OTU, OSD, kestirilen ve gerçek yetenekler arasındaki korelasyon (r), yanlılık, RMSE ve OMH değerlerindeki değişim, madde seçme yöntemlerinden MFB-KY ve MFB-KN'ye göre karşılaştırmalı olarak incelenmiştir. Elde edilen değerler maksimum madde kullanım sıklığı oranını ($r^{max}= 0,20$) aşan maddelerin yüzdesi (MKSAMY) ve bu maddelerin madde kullanım sıklıklarının ortalaması (MKSAMO) ile birlikte Tablo 6'da sunulmuştur. Sunulan değerler, her bir koşul için yapılan 30 tekrarın ortalaması alınarak hesaplanmıştır.

Tablo 6'ya göre tüm koşullarda MU yöntemi SH yöntemine kıyasla madde kullanım sıklığı kontrolü bakımından daha iyi bir performans göstermiştir. Ayrıca sınıflama kriteri, madde seçme yöntemi ve sınıflama kategori sayısı fark etmeksizin, MU'nun kullanıldığı koşullarda; hem MKSAMY'nin daha düşük olduğu hem de MKSAMO'nun istenen $r^{max}= 0,20$ değerini sağladığı söylenebilir. SH yönteminin kullanıldığı koşullarda ise MKSAMY değerlerinin daha yüksek olduğu ve MKSAMO değerinin ise istenen $r^{max}= 0,20$ değerinin oldukça üstünde olduğu görülmektedir. Örneğin, Tablo 6'da AOOT ile MFB-KY yöntemlerinin birlikte kullanıldığı ve sınıflamanın dört kategoride yapıldığı koşulda SH ile kontrol edilen madde kullanım sıklığı için hesaplanan MKSAMY değeri yaklaşık 0,26 iken MKSAMO değeri 0,30 olarak hesaplanmıştır. Başka bir ifadeyle havuzdaki maddelerin yaklaşık %26'sı maksimum madde kullanım sıklığı oranını ($r^{max}= 0,20$) aşmış ve bu maddelerin madde kullanım sıklığı ortalaması yaklaşık 0,30 olarak hesaplanmıştır.

Tablo 6

Yetenek Kestirim Yöntemi AOK Olduğunda Alt Problem 1.3. Koşullarına Ait MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri

Sınıflama Kriteri	Madde Seçme Yöntemi	Koşullar		Sınıflama Kategorisi Sayısı	Bağımlı Değişkenler						
		Madde Kullanım Sıklığı Yöntemi	Kontrol		MKSAMY	MKSAMO	OTU	OSD	r	Yanlılık	RMSE
AOOT δ: 0,20	MFB-KY	SH ($r^{max}=0,20$)	İki	0,138	0,278	26,04	0,94	0,94	-0,012	0,359	0,272
			Üç	0,205	0,287	36,57	0,88	0,95	-0,014	0,319	0,239
			Dört	0,258	0,293	44,31	0,81	0,96	-0,017	0,295	0,220
		MU ($r^{max}=0,20$)	İki	0,092	0,201	29,13	0,93	0,94	-0,014	0,356	0,271
			Üç	0,147	0,202	40,77	0,86	0,95	-0,017	0,323	0,244
			Dört	0,181	0,202	47,97	0,78	0,96	-0,017	0,304	0,230
	MFB-KN	SH ($r^{max}=0,20$)	İki	0,123	0,339	25,26	0,94	0,90	0,014	0,435	0,326
			Üç	0,215	0,341	37,78	0,87	0,92	0,011	0,387	0,285
			Dört	0,250	0,366	46,54	0,80	0,93	0,009	0,358	0,262
		MU ($r^{max}=0,20$)	İki	0,114	0,202	29,55	0,93	0,90	0,007	0,434	0,328
			Üç	0,184	0,202	43,04	0,85	0,92	0,006	0,392	0,293
			Dört	0,231	0,202	50,70	0,76	0,93	0,003	0,373	0,278
GA %90 güven düzeyi	MFB-KY	SH ($r^{max}=0,20$)	İki	0,052	0,282	11,35	0,89	0,90	0,019	0,467	0,360
			Üç	0,055	0,282	12,53	0,79	0,90	0,016	0,459	0,353
			Dört	0,060	0,283	14,08	0,70	0,91	0,015	0,449	0,343
		MU ($r^{max}=0,20$)	İki	0,032	0,202	11,36	0,88	0,90	0,017	0,479	0,370
			Üç	0,031	0,202	12,67	0,78	0,90	0,017	0,468	0,360
			Dört	0,037	0,202	14,13	0,69	0,90	0,020	0,461	0,352
	MFB-KN	SH ($r^{max}=0,20$)	İki	0,050	0,407	11,55	0,89	0,87	0,011	0,495	0,385
			Üç	0,054	0,401	12,72	0,79	0,87	0,010	0,489	0,379
			Dört	0,058	0,395	14,05	0,70	0,88	0,013	0,480	0,371
		MU ($r^{max}=0,20$)	İki	0,050	0,203	11,71	0,89	0,86	0,011	0,511	0,400
			Üç	0,052	0,203	13,24	0,78	0,87	0,005	0,500	0,390
			Dört	0,057	0,203	15,81	0,66	0,87	0,009	0,486	0,376

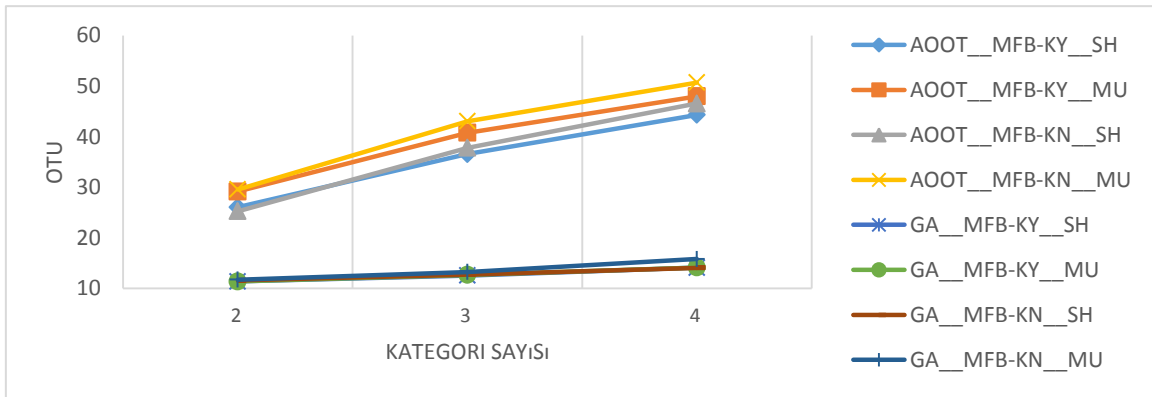
Not: MKSAMY= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Yüzdesi, MKSAMO= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Madde Kullanım Sıklıklarının Ortalaması, OTU= Ortalama Test Uzunluğu, OSD= Ortalama Sınıflama Doğruluğu, r= Gerçek ve Kestirilen Yetenek Düzeyleri Arasındaki Korelasyon.

Tablo 6 incelendiğinde, madde kullanım sıklığı kontrol yöntemi ve madde seçme yöntemi fark etmeksizin en düşük OTU'nun (yaklaşık 11 madde) GA sınıflama kriterinin kullanıldığı koşuldan, en yüksek OSD'nin (0,94) ise AOOT'nin kullanıldığı koşuldan elde edildiği görülmektedir. Dolayısıyla, Tablo 4'ten elde edilen bulguya paralel olarak, OTU bakımından GA sınıflama kriterinin, OSD bakımından ise AOOT sınıflama kriterinin daha iyi performans gösterdiği söylenebilir. Ancak OTU ve OSD'nin bir bütün olarak ele alındığı test etkililiği bakımından GA sınıflama kriterinin kısmen daha avantajlı olduğu yorumu yapılabilir. Tablo 6'ya göre, aynı sınıflama kriteri ve aynı madde seçme yöntemi üzerinden bir karşılaştırma yapıldığında, özellikle AOOT sınıflama kriterinin tercih edildiği durumda, MU madde kullanım sıklığı kontrol yönteminin SH'ye kıyasla OTU ve OSD'yi daha olumsuz etkilediği söylenebilir. Örneğin, iki kategoride yapılan sınıflamada, AOOT, MFB-KN ve SH birlikte kullanıldığında OTU= 26,04 ve OSD= 0,94 olarak, AOOT, MFB-KN ve MU birlikte kullanıldığında ise OTU= 29,13 ve OSD= 0,93 olarak hesaplanmıştır. Buna göre, MU yöntemi madde kullanım sıklığını daha iyi kontrol etmesine rağmen test etkililiği üzerinde SH'ye kıyasla daha olumsuz bir etkiye sahiptir. Öyle ki, MU'nun tercih edildiği koşullarda SH'ye kıyasla daha yüksek OTU ve daha düşük OSD'ler elde edilmiştir. Araştırmanın bu bulgusu Huebner'in (2012) araştırma sonuçlarıyla da desteklenmektedir. Huebner (2012) tarafından yapılan çalışmada, madde kullanım sıklığı kontrolü bakımından MU yönteminin SH'den daha başarılı çalıştığı, buna karşılık MU yöntemi kullanıldığında SH yöntemine kıyasla testin benzer OTU ve daha düşük OSD ile sonlandığı ifade edilmiştir.

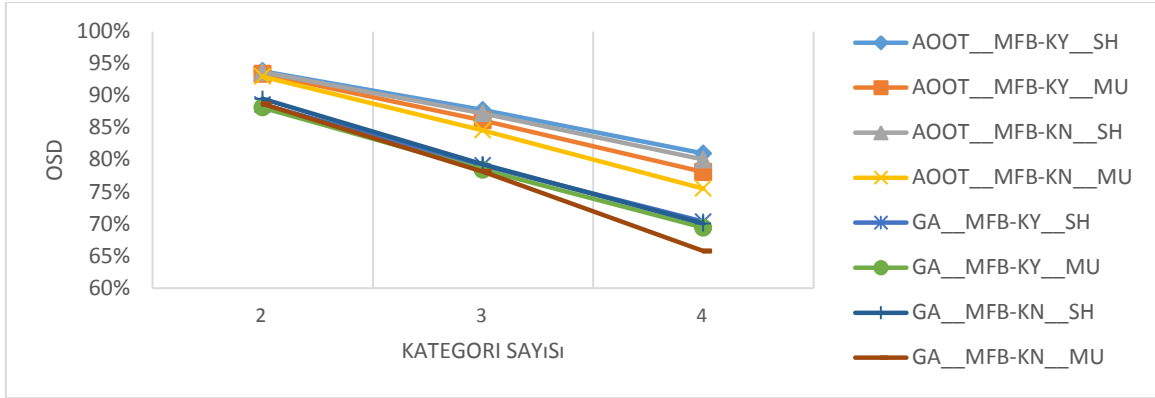
Tablo 6'ya göre madde seçme yöntemi ve madde kullanım sıklığı kontrol yöntemi fark etmeksizin, AOOT üzerinden hesaplanan MKSAMY değerleri yaklaşık %10 ile %25 arasında, MKSAMO değerleri ise 0,20 ile 0,37 arasında değişirken, GA üzerinden hesaplanan MKSAMY değerleri yaklaşık %3 ile %6 arasında, MKSAMO değerleri ise 0,20 ile 0,41 arasında değişmektedir. Bunun yanı sıra Tablo 6 incelendiğinde, SH'nin GA ile birlikte kullanıldığı koşullarda, AOOT ile birlikte kullanıldığı koşullara kıyasla daha düşük MKSAMY ancak daha yüksek MKSAMO değerleri hesaplandığı görülmektedir. Bu demek oluyor ki, SH yöntemi GA sınıflama kriteriyle birlikte kullanıldığında daha az sayıda madde, kullanım sıklığı oranını

aşmasına (daha düşük MKSAMY) rağmen aynı maddelerin çok fazla kullanılmasından dolayı MKSAMYO değeri oldukça yükselmektedir. Öyle ki, SH yönteminin kullanıldığı koşullarda; en yüksek MKSAMO değeri ortalama 0,41 (MKSAMY ise 0,05) olarak GA sınıflama kriteri ve MFB-KN madde seçme yönteminin birlikte kullanıldığı iki kategorili sınıflamalarda hesaplanmış iken en düşük MKSAMO değeri ortalama 0,28 (MKSAMY ise 0,14) olarak AOOT ve MFB-KY yöntemlerinin birlikte kullanıldığı iki kategorili sınıflamalarda hesaplanmıştır. Buna karşılık Tablo 6'ya göre, madde kullanım sıklığı MU ile kontrol edildiğinde, madde seçme yöntemi fark etmeksizin, AOOT ve GA yöntemleri için birbirine benzer MKSAMO değerleri ve SH'ye kıyasla daha düşük MKSAMY değerleri hesaplandığı görülmektedir.

Tablo 6'ya göre test etkililiği bakımından elde edilen en iyi sonuçların (OTU= 11,35 ve OSD= 0,89) GA'nın MFB-KY ve SH ile birlikte kullanıldığı ve sınıflamanın iki kategoride yapıldığı koşulda, en kötü sonuçların (OTU= 50,70 ve OSD= 0,76) ise AOOT'nin MFB-KN ve MMM ile birlikte kullanıldığı ve sınıflamanın dört kategoride yapıldığı koşulda gözlemlendiği görülmektedir. Tablo 6'da görülen değerlerin yanı sıra Şekil 11 ve Şekil 12'de sunulan grafiklerden de görüldüğü gibi genel olarak sınıflama kriteri, madde seçme yöntemi ve madde kullanım sıklığı kontrol yöntemi fark etmeksizin, sınıflama kategori sayısı arttıkça OTU'nun arttığı OSD'nin ise azaldığı görülmektedir. Eggen (1999) ile Nydick ve diğerlerinin (2012) tarafından yapılan çalışmalarda da araştırmanın bu bulgusunu destekler nitelikte, kategori sayısındaki artışın OTU'yu artırırken OSD'yi azalttığı görülmüştür.



Şekil 11. Sınıflama kriterleri, madde seçme yöntemleri ve madde kullanım sıklığı kontrol yöntemlerinin çaprazlandığı OTU değişim grafiği.



Şekil 12. Sınıflama kriterleri, madde seçme yöntemleri ve madde kullanım sıklığı kontrol yöntemlerinin çaprazlandığı OSD değişim grafiği.

Tablo 6'da sunulan korelasyon (r), yanlılık, RMSE ve OMH değerleri incelendiğinde; Tablo 4'ten elde edilen bulgulara paralel olarak, AOOT'nin GA'dan, MFB-KY'nin ise MFB-KN'den daha başarılı bir performans gösterdiği görülmektedir. Öyle ki, AOOT'nin kullanıldığı koşullar için hesaplanan korelasyon (r) değerleri 0,90 ile 0,96 aralığında iken GA'nın kullanıldığı koşullar için 0,86 ile 0,91 aralığında hesaplanmıştır. Bunun yanı sıra, AOOT için hesaplanan yanlılık değerleri (-0,017, 0,014) aralığında değişirken GA için hesaplanan yanlılık değerleri (0,005, 0,020) aralığında değişmektedir. Tablo 6'ya göre en yüksek RMSE değeri (0,511) GA, MFB-KN, MU yöntemlerinin birlikte kullanıldığı ve sınıflamanın iki kategoride yapıldığı koşulda gözlenirken en düşük RMSE değeri (0,295) ise AOOT, MFB-KY, SH yöntemlerinin birlikte kullanıldığı ve sınıflamanın dört kategoride yapıldığı koşulda gözlenmiştir. Benzer şekilde en yüksek OMH değeri (0,400) GA, MFB-KN, MU yöntemlerinin birlikte kullanıldığı ve sınıflamanın iki kategoride yapıldığı koşulda gözlenirken en düşük OMH değeri (0,220) ise AOOT, MFB-KY, SH yöntemlerinin birlikte kullanıldığı ve sınıflamanın dört kategoride yapıldığı koşulda gözlenmiştir. Özetle, AOOT sınıflama kriterinin veya MFB-KY madde seçme yönteminin kullanıldığı koşullarda, daha yüksek korelasyon değerleri, daha düşük yanlılık, RMSE ve OMH değerleri elde edilmiştir. Dolayısıyla AOOT sınıflama kriteri veya MFB-KY madde seçme yöntemi kullanıldığında bireylerin son yetenek düzeylerinin daha hassas kestirilmiş olduğu yorumu yapılabilir. Bu durumun kaynağı olarak, madde havuzunun $\theta = 0$ kesme noktası ve etrafında yüksek bilgi veren maddelerden oluşturulmuş olması

görülebilmektedir. Ayrıca tüm koşullarda, artan sınıflama kategori sayısına karşılık, korelasyon ve yanlılık değerleri benzer sonuçlar vermişken RMSE ve OMH değerleri nispeten düşmüştür. Buna göre artan sınıflama kategori sayısının, OTU'yu ve dolayısıyla yetenek kestirimi için kullanılan madde sayısını arttırması sebebiyle son yetenek düzeylerinin daha hassas kestirilmesine katkı sağladığı söylenebilir.

Tablo 4 ile Tablo 6 karşılaştırıldığında, araştırma koşullarına madde kullanım sıklığı kontrol yöntemlerinin dahil edilmesine rağmen, sınıflama kriterleri ve madde seçme yöntemlerinin performansları bakımından paralel sonuçlar elde edildiği söylenebilir. Ayrıca araştırma koşullarına dahil edilen SH ve MU madde kullanım sıklığı kontrol yöntemlerinin beklentilere uygun şekilde OTU'yu bir miktar artırdığı, OSD'yi ise bir miktar düşürdüğü görülmüştür. Elde edilen bu bulgu Thompson'ın (2007b) çalışmasıyla da kısmen desteklenmektedir. Thompson'a (2007b) göre içerik dengeleme ve madde kullanım sıklığı kısıtlamaları genellikle yalnızca OTU'da artışa neden olmaktadır. BBST uygulamasında madde kullanım sıklığı kontrolü yapıldığında OTU'da gözlenen artış ve OSD'de gözlenen düşüşün, bireyin kestirilen geçici yetenek düzeyinde en yüksek bilgiyi sağlayan maddenin, maksimum madde kullanım sıklığı oranına ulaşmış olmasından dolayı bireye uygulanamamasından kaynaklanmış olabileceği yorumu yapılabilir. Bu sorunun çözümü için maksimum madde kullanım sıklığını aşan maddelerin yoğunlaştığı yetenek ranjındaki madde sayıları arttırılarak madde havuzu genişletilebilir. Ayrıca Tablo 4 ile Tablo 6 karşılaştırıldığında elde edilen bir diğer önemli bulgu olarak; araştırma koşullarına dahil edilen SH ve MU madde kullanım sıklığı kontrol yöntemlerinin, gerçek ve kestirilen yetenekler arasındaki korelasyonu (r) etkilemezken, yanlılığı düşürdüğü, RMSE ve OMH değerlerini ise bir miktar artırdığı görülmüştür. Tablo 4, Tablo 5 ve Tablo 6 karşılaştırıldığında ise araştırma koşullarına dahil edilen madde kullanım sıklığı kontrolünün içerik dengelemeye kıyasla test etkililiğini daha olumsuz etkilediği, başka bir ifadeyle OTU'yu daha fazla artırdığı ve buna ek olarak daha düşük OSD değerleri verdiği görülmüştür. Eggen ve Straetmans'ın (2000) araştırma sonuçları da araştırmanın bu bulgusunu destekler niteliktedir. Eggen ve Straetmans (2000) tarafından bireyleri üç kategoride sınıflamak amacıyla yapılan çalışma sonucunda, GA ve AOOT sınıflama kriterleri için farklı güven düzeyleri ve δ değerlerinde

araştırma koşullarına dahil edilen içerik dengelemenin madde kullanım sıklığı kontrolüne kıyasla sınıflama hatası bakımından daha iyi performans gösterdiği ve OSD'yi daha az düşürdüğü ifade edilmiştir.

1.4. MFB-KY, MFB-KN madde seçme yöntemleri ile KBBT, MMM içerik dengeleme yöntemleri ve SH, MU madde kullanım sıklığı kontrolü yöntemlerinin bir arada ele alındığı koşullarda OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

Araştırma koşullarına içerik dengelemesi için KBBT veya MMM, madde kullanım sıklığı kontrolü için ise SH veya MU yöntemleri dahil edildiğinde OTU, OSD, kestirilen ve gerçek yetenekler arasındaki korelasyon (r), yanlılık, RMSE ve OMH değerlerindeki değişim, madde seçme yöntemlerinden MFB-KY ve MFB-KN'ye göre karşılaştırmalı olarak incelenmiştir. Elde edilen değerler dört içerik alanı için istenen içerik oranları (birinci içerik alanı için %40, ikinci içerik alanı için %30, üçüncü içerik alanı için %20, dördüncü içerik alanı için %10), uygulanan içerik oranları, maksimum madde kullanım sıklığı oranını ($r^{max}= 0,20$) aşan maddelerin yüzdesi (MKSAMY) ve bu maddelerin madde kullanım sıklıklarının ortalaması (MKSAMO) ile birlikte Tablo 7'de sunulmuştur. Sunulan değerler, her bir koşul için 30 tekrarın ortalaması alınarak hesaplanmıştır.

Tablo 7'ye göre tüm koşullarda, içerik dengeleme bakımından MMM yöntemi KBBT'ye kıyasla daha iyi bir performans göstermişken, madde kullanım sıklığı kontrolünde MU yönteminin SH'den daha iyi bir performans gösterdiği görülmektedir. Araştırmanın bu bulgusu Lin (2011) ve Huebner'in (2012) çalışmalarıyla da desteklenmektedir. Lin (2011), KBBT'nin BBST'lerde en sık tercih edilen içerik dengeleme yöntemlerinden olmasına rağmen alanyazında daha çok BBT'lerde kullanılmış olan MMM yönteminin istenen içerik dengesini sağlamada daha başarılı olduğunu ifade etmiştir. Huebner (2012) ise çalışmasında, madde kullanım sıklığı kontrolü bakımından MU yönteminin SH'den daha başarılı çalıştığı sonucuna ulaşmıştır.

Tablo 7

Yetenek Kestirim Yöntemi AOK Olduğunda Alt Problem 1.4. Koşullarına Ait Uygulanan İçerik Oranı, MKSAMY, MKSAMO, OTU, OSD, r , Yanlılık, RMSE ve OMH Değerleri

SK	Koşullar				Uygulanan İçerik Oranı	Bağımlı Değişkenler										
	MSY	İDY	MKSKY	SKS		MKS AMY	MKS AMO	OTU	OSD	r	Yanlılık	RMSE	OMH			
	İçerik Alanları İstenen Oranları					1	2	3	4							
					%40	%30	%20	%10								
AOT $\delta: 0,20$	MFB-KY	KBBT	SH $r^{max}=0,20$	İki	36,27	29,61	21,78	12,34	0,138	0,277	26,91	0,94	0,94	-0,014	0,363	0,275
				Üç	34,34	28,95	22,56	14,15	0,208	0,285	37,50	0,87	0,95	-0,015	0,323	0,242
				Dört	33,32	28,65	22,94	15,09	0,253	0,291	44,69	0,81	0,96	-0,017	0,299	0,223
			MU $r^{max}=0,20$	İki	35,56	29,38	22,09	12,97	0,088	0,202	29,87	0,93	0,94	-0,017	0,365	0,277
				Üç	33,71	28,77	22,79	14,74	0,153	0,202	41,71	0,86	0,95	-0,018	0,328	0,248
				Dört	32,81	28,47	23,15	15,57	0,175	0,202	48,50	0,78	0,96	-0,018	0,308	0,232
	MMM	SH $r^{max}=0,20$	İki	39,82	29,98	20,09	10,10	0,143	0,278	27,42	0,94	0,94	-0,015	0,366	0,278	
			Üç	39,91	29,98	20,04	10,07	0,209	0,287	37,86	0,87	0,95	-0,015	0,327	0,245	
			Dört	39,92	29,99	20,05	10,04	0,255	0,294	45,35	0,80	0,96	-0,017	0,301	0,225	
		MU $r^{max}=0,20$	İki	39,80	30,03	20,06	10,11	0,102	0,202	30,82	0,93	0,94	-0,015	0,366	0,278	
			Üç	39,84	30,01	20,08	10,08	0,151	0,202	42,27	0,85	0,95	-0,016	0,334	0,252	
			Dört	39,86	30,01	20,07	10,07	0,172	0,202	49,01	0,77	0,96	-0,018	0,320	0,241	
MFB-KN	KBBT	SH $r^{max}=0,20$	İki	37,00	29,71	21,46	11,83	0,129	0,331	25,70	0,94	0,90	0,009	0,435	0,327	
			Üç	34,41	28,98	22,49	14,12	0,209	0,341	38,35	0,87	0,92	0,009	0,386	0,284	
			Dört	32,96	28,56	23,07	15,41	0,249	0,358	47,02	0,79	0,93	0,006	0,360	0,264	
		MU $r^{max}=0,20$	İki	35,83	29,40	21,96	12,81	0,111	0,202	30,55	0,93	0,90	0,002	0,438	0,330	
			Üç	33,46	28,74	22,90	14,91	0,185	0,202	43,94	0,84	0,92	0,003	0,402	0,298	
			Dört	32,40	28,39	23,30	15,92	0,228	0,203	51,14	0,75	0,93	0,000	0,379	0,281	
	MMM	SH $r^{max}=0,20$	İki	39,81	29,96	20,12	10,11	0,133	0,327	26,18	0,94	0,90	0,009	0,432	0,325	
			Üç	39,88	30,00	20,08	10,04	0,210	0,339	38,61	0,87	0,92	0,008	0,387	0,286	
			Dört	39,92	30,01	20,03	10,05	0,251	0,356	47,36	0,79	0,93	0,006	0,360	0,264	

Tablo 7 (devamı)

GA %90 güven düzeyi	MFB-KY	KBBT	MU $r^{max}=0,20$	İki	39,75	29,95	20,16	10,13	0,118	0,202	31,01	0,93	0,90	0,005	0,440	0,331
				Üç	39,83	30,00	20,07	10,10	0,193	0,202	44,85	0,84	0,92	0,003	0,404	0,300
				Dört	39,84	30,01	20,05	10,09	0,226	0,203	51,93	0,75	0,93	0,006	0,387	0,288
		SH $r^{max}=0,20$	İki	42,13	30,95	19,09	7,83	0,044	0,274	11,13	0,88	0,89	0,011	0,480	0,371	
			Üç	41,60	30,85	19,33	8,22	0,048	0,274	12,24	0,78	0,90	0,014	0,475	0,366	
			Dört	41,01	30,69	19,64	8,66	0,052	0,273	13,37	0,70	0,90	0,013	0,467	0,357	
		MU $r^{max}=0,20$	İki	42,04	30,90	19,17	7,90	0,025	0,202	11,19	0,88	0,89	0,016	0,494	0,381	
			Üç	41,56	30,80	19,40	8,24	0,026	0,202	12,25	0,78	0,89	0,013	0,486	0,374	
			Dört	41,03	30,70	19,64	8,63	0,029	0,202	13,39	0,69	0,90	0,014	0,478	0,367	
	MMM	SH $r^{max}=0,20$	İki	39,97	30,04	19,96	10,04	0,044	0,267	11,18	0,88	0,89	0,015	0,480	0,371	
			Üç	39,95	30,01	20,01	10,02	0,048	0,267	12,28	0,78	0,90	0,011	0,472	0,363	
			Dört	40,07	29,93	20,00	9,99	0,054	0,267	13,45	0,70	0,90	0,012	0,466	0,358	
	MU $r^{max}=0,20$	İki	40,06	29,94	20,01	9,99	0,025	0,201	11,21	0,88	0,89	0,012	0,489	0,378		
		Üç	40,05	30,01	19,95	9,98	0,026	0,201	12,31	0,78	0,90	0,012	0,482	0,372		
		Dört	40,01	30,00	19,99	9,99	0,028	0,201	13,51	0,69	0,90	0,013	0,475	0,366		
	MFB-KN	KBBT	SH $r^{max}=0,20$	İki	41,96	30,93	19,19	7,92	0,053	0,364	11,36	0,89	0,86	0,012	0,503	0,392
				Üç	41,49	30,77	19,40	8,34	0,060	0,355	12,46	0,79	0,87	0,010	0,494	0,385
				Dört	40,88	30,64	19,68	8,80	0,064	0,355	13,69	0,70	0,87	0,010	0,484	0,375
MU $r^{max}=0,20$		İki	42,02	30,90	19,13	7,96	0,045	0,203	11,42	0,88	0,85	0,009	0,522	0,406		
		Üç	41,43	30,75	19,46	8,37	0,048	0,203	12,56	0,77	0,86	0,004	0,511	0,400		
		Dört	40,55	30,55	19,86	9,03	0,053	0,203	14,60	0,66	0,86	0,006	0,505	0,393		
MMM		SH $r^{max}=0,20$	İki	39,97	30,01	20,00	10,01	0,056	0,350	11,37	0,89	0,86	0,012	0,502	0,391	
			Üç	39,91	30,02	20,07	9,99	0,060	0,350	12,56	0,79	0,87	0,009	0,491	0,382	
			Dört	39,97	30,02	20,00	10,01	0,066	0,347	13,65	0,69	0,87	0,013	0,487	0,378	
MU $r^{max}=0,20$	İki	40,00	30,02	19,99	9,98	0,045	0,203	11,49	0,88	0,86	0,006	0,519	0,407			
	Üç	40,04	29,98	20,01	9,97	0,048	0,203	12,74	0,78	0,86	0,008	0,509	0,397			
	Dört	39,99	29,98	20,02	10,00	0,054	0,203	14,42	0,67	0,87	0,009	0,504	0,392			

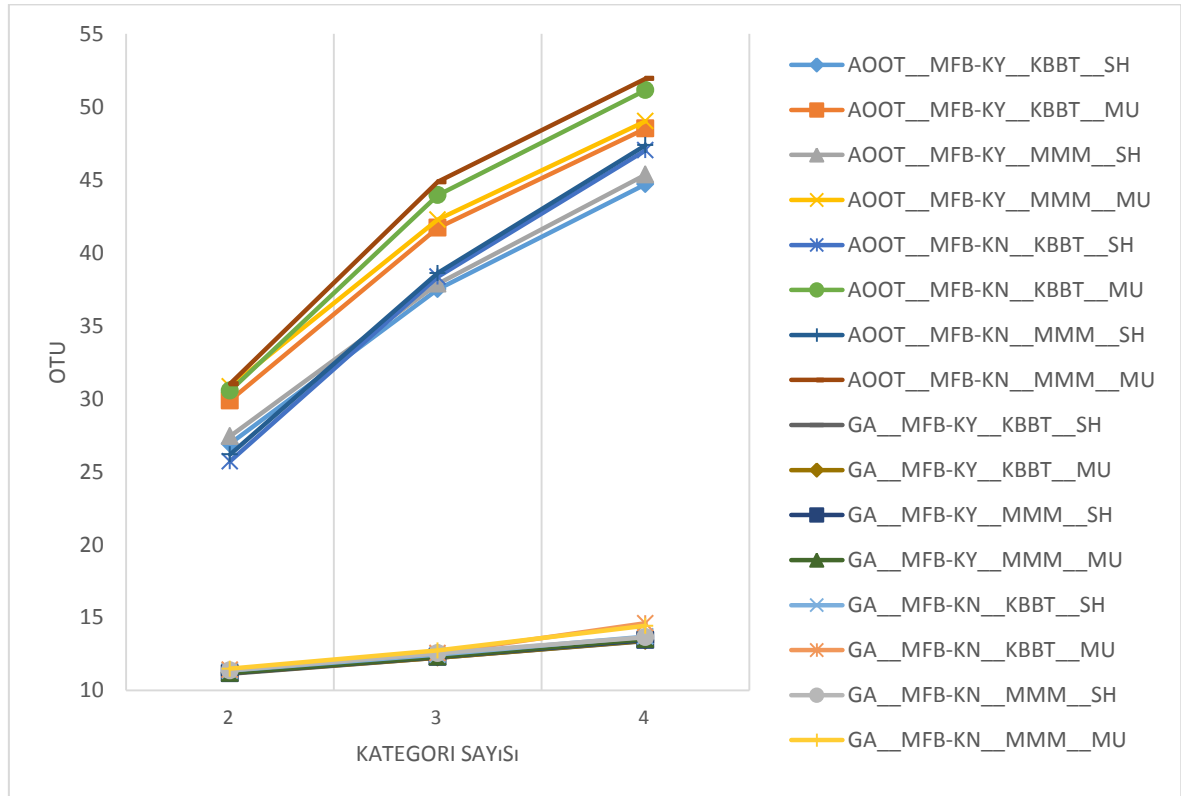
Not: SK= Sınıflama Kriteri, MSY= Madde Seçme Yöntemi, İDY= İçerik Dengeleme Yöntemi, MKSKY= Madde Kullanım Sıklığı Kontrol Yöntemi, SKS= Sınıflama Kategori Sayısı, MKSAM Y= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Yüzdesi, MKSAMO= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Madde Kullanım Sıklıklarının Ortalaması, OTU= Ortalama Test Uzunluğu, OSD= Ortalama Sınıflama Doğruluğu, r= Gerçek ve Kestirilen Yetenek Düzeyleri Arasındaki Korelasyon.

Tablo 7'ye göre MMM içerik dengeleme yönteminin kullanıldığı tüm koşullarda; uygulanan içerik oranlarının istenen içerik oranlarını (sırasıyla %40, %30, %20 ve %10) sağladığı görülmektedir. KBBT içerik dengeleme yönteminin kullanıldığı koşullarda ise Tablo 5'ten elde edilen bulgulara paralel olarak, uygulanan içerik oranlarının istenen içerik oranlarının kısmen üstünde veya altında değerler aldığı görülmektedir. Örneğin Tablo 7'de AOOT ile MFB-KN yöntemlerinin birlikte kullanıldığı, madde kullanım sıklığının MU ile kontrol edildiği ve sınıflamanın dört kategoride yapıldığı koşulda KBBT ile uygulanan içerik oranları sırasıyla yaklaşık %32, %28, %23 ve %16 olarak hesaplanmıştır. Ayrıca Tablo 7'ye göre MU madde kullanım sıklığı kontrol yönteminin kullanıldığı koşullarda; MKSAMY daha düşükken MKSAMO'nun istenen $r^{max}=0,20$ değerini sağladığı söylenebilir. SH madde kullanım sıklığı kontrol yönteminin kullanıldığı koşullarda ise Tablo 6'dan elde edilen bulgulara paralel olarak, MKSAMY değerlerinin daha yüksek olduğu ve MKSAMO değerlerinin ise istenen $r^{max}= 0,20$ değerinin oldukça üstünde olduğu görülmektedir. Örneğin Tablo 7'de AOOT ile MFB-KY yöntemlerinin birlikte kullanıldığı, içerik dengelemenin KBBT ile yapıldığı ve sınıflamanın dört kategoride yapıldığı koşulda SH ile kontrol edilen madde kullanım sıklığı için hesaplanan MKSAMY değeri yaklaşık 0,25 iken MKSAMO değeri 0,30 olarak hesaplanmıştır. Başka bir ifadeyle maddelerin yaklaşık %25'i maksimum madde kullanım sıklığı oranını ($r^{max}= 0,20$) aşmış ve bu maddelerin madde kullanım sıklığı ortalaması yaklaşık 0,30 olarak hesaplanmıştır.

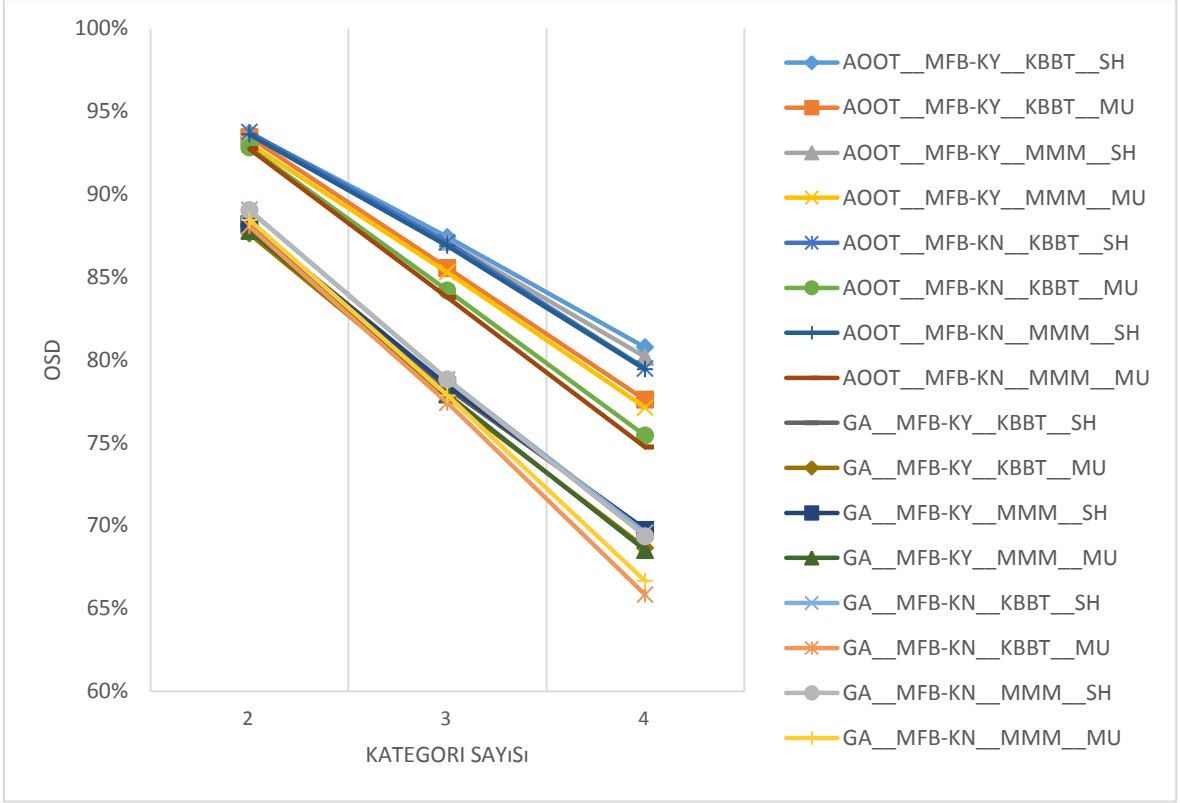
Tablo 7'ye göre, aynı sınıflama kriteri ve aynı madde seçme yöntemi üzerinden bir karşılaştırma yapıldığında, test etkililiği bakımından çok küçük farklarla KBBT içerik dengeleme yöntemi daha iyi performans göstermiş olsa da genel olarak MMM ile benzer sonuçlar verdikleri, buna karşılık SH madde kullanım sıklığı kontrol yönteminin MU'ya kıyasla daha iyi bir performans gösterdiği görülmektedir. Tablo 7'deki sınıflama kriterleri test etkililiği bakımından karşılaştırıldığında, AOOT sınıflama kriterinin tercih edildiği ve sınıflamanın iki kategoride yapıldığı durumda, MFB-KN madde seçme yöntemi ile KBBT içerik dengeleme yöntemi ve SH madde kullanım sıklığı kontrol yönteminin birlikte kullanıldığı koşullarda en iyi performansın (OTU= 25,70 OSD= 0,94) gözleendiği, MFB-KN, MMM ve MU yöntemlerinin birlikte kullanıldığı koşullarda ise en kötü performansın (OTU= 31,01 OSD= 0,93) gözleendiği görülmektedir. Buna karşılık

GA sınıflama kriterinin tercih edildiği ve sınıflamanın iki kategoride yapıldığı durumda, MFB-KY, KBBT ve SH yöntemlerinin birlikte kullanıldığı koşullarda en iyi performansın (OTU= 11,13 OSD= 0,88) gözleendiği, MFB-KN, MMM ve MU yöntemlerinin birlikte kullanıldığı koşullarda ise en kötü performansın (OTU= 11,49 OSD= 0,88) gözleendiği görülmektedir. Buna göre Tablo 4'ten elde edilen bulguya paralel olarak, OTU bakımından GA, OSD bakımından ise AOOT'nin daha iyi performans gösterdiği görülmüştür. Ancak hesaplanan OSD'lerin her iki sınıflama kriteri için de yeterince yüksek olduğu düşünülecek olursa test etkililiği bakımından GA sınıflama kriterinin daha avantajlı olabileceği söylenebilir.

Tablo 7'de görülen değerlerin yanı sıra Şekil 13 ve Şekil 14'te sunulan grafiklerden de görüldüğü gibi genel olarak tüm koşullarda, sınıflama kategori sayısı arttıkça OTU'nun arttığı ve OSD'nin ise azaldığı görülmektedir. Eggen (1999) ile Nydick ve diğerlerinin (2012) tarafından yapılan çalışmalarda da araştırmanın bu bulgusunu destekler nitelikte, kategori sayısındaki artışın OTU'yu artırırken OSD'yi azalttığı görülmüştür.



Şekil 13. Sınıflama kriterleri, madde seçme yöntemleri, içerik dengeleme yöntemleri ve madde kullanım sıklığı kontrol yöntemlerinin çaprazlandığı OTU değişim grafiği.



Şekil 14. Sınıflama kriterleri, madde seçme yöntemleri, içerik dengeleme yöntemleri ve madde kullanım sıklığı kontrol yöntemlerinin çaprazlandığı OSD değişim grafiği.

Test etkililiği bakımından elde edilen en iyi sonuçların (OTU= 11,13 ve OSD= 0,88) GA ve MFB-KY'nin KBBT ve SH ile birlikte kullanıldığı ve sınıflamanın iki kategoride yapıldığı koşulda, en kötü sonuçların (OTU= 51,93 ve OSD= 0,75) ise AOOT ve MFB-KN'nin, MMM ve MU ile birlikte kullanıldığı ve sınıflamanın dört kategoride yapıldığı koşulda gözlemlendiği görülmektedir. Başka bir ifadeyle, en iyi ve en kötü sonuçlar arasında OTU'nun yaklaşık 5 katına çıktığı söylenebilir.

Tablo 7'de sunulan korelasyon (r), yanlılık, RMSE ve OMH değerleri incelendiğinde; Tablo 4'ten elde edilen bulgulara paralel olarak AOOT'nin GA'dan, MFB-KY'nin ise MFB-KN'den daha başarılı bir performans gösterdiği söylenebilir. AOOT'nin kullanıldığı koşullar için hesaplanan korelasyon (r) değerleri 0,90 ile 0,96 aralığında iken GA'nın kullanıldığı koşullar için 0,85 ile 0,90 aralığında hesaplanmıştır. Tablo 7'de görülen yanlılık değerlerinin ise AOOT'nin kullanıldığı koşullarda (-0,014, 0,009) aralığında değişirken GA'nın kullanıldığı koşullarda (0,011, 0,016) aralığında değiştiği görülmektedir. Tablo 7'ye göre en yüksek

RMSE deęeri (0,504) GA, MFB-KN, MMM, MU yöntemlerinin birlikte kullanıldığı ve sınıflamanın iki kategoride yapıldığı koşulda gözlenirken en düşük RMSE deęeri (0,299) ise AOOT, MFB-KY, KBBT, MU yöntemlerinin birlikte kullanıldığı ve sınıflamanın dört kategoride yapıldığı koşulda gözlenmiştir. Benzer şekilde en yüksek OMH deęeri (0,407) GA, MFB-KN, MMM, MU yöntemlerinin birlikte kullanıldığı ve sınıflamanın iki kategoride yapıldığı koşulda gözlenirken en düşük OMH deęeri (0,223) ise AOOT, MFB-KY, KBBT, SH yöntemlerinin birlikte kullanıldığı ve sınıflamanın dört kategoride yapıldığı koşulda gözlenmiştir. Kısacası, AOOT sınıflama kriterinin veya MFB-KY madde seçme yönteminin kullanıldığı koşullarda daha yüksek korelasyon deęerleri, daha düşük yanlılık, RMSE ve OMH deęerleri hesaplanmıştır. Elde edilen bu bulgu Tablo 4, Tablo 5 ve Tablo 6'da sunulan bulgularla paralellik göstermektedir. Buna göre, sınıflama kriteri AOOT olduğunda veya madde seçme yöntemi MFB-KY olduğunda bireylerin son yetenek düzeylerinin daha hassas ve gerçek yetenek düzeylerine daha yakın kestirildiği yorumu yapılabilir. Bu durumun kaynağı olarak, madde havuzunun $\theta = 0$ kesme noktası ve etrafında yüksek bilgi veren maddelerden oluşturulmuş olması görülebilir. Ayrıca tüm koşullar incelendiğinde, artan sınıflama kategori sayısına karşılık, korelasyon ve yanlılık deęerleri benzer sonuçlar vermişken RMSE ve OMH deęerleri nispeten düşmüştür. Buna göre artan sınıflama kategori sayısının OTU'yu ve dolayısıyla yetenek kestirimi için kullanılan madde sayısını arttırması sebebiyle son yetenek düzeylerinin daha hassas kestirilmesine katkı sağladığı söylenebilir.

Tablo 4, Tablo 5, Tablo 6 ve Tablo 7'de sunulan bulgular karşılaştırıldığında; test etkililiği üzerinde, madde kullanım sıklığı kontrol yöntemlerinin görel olarak içerik dengeleme yöntemlerinden daha olumsuz bir etkiye sahip olduğu buna ek olarak hem içerik dengeleme hem de madde kullanım sıklığı kontrolü araştırma koşullarında dahil edildiğinde beklentiye uygun olarak kısmen daha yüksek OTU ve daha düşük OSD elde edildiği görülmüştür. BBST uygulamasında içerik dengeleme ve madde kullanım sıklığı kontrolü yapıldığında OTU'da gözlenen artış ve OSD'de gözlenen düşüşün, uygulanan içerik alanında birey hakkında yeterince bilgi sağlayan ve madde kullanım sıklığını oranını aşmamış bir maddenin bulunmamasından kaynaklanmış olabileceği yorumu yapılabilir. Bu sorunun çözümü için her bir içerik alanında maksimum madde

kullanım sıklığını aşan maddelerin yoğunlaştığı yetenek ranjındaki madde sayıları artırılarak madde havuzu genişletilebilir. Bunun yanı sıra içerik dengeleme yöntemlerinden MMM, istenen içerik oranını sağlamada KBBT'ye kıyasla daha iyi performans gösterirken test etkililiği açısından az da olsa daha kötü bir performans göstermiştir. Buna göre BBST uygulamasında içerik dengeleme daha kritik bir öneme sahipse MMM yönteminin tercih edilmesinin, test etkililiği daha kritik bir öneme sahipse de KBBT yönteminin tercih edilmesinin daha avantajlı olacağı söylenebilir. Madde kullanım sıklığı kontrol yöntemlerinden MU, SH'ye kıyasla madde kullanım sıklığı kontrolünde daha iyi performans gösterirken test etkililiği açısından daha kötü bir performans göstermiştir. Buna göre, BBST uygulamasında madde kullanım sıklığının kontrolü daha kritik bir öneme sahipse MU yönteminin tercih edilmesinin, test etkililiği daha kritik bir öneme sahipse de SH yönteminin tercih edilmesinin daha avantajlı olacağı söylenebilir. Test etkililiğiyle ilgili olarak bahsedilen performans farklılığı AOOT sınıflama kriterinin kullanıldığı koşullarda daha belirginken GA sınıflama kriterinin kullanıldığı koşullarda çok küçük farklılıklar olarak göze çarpmaktadır. GA'nın bu bakımdan da daha avantajlı olduğu söylenebilir. Buna ek olarak, Tablo 4 ile Tablo 7 arasında yapılan karşılaştırma sonucunda, araştırma koşullarına dahil edilen içerik dengeleme ve madde kullanım sıklığı kontrol yöntemlerinin, araştırmanın bağımlı değişkenlerinden; korelasyon (r), yanlılık, RMSE ve OMH değerlerini çok az etkilediği söylenebilir. Öyle ki, araştırma koşullarına KBBT ve MMM içerik dengeleme yöntemleri ile SH ve MU madde kullanım sıklığı kontrol yöntemleri dahil edildiğinde korelasyon değerlerinin etkilenmediği, yanlılığın düştüğü, RMSE ve OMH değerlerinin ise bir miktar arttığı gözlenmiştir. Tek başına madde kullanım sıklığı kontrol yöntemleri içerik dengeleme yöntemlerine kıyasla daha büyük değişime sebep olsa da her iki kısıtlamanın da dahil edildiği Tablo 7'de en düşük yanlılık, en yüksek RMSE ve OMH değerleri hesaplanmıştır. Bununla birlikte, bireylerin kestirilen ve gerçek yetenek düzeyleri arasındaki korelasyon (r), yanlılık, RMSE ve OMH değerleri bakımından; AOOT sınıflama kriterinin kullanıldığı koşullarda, GA sınıflama kriterinin kullanıldığı koşullara kıyasla görece olarak daha başarılı bir performans gözlemlendiği söylenebilir. Bu doğrultuda, BBST uygulamasının odak noktası OSD iken içerik dengeleme ve madde kullanım sıklığı kontrolü de kritik öneme sahipse, Tablo 7'ye göre korelasyon, yanlılık, RMSE ve OMH değerleri bakımından da daha iyi performans gösteren AOOT sınıflama

kriterinin, MFB-KY madde seçme yöntemi, MMM içerik dengeleme yöntemi ve MU madde kullanım sıklığı kontrol yöntemiyle birlikte kullanıldığı koşulların en avantajlı olacağı; bunun yanı sıra BBST uygulamasının odak noktası test etkililiği (OTU ve OSD) iken AOOT, MFB-KY, KBBT ve SH yöntemlerinin birlikte kullanıldığı koşulların en avantajlı olacağı söylenebilir. Buna karşılık BBST uygulamasının odak noktası OTU iken içerik dengeleme ve madde kullanım sıklığı kontrolü de yapılacaksa bütün koşullarda OTU bakımından benzer ve AOOT'den daha başarılı performans gösteren GA sınıflama kriteri ile MFB-KY, MMM ve MU yöntemlerinin birlikte kullanımının en avantajlı olacağı söylenebilir.

Alt Problem 2'ye İlişkin Bulgular ve Yorumlar

AOOT ($\delta: 0,20$) ve GA (%90) sınıflama kriterlerine göre OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

Araştırmanın ikinci alt probleminde, bağımsız değişkenlerden AOOT ve GA sınıflama kriterleri; OTU, OSD, korelasyon (r), yanlılık, RMSE ve OMH değerleri bakımından incelenmiştir. Sınıflama kriterlerine göre bağımlı değişkenlerdeki değişim; dört içerik alanı için istenen içerik oranları, uygulanan içerik oranları, maksimum madde kullanım sıklığı oranını ($r^{max}=0,20$) aşan maddelerin yüzdesi (MKSAMY) ve bu maddelerin madde kullanım sıklıklarının ortalaması (MKSAMO) ile birlikte Tablo 8'de sunulmuştur. Sunulan değerlerin tümü alt problem 1.4.'te elde edilen değerlerin ortalaması alınarak hesaplanmıştır.

Tablo 8'e göre sınıflama kriterleri dışındaki diğer bağımsız değişkenler dikkate alınmaksızın; GA yönteminin yaklaşık 12 maddeyle ve 0,78 sınıflama doğruluğuyla, AOOT yönteminin ise yaklaşık 39 maddeyle ve 0,86 sınıflama doğruluğuyla sınıflama yapabildiği görülmektedir. Bu bulgunun, BBST uygulamalarında OTU bakımından GA yönteminin daha kullanışlı olacağına işaret ettiği söylenebilir. Gündeğer (2017), Nydick ve diğerleri (2012), Thompson (2011) ile Thompson ve Ro'nun (2007) çalışma sonuçları, bu bulguyu kısmen desteklemektedir. Bu çalışmalarda test etkililiği (OTU ve OSD) bakımından GA'nın AOOT'den daha kullanışlı bir sınıflama kriteri olduğu ifade edilmiştir. Ancak yüksek riskli testler düşünüldüğünde artan madde sayısına rağmen daha yüksek sınıflama doğruluğu sağlayan AOOT yönteminin daha avantajlı olacağı söylenebilir.

Tablo 8

AOOT ve GA Sınıflama Kriterlerinin Uygulanan İçerik Oranı, MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri

Bağımsız Değişken	Uygulanan İçerik Oranı				MKSAMY	MKSAMO	OTU	OSD	r	Yanlılık	RMSE	OMH
	İçerik Alanları İstenen Oranları											
	1 %40	2 %30	3 %20	4 %10								
AOOT	37,09	29,48	21,31	12,12	0,179	0,258	39,11	0,86	0,93	-0,005	0,367	0,275
GA	40,74	30,39	19,71	9,16	0,046	0,257	12,49	0,78	0,88	0,011	0,491	0,380

Not: MKSAMY= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Yüzdesi, MKSAMO= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Madde Kullanım Sıklıklarının Ortalaması, OTU= Ortalama Test Uzunluğu, OSD= Ortalama Sınıflama Doğruluğu, r= Gerçek ve Kestirilen Yetenek Düzeyleri Arasındaki Korelasyon.

Tablo 8'de bireylerin gerçek ve kestirilen son yetenek düzeyleri arasındaki korelasyonlar (r) bakımından AOOT sınıflama kriterinin GA'dan daha yüksek ilişkiler sağladığı görülmekle birlikte her iki sınıflama kriterinin de iyi performans gösterdikleri söylenebilir. Benzer şekilde yanlılık, RMSE ve OMH bakımından da AOOT'nin GA'ya kıyasla daha iyi performans gösterdiği görülmektedir. Bu sonuçlar test etkililiğiyle birlikte düşünüldüğünde, OTU bakımından dezavantajlı olsa da daha yüksek sınıflama doğruluğu sağlayan AOOT sınıflama kriterinin uygulamada daha avantajlı olacağı söylenebilir.

Tablo 8'e göre GA sınıflama kriteri kullanıldığında; uygulanan içerik oranlarının istenen içerik oranlarını (sırasıyla %40, %30, %20 ve %10) sağladığı, AOOT sınıflama kriteri kullanıldığında ise uygulanan içerik oranlarının istenen içerik oranlarının kısmen üstünde veya altında kaldığı görülmektedir. Bunun yanı sıra, GA sınıflama kriteri kullanıldığında; maksimum madde kullanım sıklığı oranını ($r^{max}=0,20$) aşan maddelerin yüzdesi (MKSAMY) daha düşükken bu maddelerin madde kullanım sıklığı ortalamasının (MKSAMO) GA ve AOOT yöntemleri için benzer olduğu görülmektedir. Buna göre, içerik dengeleme ve madde kullanım sıklığı kontrolünün yapılacağı durumlarda GA sınıflama kriterinin nispeten daha avantajlı olacağı yorumu yapılabilir.

Alt Problem 3'e İlişkin Bulgular ve Yorumlar

MFB-KY ve MFB-KN madde seçme yöntemlerine göre OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

Araştırmanın üçüncü alt probleminde, bağımsız değişkenlerden MFB-KY ve MFB-KN madde seçme yöntemleri; OTU, OSD, korelasyon (r), yanlılık, RMSE ve OMH değerleri bakımından incelenmiştir. Madde seçme yöntemlerine göre bağımlı değişkenlerdeki değişim; dört içerik alanı için istenen içerik oranları, uygulanan içerik oranları, maksimum madde kullanım sıklığı oranını ($r^{max}=0,20$) aşan maddelerin yüzdesi (MKSAMY) ve bu maddelerin madde kullanım sıklıklarının ortalaması (MKSAMO) ile birlikte Tablo 9'da sunulmuştur. Sunulan değerlerin tümü alt problem 1.4.'ye elde edilen değerlerin ortalaması alınarak hesaplanmıştır.

Tablo 9

MFB-KY ve MFB-KN Madde Seçme Yöntemlerinin Uygulanan İçerik Oranı, MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri

Bağımsız Değişken	Uygulanan İçerik Oranı				MKSAMY	MKSAMO	OTU	OSD	r	Yanlılık	RMSE	OMH
	İçerik Alanları İstenen Oranları											
	1 %40	2 %30	3 %20	4 %10								
MFB-KY	38,94	29,94	20,50	10,62	0,104	0,240	25,39	0,82	0,92	-0,002	0,406	0,310
MFB-KN	38,89	29,93	20,52	10,66	0,121	0,275	26,21	0,82	0,89	0,007	0,452	0,345

Not: MKSAMY= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Yüzdesi, MKSAMO= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Madde Kullanım Sıklıklarının Ortalaması, OTU= Ortalama Test Uzunluğu, OSD= Ortalama Sınıflama Doğruluğu, r= Gerçek ve Kestirilen Yetenek Düzeyleri Arasındaki Korelasyon.

Tablo 9'a göre madde seçme yöntemleri dışındaki diğer bağımsız değişkenler dikkate alınmaksızın; MFB-KY yönteminin yaklaşık 25 maddeyle ve 0,82 sınıflama doğruluğuyla, MFB-KN yönteminin ise yaklaşık 26 maddeyle ve 0,82 sınıflama doğruluğuyla sınıflama yapabildiği görülmektedir. Buna göre, test etkililiği bakımından MFB-KY yönteminin nispeten daha kullanışlı olacağı söylenebilir. Bunun yanı sıra, bireylerin gerçek ve kestirilen son yetenek düzeyleri arasındaki korelasyon (r), yanlılık, RMSE ve OMH değerleri bakımından MFB-KY madde seçme yönteminin MFB-KN'ye kıyasla daha iyi performans gösterdiği görülmektedir. Dolayısıyla, MFB-KY madde seçme yönteminin hatadan daha arınık ve test etkililiği yüksek BBST uygulamaları için avantajlı olacağı söylenebilir.

Tablo 9 incelendiğinde, her iki madde seçme yöntemi için de uygulanan içerik oranlarının istenen içerik oranlarına (sırasıyla %40, %30, %20 ve %10) yakın olduğu görülmektedir. MFB-KY madde seçme yöntemi kullanıldığında; maksimum madde kullanım sıklığı oranını aşan maddelerin yüzdesi (MKSAMY) daha düşük ve bu maddelerin madde kullanım sıklığı ortalaması (MKSAMO) belirlenen $r^{max}=0,20$ değerine daha yakındır. Buna göre BBST uygulamalarında MFB-KY madde seçme yöntemi tercih edildiğinde; içerik dengeleme, madde kullanım sıklığı kontrolü, test etkililiği ve ölçme kesinliği olmak üzere tüm koşullar için avantaj sağlanacağı söylenebilir.

Alt Problem 4'e İlişkin Bulgular ve Yorumlar

KBBT ve MMM içerik dengeleme yöntemlerine göre OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

Araştırmanın dördüncü alt probleminde, bağımsız değişkenlerden KBBT ve MMM içerik dengeleme yöntemleri; OTU, OSD, korelasyon (r), yanlılık, RMSE ve OMH değerleri bakımından incelenmiştir. İçerik dengeleme yöntemlerine göre bağımlı değişkenlerdeki değişim; dört içerik alanı için istenen içerik oranları, uygulanan içerik oranları, maksimum madde kullanım sıklığı oranını ($r^{max}=0,20$) aşan maddelerin yüzdesi (MKSAMY) ve bu maddelerin madde kullanım sıklıklarının ortalaması (MKSAMO) ile birlikte Tablo 10'da sunulmuştur. Sunulan değerlerin tümü alt problem 1.4.'te elde edilen değerlerin ortalaması alınarak hesaplanmıştır.

Tablo 10

KBBT ve MMM İçerik Dengeleme Yöntemlerinin Uygulanan İçerik Oranı, MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri

Bağımsız Değişken	Uygulanan İçerik Oranı				MKSAMY	MKSAMO	OTU	OSD	r	Yanlılık	RMSE	OMH
	İçerik Alanları İstenen Oranları											
	1 %40	2 %30	3 %20	4 %10								
KBBT	37,91	29,88	20,98	11,24	0,111	0,258	25,65	0,82	0,91	0,003	0,429	0,327
MMM	39,92	30,00	20,04	10,04	0,113	0,256	25,95	0,82	0,91	0,003	0,429	0,328

Not: MKSAMY= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Yüzdesi, MKSAMO= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Madde Kullanım Sıklıklarının Ortalaması, OTU= Ortalama Test Uzunluğu, OSD= Ortalama Sınıflama Doğruluğu, r= Gerçek ve Kestirilen Yetenek Düzeyleri Arasındaki Korelasyon.

Tablo 10'a göre içerik dengeleme yöntemleri dışındaki diğer bağımsız değişkenler dikkate alınmaksızın; KBBT yönteminin de MMM yönteminin de yaklaşık 26 maddeyle ve 0,82 sınıflama doğruluğuyla sınıflama yapabildiği görülmektedir. Ayrıca bireylerin gerçek yetenek düzeyleri ile kestirilen yetenek düzeyleri arasındaki korelasyon (r), yanlılık, RMSE ve OMH değerleri bakımından da KBBT ve MMM içerik dengeleme yöntemlerinin benzer sonuçlar verdiği görülmektedir.

Tablo 10 incelendiğinde, MMM içerik dengeleme yöntemi kullanıldığında; uygulanan içerik oranlarının istenen içerik oranlarını (sırasıyla %40, %30, %20 ve %10) sağladığı, KBBT yöntemi kullanıldığında ise uygulanan içerik oranlarının istenen içerik oranlarının kısmen üstünde veya altında kaldığı görülmektedir.

Tablo 10'a göre, her iki içerik dengeleme yöntemi için de maksimum madde kullanım sıklığı oranını ($r^{max}=0,20$) aşan maddelerin yüzdesinin (MKSAMY) ve bu maddelerin madde kullanım sıklığı ortalamasının (MKSAMO) benzer değerlere sahip olduğu görülmektedir. Tablo 10'dan elde edilen tüm bulgular dikkate alındığında, özellikle içerik dengelemenin kritik öneme sahip olduğu durumlarda, MMM içerik dengeleme yönteminin KBBT yöntemine kıyasla daha iyi performans gösterdiği ve uygulamada daha avantajlı olduğu söylenebilir.

Alt Problem 5'e İlişkin Bulgular ve Yorumlar

SH ve MU madde kullanım sıklığı kontrol yöntemlerine göre OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

Araştırmanın beşinci alt probleminde, bağımsız değişkenlerden SH ve MU madde kullanım sıklığı kontrol yöntemleri; OTU, OSD, korelasyon (r), yanlılık, RMSE ve OMH değerleri bakımından incelenmiştir. Madde kullanım sıklığı kontrol yöntemlerine göre bağımlı değişkenlerdeki değişim; dört içerik alanı için istenen içerik oranları, uygulanan içerik oranları, maksimum madde kullanım sıklığı oranını ($r^{max}=0,20$) aşan maddelerin yüzdesi (MKSAMY) ve bu maddelerin madde kullanım sıklıklarının ortalaması (MKSAMO) ile birlikte Tablo 11'de sunulmuştur. Sunulan değerlerin tümü alt problem 1.4.'te elde edilen değerlerin ortalaması alınarak hesaplanmıştır.

Tablo 11

SH ve MU Madde Kullanım Sıklığı Kontrol Yöntemlerinin Uygulanan İçerik Oranı, MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri

Bağımsız Değişken	Uygulanan İçerik Oranı				MKSAMY	MKSAMO	OTU	OSD	r	Yanlılık	RMSE	OMH
	İçerik Alanları İstenen Oranları											
	1 %40	2 %30	3 %20	4 %10								
SH	39,02	29,97	20,46	10,55	0,126	0,313	24,65	0,83	0,91	0,004	0,423	0,322
MU	38,81	29,90	20,56	10,73	0,098	0,202	26,95	0,81	0,91	0,002	0,435	0,333

Not: MKSAMY= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Yüzdesi, MKSAMO= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Madde Kullanım Sıklıklarının Ortalaması, OTU= Ortalama Test Uzunluğu, OSD= Ortalama Sınıflama Doğruluğu, r= Gerçek ve Kestirilen Yetenek Düzeyleri Arasındaki Korelasyon.

Tablo 11'e göre madde kullanım sıklığı kontrol yöntemleri dışındaki diğer bağımsız değişkenler dikkate alınmaksızın; SH yönteminin yaklaşık 25 maddeyle ve 0,83 sınıflama doğruluğuyla, MU yönteminin ise yaklaşık 27 maddeyle ve 0,81 sınıflama doğruluğuyla sınıflama yapabildiği görülmektedir. Bu bulgunun, BBST uygulamalarında test etkililiği bakımından SH yönteminin nispeten daha kullanışlı olacağına işaret ettiği söylenebilir.

Tablo 11'de bireylerin gerçek yetenek düzeyleri ile kestirilen son yetenek düzeyleri arasındaki korelasyon (r), yanlışlık, RMSE ve OMH değerleri bakımından SH ve MU madde kullanım sıklığı kontrol yöntemlerinin benzer sonuçlar verdiği görülmektedir. Bu sonuçlar test etkililiğiyle birlikte düşünüldüğünde, SH madde kullanım sıklığı kontrol yönteminin uygulamada daha avantajlı olacağı söylenebilir.

Tablo 11'e göre her iki madde kullanım sıklığı kontrol yöntemi için de uygulanan içerik oranları ile istenen içerik oranlarının (sırasıyla %40, %30, %20 ve %10) benzer olduğu, uygulanan içerik oranlarının istenen içerik oranlarına oldukça yakın olduğu görülmektedir. MU madde kullanım sıklığı kontrol yöntemi kullanıldığında; maksimum madde kullanım sıklığı oranını aşan maddelerin yüzdesinin (MKSAMY) SH yöntemine kıyasla daha düşük olduğu ve bu maddelerin madde kullanım sıklığı ortalamasının (MKSAMO) ise belirlenen $r^{max}=0,20$ değerini sağladığı görülmektedir. SH madde kullanım sıklığı kontrol yöntemi kullanıldığında ise hem hesaplanan maksimum madde kullanım sıklığı oranını aşan maddelerin yüzdesinin (MKSAMY) daha yüksek olduğu hem de bu maddelerin madde kullanım sıklığı ortalamasının (MKSAMO) belirlenen $r^{max}=0,20$ değerinin oldukça üstünde olduğu görülmektedir.

Tablo 11'den elde edilen tüm bulgular dikkate alındığında, BBST uygulamasında madde kullanım sıklığı kontrolü daha kritik bir öneme sahipse MU yönteminin tercih edilmesinin daha avantajlı olduğu, buna karşılık test etkililiği daha kritik bir öneme sahipse de SH yönteminin tercih edilmesinin daha avantajlı olduğu söylenebilir. Araştırmanın bu bulgusu, madde kullanım sıklığı kontrolünde MU yönteminin SH karşısında sağladığı avantaj bakımından Huebner'in (2012) araştırma sonuçlarıyla da desteklenmektedir. Huebner (2012) çalışmasında, madde kullanım sıklığı kontrolü bakımından MU yönteminin SH'den daha başarılı çalıştığı sonucuna ulaşmıştır.

Alt Problem 6'ya İlişkin Bulgular ve Yorumlar

Sınıflama kategori sayısına göre OTU, OSD, gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri nasıl değişmektedir?

Araştırmanın altıncı alt probleminde, bağımsız değişkenlerden sınıflama kategori sayısı (iki, üç veya dört); OTU, OSD, korelasyon (r), yanlılık, RMSE ve OMH değerleri bakımından incelenmiştir. Sınıflama kategori sayılarına göre bağımlı değişkenlerdeki değişim; dört içerik alanı için istenen içerik oranları, uygulanan içerik oranları, maksimum madde kullanım sıklığı oranını ($r^{max}=0,20$) aşan maddelerin yüzdesi (MKSAMY) ve bu maddelerin madde kullanım sıklıklarının ortalaması (MKSAMO) ile birlikte Tablo 12'de sunulmuştur. Sunulan değerlerin tümü alt problem 1.4.'te elde edilen değerlerin ortalaması alınarak hesaplanmıştır.

Tablo 12'ye göre sınıflama kategori sayıları dışındaki diğer bağımsız değişkenler dikkate alınmaksızın; BBST uygulamasında bireylerin sınıflanıp testin sonlanması için iki kategoride yaklaşık 20 maddeyle ve 0,91 sınıflama doğruluğuyla, üç kategoride yaklaşık 27 maddeyle ve 0,82 sınıflama doğruluğuyla, dört kategoride yaklaşık 31 maddeyle ve 0,73 sınıflama doğruluğuyla, sınıflama yapılabildiği görülmektedir. Bu bulgu, sınıflama kategori sayısındaki artışın test etkililiğini düşürdüğünü, başka bir ifadeyle OTU'yu artırırken OSD'yi azalttığını göstermektedir. Araştırmanın bu bulgusu Eggen (1999) ile Nydick ve diğerlerinin (2012) araştırma sonuçlarıyla da örtüşmektedir. Eggen (1999) ile Nydick ve diğerleri (2012) tarafından yapılan çalışmalarda, artan kategori sayısının OTU'yu artırırken OSD'yi azalttığı sonucuna ulaşılmıştır.

Tablo 12'de sınıflama kategori sayısındaki artışa rağmen korelasyon (r) ve yanlılık değerleri benzer sonuçlar verirken RMSE ve OMH değerleri genel olarak nispeten düşmüştür. Bu sonuçlar test etkililiğiyle birlikte düşünüldüğünde, BBST uygulamalarında mümkün olduğunca az kategoride yapılacak sınıflamaların daha avantajlı olacağı söylenebilir.

Tablo 12

İki, Üç ve Dört Kategoride Yapılan Sınıflamaların Uygulanan İçerik Oranı, MKSAMY, MKSAMO, OTU, OSD, r, Yanlılık, RMSE ve OMH Değerleri

Bağımsız Değişken	Uygulanan İçerik Oranı				MKSAMY	MKSAMO	OTU	OSD	r	Yanlılık	RMSE	OMH
	İçerik Alanları İstenen Oranları											
	1 %40	2 %30	3 %20	4 %10								
İki Kategori	39,50	30,11	20,27	10,13	0,081	0,255	19,93	0,91	0,90	0,004	0,450	0,345
Üç Kategori	38,84	29,91	20,54	10,71	0,118	0,257	26,53	0,82	0,91	0,002	0,426	0,324
Dört Kategori	38,41	29,79	20,72	11,08	0,138	0,260	30,94	0,73	0,92	0,002	0,411	0,313

Not: MKSAMY= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Yüzdesi, MKSAMO= Maksimum Madde Kullanım Sıklığı Oranını ($r^{max}=0,20$) Aşan Maddelerin Madde Kullanım Sıklıklarının Ortalaması, OTU= Ortalama Test Uzunluğu, OSD= Ortalama Sınıflama Doğruluğu, r= Gerçek ve Kestirilen Yetenek Düzeyleri Arasındaki Korelasyon.

Tablo 12'ye göre sınıflama kategori sayısındaki artışın; uygulanan içerik oranlarının istenen içerik oranlarından (sırasıyla %40, %30, %20 ve %10) uzaklaşmasına neden olduğu, istenen içerik oranlarını sağlamada en iyi performansın iki kategoride yapılan sınıflamalarda görüldüğü, en kötü performansın ise dört kategoride yapılan sınıflama görüldüğü söylenebilir. Bununla birlikte, sınıflama kategori sayısındaki artışın; maksimum madde kullanım sıklığı oranını ($r^{max}=0,20$) aşan maddelerin yüzdesini (MKSAMY) artırırken bu maddelerin madde kullanım sıklığı ortalamasını (MKSAMO) ise yükselttiği, başka bir ifadeyle madde kullanım sıklığı kontrolü üzerinde olumsuz bir etki yaptığı görülmektedir.

BBST uygulamalarında amaç, bireylerin olabildiğince az sayıda maddeyle ve yüksek doğruluk oranıyla sınıflanmasıdır. Bu açıdan bakıldığında ve Tablo 12'den elde edilen tüm bulgular birlikte düşünüldüğünde mümkün olduğunca az kategoride yapılacak sınıflamaların daha avantajlı olacağı söylenebilir.

Bölüm 5

Sonuç, Tartışma ve Öneriler

Bu bölümde araştırma bulgularına dayalı olarak çıkarılan sonuçlara ve bu sonuçlar doğrultusunda geliştirilen önerilere yer verilmiştir.

Sonuçlar

Bu araştırmada, BBST uygulamalarında kullanılan farklı sınıflama kriterleri ve madde seçme yöntemlerinin performansı, AOK yetenek kestirim yönteminin kullanıldığı durumda, farklı içerik dengeleme yöntemleri, madde kullanım sıklığı kontrol yöntemleri ve sınıflama kategori sayılarına göre OTU, OSD, bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri bakımından MC simülasyonu aracılığıyla incelenmiştir. Bu doğrultuda, oluşturulan 48 koşul için karşılaştırmalar yapılmış ve araştırmanın altı alt problemine yanıt aranmıştır. Araştırmadan elde edilen sonuçlar aşağıda sunulmuştur.

- Bu araştırmanın temel odak noktası sınıflama doğruluğu olduğundan farklı koşullarda OSD için hesaplanan değerler, sunulan bulguların yorumlanmasında büyük öneme sahiptir. Tüm araştırma koşullarında; madde seçme yöntemi, içerik dengeleme yöntemi, madde kullanım sıklığı kontrol yöntemi ve sınıflama kategori sayısı fark etmeksizin OSD bakımından yüksek değerler elde edilmesinin yanı sıra AOOT sınıflama kriterinin GA'ya kıyasla daha iyi performans gösterdiği ve bireyleri doğru kategorilere sınıflama oranının daha yüksek olduğu sonucuna ulaşılmıştır.
- GA sınıflama kriterinin, tüm araştırma koşullarında; madde seçme yöntemi, içerik dengeleme yöntemi, madde kullanım sıklığı kontrol yöntemi ve sınıflama kategori sayısı fark etmeksizin OTU bakımından AOOT'ye kıyasla daha iyi performans gösterdiği ve bireyleri sınıflamada daha az sayıda madde gerektirdiği belirlenmiştir.
- Bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri bakımından ise tüm araştırma koşullarında, AOOT sınıflama kriterinin görece olarak GA'dan daha iyi

performans gösterdiği, MFB-KY madde seçme yönteminin ise MFB-KN'den daha iyi performans gösterdiği tespit edilmiştir.

- Araştırma sonucunda, ele alınan tüm koşullarda, bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon değerlerinin oldukça yüksek olduğu gözlenmiştir. Buradan yola çıkarak, sınıflama kriteri, madde seçme yöntemi, içerik dengeleme yöntemi ve madde kullanım sıklığı kontrol yöntemi fark etmeksizin yetenek kestiriminde kullanılan AOK yönteminin oldukça başarılı bir performans gösterdiği söylenebilir. Buna ek olarak AOK yetenek kestirim yönteminin AOOT sınıflama kriteri ve MFB-KY madde seçme yöntemiyle birlikte kullanıldığı koşullarda daha başarılı sonuçlar verdiği belirlenmiştir.
- Araştırmada ele alınan sınıflama kriterleri, içerik dengeleme yöntemleri, madde kullanım sıklığı kontrol yöntemleri ve sınıflama kategori sayıları fark etmeksizin, test etkililiği ve ölçme kesinliği bakımından MFB-KY madde seçme yönteminin MFB-KN madde seçme yöntemine kıyasla nispeten daha başarılı sonuçlar verdiği belirlenmiştir. Başka bir ifadeyle MFB-KY madde seçme yönteminin, daha düşük OTU, benzer OSD oranları, daha yüksek korelasyon ve daha düşük yanlılık, RMSE, OMH değerleri verdiği görülmüştür.
- Araştırma koşullarına dahil edilen madde kullanım sıklığı kontrol yöntemlerinin içerik dengeleme yöntemlerine kıyasla OTU'yu daha fazla artırdığı, OSD'de ise her iki kısıtlamanın da önemli bir değişime sebep olmadığı tespit edilmiştir.
- İçerik dengeleme yöntemlerinden MMM'in, istenen içerik oranını sağlamada KBBT'ye kıyasla daha iyi performans gösterdiği, buna karşılık test etkililiği söz konusu olduğunda, GA sınıflama kriterinin kullanıldığı koşullarda çok küçük değişiklikler olsa da AOOT'nin kullanıldığı koşullarda, özellikle OTU bakımından KBBT'nin daha iyi performans gösterdiği sonucuna ulaşılmıştır.
- Madde kullanım sıklığını, MU yönteminin SH yöntemine kıyasla daha iyi kontrol ettiği, buna karşılık test etkililiği bakımından, özellikle AOOT sınıflama kriterinin kullanıldığı koşullarda, SH'nin performansının daha

iyi olduđu görülmüştür. Öyle ki, SH'nin kullanıldığı koşullarda daha düşük OTU ve daha yüksek OSD'nin gözlemlendiği belirlenmiştir.

- Araştırma koşullarına dahil edilen içerik dengeleme ve madde kullanım sıklığı kontrol yöntemleri, bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyonu deđiştirmezken, yanlılıkta düşüşe, RMSE ve OMH deđerlerinde ise bir miktar artışa sebep olmuştur. Ayrıca madde kullanım sıklığı kontrolünün içerik dengelemeye kıyasla; yanlılık, RMSE ve OMH deđerlerinde daha fazla deđişime sebep olduđu belirlenmiştir. Buna ek olarak en düşük yanlılık, en yüksek RMSE ve OMH deđerlerinin, içerik dengeleme ve madde kullanım kontrolünün araştırma koşullarına birlikte dahil edildiđi durumlarda gözlemlendiği belirlenmiştir.
- Tüm araştırma koşullarında; sınıflama kriteri, madde seçme yöntemi, içerik dengeleme yöntemi ve madde kullanım sıklığı kontrol yöntemi fark etmeksizin artan kategori sayısının test etkililiđini düşürdüđü, başka bir ifadeyle OTU'yu artırırken OSD'yi düşürdüđü görülmüştür. Bunun yanı sıra kategori sayısındaki artışa rağmen bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon ve yanlılık deđerleri bakımından benzer sonuçlar elde edildiđi, RMSE ve OMH deđerlerinin ise genel olarak nispeten düştüđü belirlenmiştir.

Öneriler

Bu kısımda, araştırma sonuçlarına dayalı olarak geliştirilen araştırmaya yönelik önerilere ve uygulamaya yönelik önerilere yer verilmiştir.

Araştırmaya yönelik öneriler. Araştırmaya yönelik öneriler aşağıda sunulmuştur.

- Bu çalışmada, madde havuzu tek boyutlu olacak şekilde simüle edilmiştir. Benzer BBST çalışmaları çok boyutlu madde havuzları için de gerçekleştirilebilir.
- Bu çalışmada, MTK temel alınmış ve iki kategorili puanlanan maddelerden oluşturulan madde havuzu 3 PLM ile kalibre edilmiştir. İlerideki çalışmalarda, çok kategorili puanlanan veya karma

maddelerden oluşan madde havuzlarında BBST uygulamaları yapılabilir.

- Tek boyutluluğun ihlal edildiği veya model uyumunun sağlanmadığı durumlar için sınıflama kriterlerinin performansının incelendiği yeni çalışmalar yapılabilir.
- Sahip olduğu madde sayısı bakımından farklı büyüklükteki, sivri, basık vb. özellikler taşıyan madde havuzlarıyla da benzer çalışmalar gerçekleştirilebilir.
- Bu çalışmada sınıflama kriterlerinden AOOT ve GA ele alınmıştır. GOO ve BKK gibi diğer sınıflama kriterleriyle benzer bir çalışma yürütülebilir.
- Bu çalışmanın bir sınırlılığı, yetenek kestiriminde sadece AOK yönteminin kullanılmış olmasıdır. İleride yapılacak çalışmalarda MOK, BSD, MSD gibi farklı yetenek kestirim yöntemleri diğer araştırma koşullarıyla çaprazlanarak incelenebilir.
- Bu çalışmada madde seçme yöntemlerinden sadece MFB-KY ve MFB-KN kullanılmıştır. Yapılacak benzer çalışmalarda diğer madde seçme yöntemleri de ele alınabilir.
- Bu çalışmada içerik dengeleme yöntemlerinden KBBT ve MMM kullanılmıştır. İleride yapılacak benzer çalışmalarda diğer içerik dengeleme yöntemlerinin BBST uygulamalarındaki performansı incelenebilir.
- Bu çalışmada madde kullanım sıklığı kontrolünde SH ve MU yöntemlerinden yararlanılmıştır. Yapılacak benzer çalışmalarda diğer madde kullanım sıklığı kontrol yöntemlerinin BBST uygulamalarındaki performansı karşılaştırılabilir.
- Bu çalışma sadece MC simülasyonu ile sınırlıdır. Benzer bir çalışmada, temel BBST bileşenlerinin performansları, gerçek veri üzerinden de karşılaştırılabilir.

Uygulamaya yönelik öneriler. Uygulamaya yönelik öneriler aşağıda sunulmuştur.

- Araştırmada ele alınan AOOT ve GA sınıflama kriterleri, araştırmanın odak noktası olan yüksek sınıflama doğruluğu bakımından değerlendirildiğinde, daha yüksek OSD'lerin hesaplandığı AOOT yönteminin tercih edilmesi önerilebilir. Buna karşılık OTU bakımından yapılan değerlendirme sonucu bireylerin sınıflanıp testin sonlanabilmesi için daha az sayıda madde gerektiren GA sınıflama kriterinin tercih edilmesi söz konusu olabilir. Ancak yüksek riskli testler (örneğin, Tıp alanında uygulanan ve doğrudan insan hayatıyla ilgili olan testler) dikkate alındığında artan madde sayısına rağmen daha yüksek sınıflama doğruluğu sağlayan yöntemin tercih edilmesinin çok daha önemli olduğu unutulmamalıdır.
- BBST uygulamalarında, test etkililiği adı altında OTU ve OSD genellikle birlikte değerlendirilmektedir. Dolayısıyla test etkililiği bakımından en başarılı sınıflama kriterinin tercih edilmesi gerektiğinde, her iki sınıflama kriterinin de (AOOT ve GA) iyi düzeyde sınıflama doğruluğu sağladığı düşünülecek olursa, GA sınıflama kriterinin kullanılması önerilebilir.
- Bunun yanı sıra yapılan BBST uygulamasının odak noktası; bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık, RMSE ve OMH değerleri ise AOOT sınıflama kriterinin tercih edilmesinin daha avantajlı olacağı söylenebilir.
- Araştırma sonucunda, sınıflama kriteri fark etmeksizin MFB-KY madde seçme yönteminin MFB-KN'den daha kullanışlı olduğu görülmüştür. Dolayısıyla MFB-KY madde seçme yönteminin tercih edilmesi önerilebilir.
- Yapılan BBST uygulamasında, yüksek riskli testlerde olduğu gibi içerik dengeleme kritik öneme sahipse MMM yönteminin, test etkililiği daha kritik öneme sahipse de KBBT yönteminin kullanımı önerilebilir.
- BBST uygulamasında madde kullanım sıklığının kontrolü başka bir ifadeyle sınavın/madde havuzunun güvenliği kritik öneme sahipse MU yönteminin, test etkililiği daha kritik öneme sahipse de SH yönteminin kullanımı önerilebilir.

- BBST uygulamasında, içerik dengeleme ve madde kullanım sıklığı kontrol yöntemleri araştırma koşullarına dahil edildiğinde GA sınıflama kriteri ile elde edilen sonuçların çok az etkilendiği görülmüştür. Sağladığı bu avantajdan dolayı GA'nın tercih edilmesi önerilebilir.
- BBST uygulamasında artan kategori sayısının OTU'yu artırırken OSD'yi azalttığı, başka bir ifadeyle test etkililiğini düşürdüğü görülmüştür. Dolayısıyla test etkililiği bakımından düşünüldüğünde, sınıflama kategori sayısının mümkün olduğunca az belirlenmesi önerilebilir. Bunun yanı sıra, artan kategori sayısına karşılık; bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon ve yanlılık değerlerinin benzer olduğu, RMSE ve OMH değerlerinin ise nispeten düştüğü sonucu da dikkate alınarak sınıflama kategori sayısının daha ideal şekilde belirlenebileceği söylenebilir.
- Araştırmadan elde edilen sonuçlara göre, OTU bakımından GA sınıflama kriterinin bütün koşullarda benzer ve AOOT'den daha iyi performans gösterdiği görülmüştür. Dolayısıyla BBST uygulamasının odak noktası OTU iken içerik dengeleme ve madde kullanım sıklığı kontrolü de yapılacaksa GA sınıflama kriteri, MFB-KY madde seçme yöntemi, MMM içerik dengeleme yöntemi ve MU madde kullanım sıklığı kontrol yönteminin birlikte kullanılması önerilebilir.
- Tüm araştırma sonuçları genel bir çerçeveden değerlendirildiğinde, BBST uygulamasının odak noktası OSD iken içerik dengeleme ve madde kullanım sıklığı kontrolü de kritik öneme sahipse; korelasyon, yanlılık, RMSE ve OMH değerleri bakımından da daha iyi performans gösteren AOOT sınıflama kriterinin, MFB-KY madde seçme yöntemi, MMM içerik dengeleme yöntemi ve MU madde kullanım sıklığı kontrol yöntemiyle birlikte kullanılması önerilebilir. Bunun yanı sıra BBST uygulamasının odak noktası test etkililiği (OTU ve OSD) ise AOOT, MFB-KY, KBBT ve SH yöntemlerinin birlikte kullanılmasının daha avantajlı olacağı söylenebilir.

Kaynaklar

- Baker, F. B. (2001). *The basics of item response theory*. USA: ERIC Clearinghouse on Assessment and Evaluation.
- Barrada, J.A., Abad F.J., & Veldkamp, B.P. (2009). Comparison of methods for controlling maximum exposure rates in computerized adaptive testing. *Psicothema*, 21(2), 313-320. <https://www.researchgate.net/publication/24377944> adresinden erişilmiştir.
- Barret, A. F. (2015). *Facilitating variable-length computerized classification testing via automatic racing calibration heuristics*. (Unpublished Doctoral Dissertation). The University of Indiana, Indiana.
- Bartroff, J., Finkelman, M., & Lai, T. L. (2008). Modern sequential analysis and its applications to computerized adaptive testing. *Psychometrika*, 73(3), 473-486. doi: 10.1007/S11336-007-9053-9
- 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: 10.5926/jjep1953.41.1_22
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2008). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.
- Chang, Y.-C.I. (2005). Application of Sequential Interval Estimation to Adaptive Mastery Testing. *Psychometrika*, 70(4), 685-713. doi: 10.1007/S11336-005-1140-9
- Chang, H. H., & Ying, Z. (1999). a-Stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, 23(3), 211-222. doi: 10.1177/01466219922031338
- Chen, S.-Y., & Ankenmann, R. D. (2004). Effects of practical constraints on item selection rules at the early stages of computerized adaptive testing. *Journal of Educational Measurement*, 41(2), 149-174. doi: 10.1111/j.1745-3984.2004.tb01112.x

- Chen, S.Y. Lei, P.W., & Liao, W.H. (2008). Controlling item exposure and test overlap on the fly in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, 61(2), 75-91. doi: 10.1348/000711007X227067
- Cheng, Y., & Chang, H. (2009). The maximum priority index method for severely constrained item selection in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, 62(2), 369-383. doi: 10.1348/000711008X304376
- Cheng, Y., Chang, H.-H., Douglas, J., & Guo, F. (2009). Constraint-weighted a-stratification for computerized adaptive testing with nonstatistical constraints: Balancing measurement efficiency and exposure control. *Educational and Psychological Measurement*, 69(1), 35-49. doi: 10.1177/0013164408322030
- Cortina, J. M. (1993). What is coefficient Alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104. doi: 10.1037/0021-9010.78.1.98
- Davey, T., & Parshall, C. G. (1995, April). *New algorithms for the item selection and exposure control with computerized adaptive testing*. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.
- Demars, C. (2010). *Item response theory*. New York: Oxford University Press, Inc.
- Dooley, K. (2002). Simulation research methods. In J. Baum (Ed.). *Companion to organizations*, 829-848. London: Blackwell.
- Eggen, T. J. H. M. (1999). Item selection in adaptive testing with the sequential probability ratio test. *Applied Psychological Measurement*, 23, 249-261. doi: 10.1177/01466219922031365
- Eggen, T. J. H. M. (2011). Computerized classification testing with the Rasch model. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 17(5), 361-371. doi: 10.1080/13803611.2011.630528
- 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. doi:
10.1177/00131640021970862

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates Publishers.

Falissard, B. (2012). *psy: Various Procedures Used In Psychometry*. <https://cran.r-project.org/web/packages/catIrt/catIrt.pdf> adresinden erişilmiştir.

Fan, Z., Wang, C., Chang, H., & Douglas, J. (2012). Utilizing Response Time Distributions for Item Selection in CAT. *Journal of Educational and Behavioral Statistics*, 37(5), 655-670. doi: 10.3102/1076998611422912

Finkelman, M. (2008). On using stochastic curtailment to shorten the SPRT in sequential mastery testing. *Journal of Educational and Behavioral Statistics*, 33(4), 442-463. doi: 10.3102/1076998607308573

Frick, T. W. (1992). Computerized adaptive mastery tests as expert systems. *Journal of Educational Computing Research*, 8(2), 187-213. doi: 10.2190/J87V-6VWP-52G7-L4XX

Georgiadou, E., Triantafillou, E., & Economides, A. (2007). A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *Journal of Technology, Learning, and Assessment*, 5(8). <https://ejournals.bc.edu/index.php/jtla/article/download/1647/1482/> adresinden erişilmiştir.

Glas, C.A.W., & Vos, H.J. (2006). *Testlet-Based Adaptive Mastery Testing* (Research Report 99-11). Newtown, PA: Law School Admission Council.

Gündeğer, C. (2017). *Bireyselleştirilmiş bilgisayarlı sınıflama testi kriterlerinin sınıflama doğruluğu ve test uzunluğu açısından karşılaştırılması*. (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Gündeğer, C. (2018, Mayıs). *Bireyselleştirilmiş bilgisayarlı sınıflama testlerinde yetenek kestirim yöntemlerinin çok kategorili madde havuzları altında incelenmesi*. 4. Uluslararası Eğitim Bilimleri Sempozyumunda sunulan bildiri, 152-163, Alanya, Türkiye. <https://www.researchgate.net/publication/327867960> adresinden erişilmiştir.

- Gündeğer, C. ve Doğan, N. (2018). Bireyselleştirilmiş bilgisayarlı sınıflama testlerinde madde havuzu özelliklerinin test uzunluğu ve sınıflama doğruluğu üzerindeki etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 33(4), 888-896. doi: 10.16986/HUJE.2016024284
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer Nijhoff Publishing.
- He, W., Diao, Q., & Hauser, C. (2014). A comparison of four item-selection methods for severely constrained CATs. *Educational and Psychological Measurement*, 74(4), 677-696. doi: 10.1177/0013164413517503
- 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> adresinden erişilmiştir.
- Huebner, A., & Li, Z. (2012). A stochastic method for balancing item exposure rates in computerized classification tests. *Applied Psychological Measurement*, 36(3), 181-188. doi: 10.1177/0146621612439932
- 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> adresinden erişilmiştir.
- Kalohn, J. C., & Spray, J. A. (1999). The effect of model misspecification on classification decisions made using a computerized test. *Journal of Educational Measurement*, 36(1), 47-59. doi: 10.1111/j.1745-3984.1999.tb00545.x
- 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

- (Ed.), *New horizons in testing: Latent trait theory and computerized adaptive testing*, 237-254. New York: Academic Press.
- Kingsbury, G.G., & Zara, A.R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, 2(4), 359-375. doi: 10.1207/s15324818ame0204_6
- Lau, C. A. (1996). *Robustness of a unidimensional computerized testing mastery procedure with multidimensional testing data*. (Unpublished Doctoral Dissertation). University of Iowa.
- Lau, C. A., & Wang, T. (1998, Nisan). *Comparing and combining dichotomous and polytomous items with SPRT procedure in computerized classification testing*. American Educational Research Association (AERA) toplantısında sunulan bildiri, San Diego, CA, <http://iacat.org/sites/default/files/biblio/la98-01.pdf> adresinden erişilmiştir.
- 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.
- Leung, C.-K., Chang, H.-H., & Hau, K.-T. (2002). Item selection in computerized adaptive testing: Improving the a-stratified design with the Sympon–Hetter algorithm. *Applied Psychological Measurement*, 26(4), 376-392. doi: 10.1177/014662102237795
- Leung, C.-K., Chang, H.-H., & Hau, K.-T. (2003). Computerized adaptive testing: A comparison of three content balancing methods. *Journal of Technology, Learning, and Assessment*, 2(5). <https://ejournals.bc.edu/index.php/jtla/article/view/1665/1507> adresinden erişilmiştir.
- Lewis, C., & Sheenan, K. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement*, 14(4), 367-386. doi: 10.1177/014662169001400404

- Lin, C. (2011). Item selection criteria with practical constraints for computerized classification testing. *Applied Psychological Measurement* 71(1), 20-36. doi: 10.1177/0013164410387336
- 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> adresinden erişilmiştir.
- Linn, R. L., Rock, D. A., & Cleary, T. A. (1972). Sequential testing for dichotomous decisions. *Educational & Psychological Measurement*, 32, 85-95. doi: 10.1177/001316447203200108
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates Publishers.
- Lord, F. M. (1983). Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika*, 48, 233-245. doi: 10.1007/BF02294018
- Lord, F. M. & Stocking, M.L. (1988). *Item response theory*. In J.P. Keeves (ed.), *Educational Research, methodology, and measurement: an international handbook* (pp. 269-272). New York: Pergamon press.
- Miller, I., & Miller, M. (2004). *John E. Freund's Mathematical Statistics with Applications*. (7th Edition). New Jersey: Prentice Hall.
- Nydick, S. W. (2013). Multidimensional mastery testing with CAT. Unpublished Doctoral Dissertation. University of Minnesota, USA.
- Nydick, S. W. (2014). *catirt: An R Package for Simulating IRT-Based Computerized Adaptive Tests*. <https://cran.r-project.org/web/packages/catlrt/catlrt.pdf> adresinden erişilmiştir.
- 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> adresinden erişilmiştir.

- 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. doi: 10.2307/2285821
- Parshall, C.G., Spray, J.A., Kalohn, J.C., & Davey, T. (2006). *Practical considerations in computer-based testing*. New York: Springer.
- 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.
- 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.
- Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(4), 311-327. doi: 10.1111/j.1745-3984.1998.tb00541.x
- Robin, F., van der Linden, W. J., Eignor, D. R., Steffen, M., & Stocking, M. L. (2005). *A comparison of two procedures for constrained adaptive test construction* (Research Report 04-39). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/j.2333-8504.2004.tb01966.x> adresinden erişilmiştir.
- Rudner, L. M. (1998). *An online, interactive, computer adaptive testing tutorial*. <http://EdRes.org/scripts/cat> adresinden erişilmiştir.
- 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. *Psychometric Monograph No. 17*. Richmond, VA: Psychometric Society. <http://www.psychometrika.org/journal/online/MN17.pdf> adresinden erişilmiştir.

- Sheehan, K. & Lewis, C. (1992). Computerized mastery testing with nonequivalent testlets. *Applied Psychological Measurement*, 16(1), 65-76. doi: 10.1177/014662169201600108
- Shin, C., Chien, Y., & Way, D. (2012, Nisan). *A comparison of two content balancing methods for fixed and variable length computerized adaptive test*. The National Council on Measurement in Education (NCME) toplantısında sunulan bildiri, Vancouver, Canada. <https://pdfs.semanticscholar.org/5e81/50b70103fa9653d7eb1b76f89b69d1adeeee0.pdf> adresinden erişilmiştir.
- Sie, H., Finkelman, M. D., Riley, B., & Smits, N. (2015). Utilizing response times in computerized classification testing. *Applied Psychological Measurement*, 39(5), 389-405. doi: 10.1177/0146621615569504
- 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> adresinden erişilmiştir.
- Spray, J. A., & Reckase, M. D. (1994, Nisan). *The Selection of Test Items for Decision Making with a Computer Adaptive Test*. The National Council on Measurement in Education (NCME) toplantısında sunulan bildiri, NewOrleans, LA. <https://eric.ed.gov/?id=ED372078> adresinden erişilmiştir.
- 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. doi: 10.3102/10769986021004405
- Stocking, M. L., & Lewis, C. (1995). *A new method of controlling item exposure in computerized adaptive testing* (Research Report 95-25). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/j.2333-8504.1995.tb01660.x> adresinden erişilmiştir.
- Stocking, M. L., & Swanson, L. (1993). A method for severely constrained item selection in adaptive testing. *Applied Psychological Measurement*, 17, 277-292. doi: 10.1177/014662169301700308

- Sympson, J.B., & Hetter, R.D. (1985, Ekim). *Controlling item exposure rates in computerized adaptive testing*. 27th Military Testing Association toplantısında sunulan bildiri, 937-977. San Diego, CA: Navy Personnel Research and Development Center. <http://www.iacat.org/content/controlling-item-exposure-rates-computerized-adaptive-testing> adresinden erişilmiştir.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics*. USA: Pearson Education Inc.
- Thompson, N. A. (2007a). *A comparison of two methods of polytomous computerized classification testing for multiple cutscores*. (Unpublished Doctoral Dissertation). University of Minnesota.
- Thompson, N. A. (2007b). 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> adresinden erişilmiştir.
- Thompson, N. A. (2009). Item selection in computerized classification testing. *Educational and Psychological Measurement*, 69(5), 778-793. doi: 10.1177/0013164408324460
- Thompson, N. A. (2011). Termination criteria for computerized classification testing. *Practical Assessment, Research & Evaluation*, 16(4), 1-7. <https://pareonline.net/getvn.asp?v=16&n=4> adresinden erişilmiştir.
- 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/cat07nthompson.pdf> adresinden erişilmiştir.
- Van der Linden, W. J. (2005). A comparison of item-selection methods for adaptive tests with content constraints. *Journal of Educational Measurement*, 42(3), 283-302. doi: 10.1111/j.1745-3984.2005.00015.x
- Van der Linden, W.J., & Veldkamp, B.P. (2004). Constraining Item Exposure in Computerized Adaptive Testing With Shadow Tests. *Journal of*

- Educational and Behavioral Statistics*, 29(3), 273-291. <https://www.utwente.nl/nl/bms/omd/Medewerkers/artikelen/vdLinden/JEBS%202004%2C%20273-291-1.pdf> adresinden erişilmiştir.
- van Groen, M. M., Eggen, T. J. H. M. & Veldkamp, B. P. (2014). Item selection methods based on multiple objective approaches for classifying respondents into multiple levels. *Applied Psychological Measurement*, 38(3), 187-200. doi: 10.1177/0146621613509723
- Van Groen, M. M., Eggen, T. J. H. M., & Veldkamp, B. P. (2016). Multidimensional computerized adaptive testing for classifying examinees with within-dimensionality. *Applied Psychological Measurement*, 40(6), 387-404. doi: 10.1177/0146621616648931
- Vos, H.J. (2000). A Bayesian Procedure in the Context of Sequential Mastery Testing. *Psicológica*, 21, 191-211. https://www.researchgate.net/publication/26421570_A_bayesian_procedure_in_the_context_of_sequential_mastery_testing adresinden erişilmiştir.
- Wald, A. (1947). *Sequential analysis*. Oxford, England: John Wiley.
- 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. doi: 10.1177/01466210122032163
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427-450. doi: 10.1007/BF02294627
- 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.x> adresinden erişilmiştir.
- Weissman, A. (2006). A Feedback Control Strategy for Enhancing Item Selection Efficiency in Computerized Adaptive Testing. *Applied Psychological Measurement*, 30, 84-99. doi: 10.1177/0146621605282774
- Wouda, J. T., & Eggen, T. J. H. M. (2009). Computerized classification testing in more than two categories by using stochastic curtailment. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized*

Adaptive Testing. <http://iacat.org/sites/default/files/biblio/cat09wouda.pdf>
adresinden erişilmiştir.

Xiao, B. (1999). Strategies for computerized adaptive grading testing. *Applied Psychological Measurement*, 23, 136-146. doi:
10.1177/01466219922031266

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. doi: 10.1177/0013164405284031

**EK-A: Türetilen Madde Parametrelerinden Rastgele Seçilmiş Bir Tanesine Ait
Betimsel İstatistikler**

	a	b	c
N	500	500	500
Ortalama	1,24	-0,37	0,20
Ortanca	1,25	-0,40	0,20
Mod	0,50 ^a	-2,98 ^a	0,04 ^a
Standart Sapma	0,43	1,27	0,05
Çarpıklık	0,02	0,18	-0,09
Çarpıklığın Standart Hatası	0,11	0,11	0,11
Basıklık	-1,18	-0,48	-0,22
Basıklığın Standart Hatası	0,22	0,22	0,22
Ranj	1,49	5,92	0,29
Minimum	0,50	-2,98	0,04
Maximum	1,99	2,94	0,33

^a. Çoklu mod mevcuttur. En küçük mod değeri sunulmuştur.

**EK-B: Türetilen Yetenek Parametrelerinden Rastgele Seçilmiş Bir Tanesine
Ait Betimsel İstatistikler**

	θ
N	5000
Ortalama	0,01
Ortanca	0,00
Mod	-2,99 ^a
Standart Sapma	0,97
Çarpıklık	0,04
Çarpıklığın Standart Hatası	0,04
Basıklık	-0,13
Basıklığın Standart Hatası	0,07
Ranj	5,95
Minimum	-2,99
Maximum	2,95

^a. Çoklu mod mevcuttur. En küçük mod değeri sunulmuştur.

EK-C: Etik Komisyonu Onay Bildirimi



T.C.
HACETTEPE ÜNİVERSİTESİ
Rektörlük

Tarih: 16.01.2019 18:18
Sayı: 35853172-300-E.00000412502



E.00000412502

Sayı : 35853172-300
Konu : Seda DEMİR Hk.

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 21.12.2018 tarihli ve 51944218-300/00000380813 sayılı yazı.

Enstitünüz Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı tezli doktora programı öğrencilerinden **Seda DEMİR'in Doç. Dr. Burcu ATAR** danışmanlığında yürüttüğü "**Bireyselleştirilmiş Bilgisayarlı Sınıflama Testlerinde Kullanılan Pratik Kısıtlamalar Altında İki-Üç-Dört Kategorili Sınıflamaları Karşılaştırılması**" başlıklı tez çalışması, Üniversitemiz Senatosu Etik Komisyonunun **8 Ocak 2019** tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini saygılarımla rica ederim.

e-İmzalıdır
Prof. Dr. Rahime Meral NOHUTCU
Rektör Yardımcısı

Evrakın elektronik imzalı suretine <https://belgedogrulama.hacettepe.edu.tr> adresinden 9e251e78-b1df-4c30-8ddb-1705f8944f95 kodu ile erişebilirsiniz. Bu belge 5070 sayılı Elektronik İmza Kanunu'na uygun olarak Güvenli Elektronik İmza ile imzalanmıştır.

Hacettepe Üniversitesi Rektörlük 06100 Sıhhiye-Ankara
Telefon:0 (312) 305 3001-3002 Faks:0 (312) 311 9992 E-posta:yazimd@hacettepe.edu.tr İnternet
Adresi: www.hacettepe.edu.tr

Duygu Didem İLFRİ



Evrakın elektronik imzalı suretine <https://belgedogrulama.hacettepe.edu.tr> adresinden 5570ce9b-17d4-4fa9-8c82-2ce9db908c77 kodu ile erişebilirsiniz. Bu belge 5070 sayılı Elektronik İmza Kanunu'na uygun olarak Güvenli Elektronik İmza ile imzalanmıştır.

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 yararlanılması 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.

12/11/2019


Seda DEMİR

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

12/11/2019

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

Tez Başlığı : Bireyselleştirilmiş Bilgisayarlı Sınıflama Testlerinde Sınıflama Doğruluğunun
İncelenmesi

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
11/11 /2019	107	194501	24/10 /2019	%2	1211265363

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 Esasları'nı 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: Seda DEMİR

Öğrenci No.: N14244989

Ana Bilim Dalı: Eğitim Bilimleri

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

İmza



DANIŞMAN ONAYI



UYGUNDUR.
(Doç. Dr. Burcu ATAR)

EK-E: Thesis/Dissertation Originality Report

12/11/2019

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

Thesis Title: Investigation Of Classification Accuracy At Computerized Adaptive Classification
Tests

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
11/11 /2019	107	194501	24/10 /2019	%2	1211265363

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: Seda DEMİR
Student No.: N14244989
Department: Educational Sciences
Program: Measurement and Evaluation in Education
Status: Masters Ph.D. Integrated Ph.D.

Signature



ADVISOR APPROVAL



APPROVED

(Doç. Dr. Burcu ATAR)

EK-F: Yayınlama ve Fikri Mülkiyet Hakları Beyanı

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

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.

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

12 /11 /2019



Seda DEMİR

"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 tezin erişime açılmasının ertelenmesine karar verebilir.

(2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.

(3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü 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 tarafından karar verilir.

