



HACETTEPE ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

BİLGİSAYARDA BİREYSELLEŞTİRİLMİŞ TEST UYGULAMALARINDA KAPSAM DENGELEMENİN ÖLÇME KESİNLİĞİNE ETKİSİ

İlkay ÜÇGÜL ÖCAL

Doktora Tezi

Ankara, 2024



Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

Daha ileriye... En iyiye...



HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

BİLGİSAYARDA BİREYSELLEŞTİRİLMİŞ TEST UYGULAMALARINDA KAPSAM
DENGELEMENİN ÖLÇME KESİNLİĞİNE ETKİSİ

EFFECTS OF CONTENT BALANCING ON MEASUREMENT PRECISION IN
COMPUTERIZED ADAPTIVE TESTS

İlkay ÜÇGÜL ÖCAL

Doktora Tezi

Ankara, 2024

Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

İlkay ÜÇG¼L ÖCAL'ın hazırladıđı "Bilgisayarda Bireyselleřtirilmiř Test Uygulamalarında Kapsam Dengelemenin Ölçme Kesinliđine Etkisi" bařlıklı bu çalıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölçme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiřtir.

J¼ri Bařkanı Prof. Dr. Selahattin GELBAL İmza

J¼ri Üyesi (Danıřman) Prof. Dr. Nuri DOĐAN İmza

J¼ri Üyesi Doç. Dr. Eren Can AYBEK İmza

J¼ri Üyesi Doç. Dr. Sevda ÇETİN İmza

J¼ri Üyesi Doç. Dr. Celal Deha DOĐAN İmza

Bu tez Hacettepe Üniversitesi Lisansüstü Eđitim, Öğretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından 28 / 11 / 2024 tarihinde uygun gör¼lm¼ř ve Enstitü Yönetim Kurulunca / / tarihi itibarıyla kabul edilmiřtir.

Prof. Dr. İsmail Hakkı MİRİCİ
Eđitim Bilimleri Enstitüsü M¼d¼r¼

Öz

Bu arařtırmada, bireyselleřtirilmiř bilgisayarlı test (BBT) uygulamasında kapsam dengeleme yapılmasının ölçme kesinlięi kestirim deęerlerini; farklı örneklem büyüklüklerine, yetenek kestirim yöntemlerine, madde seęim yöntemlerine ve sonlandırma kurallarına göre nasıl etkiledięinin incelenmesi amaçlanmıřtır. Bu amaçla, simülatif olarak 250 ve 500 kiřilik iki farklı gruba ait yetenek parametreleri ile üç parametrelili lojistik (3PL) madde tepki modeli kullanılarak iki kategorili puanlanan beř farklı içerik alanı, eřit oranlarda aęırlıklandırılarak 750 maddelik madde havuzu oluşturulmuřtur. Kapsam dengelemenin yapıldıęı ve yapılmadıęı her iki durum için; 250 ve 500 kiřilik örnekleme, farklı yetenek kestirim yöntemleri, madde seęim yöntemleri ve test sonlandırma kuralları kullanılarak oluşturulan kořullar altında BBT uygulamaları geręekleřtirilmiřtir. alıřma kapsamında oluşturulan 80 kořula ait BBT uygulaması, her bir kořul için 50 replikasyon olacak řekilde R programlama dilinde yer alan "catR" paketi kullanılarak karřılařtırılmıřtır. Yapılan simülasyonlar sonucunda her bir kořul için ölçme kesinlięine ait RMSE, yanlılık ve korelasyon deęerleri hesaplanmıřtır. alıřmadan elde edilen sonuçlar kapsam dengelemenin genel olarak ortalama hata ve yanlılık deęerlerinde hafif artışlara neden olduęunu göstermiřtir. Bu küçük artışların yanı sıra kapsam dengeleme yapılan ve yapılmayan farklı test kořullarında tahmin edilen ve geręek yetenek düzeyleri arasında yüksek korelasyon elde edilmiřtir. Ayrıca RMSE, yanlılık ve korelasyon deęerleri bakımından, alıřmaya dahil edilen yetenek kestirim yöntemleri arasında ise Beklenen Sonsal Daęılım yönteminin daha etkili olduęu görölmüřtür. Arařtırmada, kapsam dengelemenin, belirli bir doęruluk eřięine ulařmak için, özellikle standart hatanın belirli bir eřik deęerin altına düřtüęü sıkı sonlandırma kuralları altında, gereken madde sayısını yaklaşık 1 madde kadar artırdıęı bulunmuřtur.

Anahtar sözcükler: bilgisayarda bireyselleřtirilmiř testler, kapsam dengeleme, madde seęim yöntemleri, yetenek kestirim yöntemleri, sonlandırma kuralları

Abstract

In this study, the aim is to examine how applying content balancing in computerized adaptive testing (CAT) affects measurement accuracy estimation values based on different sample sizes, ability estimation methods, item selection methods, and stopping rules. For this purpose, a 750-item pool was created using ability parameters for two different simulated groups of 250 and 500 individuals. This pool, weighted equally across five different content areas scored in two categories, was generated using the three-parameter logistic (3PL) item response model. CAT applications were conducted under conditions where content balancing was and was not applied for both 250- and 500-person samples, using different ability estimation methods, item selection methods, and test stopping rules. The CAT applications for the 80 conditions created in the study were compared using the “catR” package in the R programming language, with 50 replications for each condition. The simulations calculated RMSE, bias, and correlation values for measurement accuracy in each condition. Results indicated that, in general, content balancing led to slight increases in average error and bias values. Despite these small increases, a high correlation was observed between the estimated and actual ability levels across different test conditions with and without content balancing. Additionally, among the ability estimation methods included in the study, Expected a Posteriori method proved to be more effective in terms of RMSE, bias, and correlation values. The research also found that content balancing increased the required number of items by about one item to reach a specific accuracy threshold, especially under strict stopping rules where the standard error falls below a set threshold value.

Keywords: computerized adaptive test, content balancing, item selection methods, ability estimation methods, stopping rules.

Teşekkür

Öncelikle, bu uzun ve zorlu yolculuk boyunca bilgi ve deneyimleriyle bana rehberlik eden, sabır ve hoşgörüsüyle yolumu aydınlatan danışmanım Sayın Nuri DOĞAN'a sonsuz şükranlarımı sunarım. Rehberliğiniz ve verdiğiniz destek olmadan bu çalışmayı hayata geçirmek benim için çok zor olurdu.

Tez İzleme Komisyonu'mda yer alarak çalışmamın her aşamasında dikkatle değerlendirmelerde bulunan, farklı bakış açıları kazandırarak tezimin çok daha nitelikli bir hale gelmesini sağlayan Sayın Prof. Dr. Selahattin GELBAL ve Doç. Dr. Eren Can AYBEK'e içten teşekkürlerimi sunarım.

Tez savunma jürimde yer alarak çalışmamın son halinin şekillenmesinde değerli katkılar sağlayan Sayın Doç. Dr. Celal Deha DOĞAN ve Doç. Dr. Sevda ÇETİN'e çok teşekkür ederim.

Bu süreçte desteğini hiç eksik etmeyen sevgili arkadaşım Seçil YILDIRIM PALABIYIK'a hem akademik hem de duygusal anlamda yanımda olduğu için teşekkür ederim. Yorucu günlerde verdiği destek, yüzüme yerleştirdiği bir tebessüm bile benim için çok kıymetliydi.

Beni bugünlere getiren, tükenmeyen sevgi ve dualarını hep hissettiğim sevgili annem Safiye ÜÇGÜL'e ve babam Mehmet ÜÇGÜL'e, desteğini hiçbir zaman esirgemeyen biricik kardeşim Ahmet Yücel ÜÇGÜL'e çok teşekkür ederim. Her adımda yanımda olduğunuzu bilmek, bana güç verdi.

Son olarak, her zaman yanımda olan ve bu yolculuğun her anında bana destek olan sevgili eşim Altuğ ÖCAL'a, kızlarım Aylin ve Pelin ÖCAL'a en derin teşekkürlerimi sunuyorum.

Sabırla beni dinlediğiniz, her zorlukta elimden tuttuğunuz ve başarım için kendi hayatınızdan fedakârlık ettiğiniz için minnettarım. Sizlerin sevgisi ve desteği, benim için her zaman en büyük motivasyon kaynağı oldu.

Bu tez, yalnızca benim çabamın değil, birlikte başardığımız bir emeğin ürünüdür. Hepinize sonsuz teşekkürler!

İçindekiler

Kabul ve Onay.....	ii
Öz.....	iii
Abstract.....	iv
Teşekkür.....	v
Tablolar Dizini.....	ix
Şekiller Dizini.....	x
Simgeler ve Kısaltmalar Dizini.....	xii
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	3
Araştırma Problemi.....	4
Alt Problemler.....	4
Sınırlılıklar.....	4
Tanımlar.....	5
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	6
Madde Tepki Kuramı.....	6
Madde Tepki Kuramı Varsayımları.....	9
Madde Tepki Kuramı Modelleri.....	10
Bilgisayarda Bireyselleştirilmiş Testler (BBT).....	14
Bilgisayarda Bireyselleştirilmiş Testlerin Bileşenleri.....	20
Madde Havuzu.....	21
Başlangıç Kuralı.....	23
Yetenek Kestirim Yöntemleri.....	24
Madde Seçim Yöntemi.....	27
Kapsam Dengeleme.....	30

Madde Kullanım Sıklığı Kontrolü	32
Sonlandırma Kuralı	35
İlgili Araştırmalar	38
İlgili Araştırmalar Özet.....	43
Bölüm 3 Yöntem.....	45
Araştırma Yöntemi	45
Verilerin Üretilmesi.....	45
Yetenek Parametrelerinin Üretilmesi.....	45
Madde Havuzunun Oluşturulması	46
BBT Simülasyon Koşulları.....	47
Verilerin Analizi	48
Bölüm 4 Bulgular, Yorumlar ve Tartışma.....	50
Birinci Alt Probleme Ait Bulgular ve Yorumlar	50
İkinci Alt Probleme Ait Bulgular ve Yorumlar	55
Üçüncü Alt Probleme Ait Bulgular ve Yorumlar	61
Dördüncü Alt Probleme Ait Bulgular ve Yorumlar.....	66
Bölüm 5 Sonuç ve Öneriler.....	82
Sonuçlar	82
Öneriler	85
Kaynaklar	89
EK-A: Madde Havuzunun Farklı İçerik Alanlarına Ait Maddelerin Betimsel İstatistikleri.....	98
EK-B: BSD Yöntemi Kullanılan Koşullardan Elde Edilen Ölçme Kesinliğine İlişkin Değerler	99
EK-C: MOK Yöntemi Kullanılan Koşullardan Elde Edilen Ölçme Kesinliğine İlişkin Değerler	100
EK-Ç: BSD Yöntemi Kullanılan Koşullardan Elde Edilen Ölçme Kesinliğine İlişkin Değerlere Ait Standart Hatalar	101

EK-D: MOK Yöntemi Kullanılan Koşullardan Elde Edilen Ölçme Kesinliğine İlişkin Değerlere Ait Standart Hatalar	102
EK E: Araştırma Etik Komisyon İzin Muafiyeti Formu	103
EK-F: Etik Beyanı	104
EK-G: Doktora Tez Çalışması Orijinallik Raporu	105
EK-H: Dissertation Originality Report	106
EK-I: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	107



Tablolar Dizini

Tablo 1 <i>Madde Havuzundaki Maddelerin Betimsel İstatistikleri</i>	46
Tablo 2 <i>BBT Simülasyon Koşulları</i>	47



Şekiller Dizini

Şekil 1 Madde Karakteristik Eğrisi Örneği.....	8
Şekil 2 1 PLM'ye Ait Madde Karakteristik Eğrileri	12
Şekil 3 2 PLM'ye Ait Madde Karakteristik Eğrileri	13
Şekil 4 3 PLM'ye Ait Madde Karakteristik Eğrileri	14
Şekil 5 Uyarlanabilir Bir Testin Şematik Gösterimi.....	20
Şekil 6 Küçük Örneklemden Elde Edilen RMSE Değerleri.....	50
Şekil 7 Geniş Örneklemden Elde Edilen RMSE Değerleri	51
Şekil 8 Küçük Örneklemden Elde Edilen Yanlılık Değerleri	52
Şekil 9 Geniş Örneklemden Elde Edilen Yanlılık Değerleri.....	53
Şekil 10 Küçük Örneklemden Elde Edilen Uyum Değerleri.....	54
Şekil 11 Geniş Örneklemden Elde Edilen Uyum Değerleri	54
Şekil 12 Maksimum Olabilirlik Kestiriminden Elde Edilen RMSE Değerleri.....	56
Şekil 13 Beklenen Sonsal Dağılımdan Elde Edilen RMSE Değerleri	56
Şekil 14 Maksimum Olabilirlik Kestiriminden Elde Edilen Yanlılık Değerleri	58
Şekil 15 Beklenen Sonsal Dağılımdan Elde Edilen Yanlılık Değerleri.....	58
Şekil 16 Maksimum Olabilirlik Kestiriminden Elde Edilen Uyum Değerleri.....	59
Şekil 17 Beklenen Sonsal Dağılımdan Elde Edilen RMSE Değerleri	60
Şekil 18 Maksimum Fisher Bilgisinden Elde Edilen RMSE Değerleri.....	61
Şekil 19 Kullback Leibler Bilgisinden Elde Edilen RMSE Değerleri.....	62
Şekil 20 Maksimum Fisher Bilgisinden Elde Edilen Yanlılık Değerleri	63
Şekil 21 Kullback Leibler Bilgisinden Elde Edilen Yanlılık Değerleri	63
Şekil 22 Maksimum Fisher Bilgisinden Elde Edilen Uyum Değerleri.....	65
Şekil 23 Kullback Leibler Bilgisinden Elde Edilen Uyum Değerleri.....	65
Şekil 24 Kısa Testten Elde Edilen RMSE Değerleri	66
Şekil 25 Uzun Testten Elde Edilen RMSE Değerleri	67
Şekil 26 $SH \leq 30$ Kuralından Elde Edilen RMSE Değerleri.....	68
Şekil 27 $SH \leq 40$ Kuralından Elde Edilen RMSE Değerleri	68
Şekil 28 $SH \leq 50$ Kuralından Elde Edilen RMSE Değerleri	69
Şekil 29 Kısa Testten Elde Edilen Yanlılık Değerleri.....	70
Şekil 30 Uzun Testten Elde Edilen Yanlılık Değerleri	71
Şekil 31 $SH \leq 30$ Kuralından Elde Edilen Yanlılık Değerleri	72
Şekil 32 $SH \leq 40$ Kuralından Elde Edilen Yanlılık Değerleri	72

Şekil 33 <i>SH≤.50 Kuralından Elde Edilen Yanlılık Değerleri</i>	73
Şekil 34 <i>Kısa Testten Elde Edilen Uyum Değerleri</i>	74
Şekil 35 <i>Uzun Testten Elde Edilen Uyum Değerleri</i>	75
Şekil 36 <i>SH≤.30 Kuralından Elde Edilen Uyum Değerleri</i>	76
Şekil 37 <i>SH≤.40 Kuralından Elde Edilen Uyum Değerleri</i>	76
Şekil 38 <i>SH≤.50 Kuralından Elde Edilen Uyum Değerleri</i>	77
Şekil 39 <i>Farklı Standart Hata Değerlerine Ait Ortalama Madde Sayıları</i>	78



Simgeler ve Kısaltmalar Dizini

BBT: Bilgisayarda Bireyselleştirilmiş Test (CAT: Computerized Adaptive Test)

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

KBBT: Kısıtlı Bilgisayarda Bireyselleştirilmiş Test (CCAT: Constrained Computerized Adaptive Test)

KTK: Klasik Test Kuramı

MÇM: Modifiye Edilmiş Çok Terimli Model (MMM: Modified Multinomial Model)

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

MKBBT: Modifiye Edilmiş Kısıtlı Bilgisayarda Bireyselleştirilmiş Test (MCCAT: Modified Constrained Computerized Adaptive Test)

KLB: Kullback-Leibler

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

MSK: Maksimum Sonsal Kestirim (MAP: Maksimum a Posteriori)

MTK: Madde Tepki Kuramı (Item Response Theory)

OAB: Olabilirlikle Ağırlıklandırılmış Bilgi (WLE: Weighted Likelihood Estimation)

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

3PLM: Üç Parametrelili Lojistik Model

Bölüm 1

Giriş

Bu bölümde; araştırmanın temel aldığı problem durumu, amacı ve önemi, problem cümlesi, alt problemler, sayıltı ve sınırlılıklara yer verilmiştir.

Problem Durumu

Eğitimde kullanılan testler kâğıt ve kalem testleri ve performans değerlendirmelerine odaklanmışken kişisel bilgisayarların eğitimde yaygınlaşmasının başladığı 1980'lerin sonlarından bu yana, bu test formatları hızla bilgisayar tarafından verilmeye uygun formatlara dönüştürülmeye başlamıştır. Geleneksel testler tüm katılımcıların tamamen aynı maddeleri veya paralel maddeleri aldığı varsayımına dayanmaktadır. Bu durumda çoğu madde gizil özelliğın belirli aralıklarında bireyler arasında ayırım yapmada faydalı olamaması nedeniyle etkili bir psikolojik ölçüm sağlayamamaktadır. Madde tepki kuramının (MTK) ve daha hızlı ve büyük kapasiteli bilgisayar teknolojisinin geliştirilmesi, bilgisayarda bireyselleştirilmiş testlerin (BBT) geliştirilmesine olanak tanımıştır. BBT'de, testi alan tüm katılımcıların bireysel olarak iyi ölçüm sağlayacak olan testi aldığı algoritma oluşturmak için, MTK değişmezlik özelliği kullanılır. Test sürecinde kâğıt ve kalem yerine bilgisayar kullanıldığından "Bilgisayarlı" test olarak, uygulanan maddeler testi alanların yeteneklerine göre bağımsız olarak seçilip uygulandığından "Uyarlanabilir" test olarak adlandırılmıştır. BBT, testi alanların cevaplarına göre maddelerin belirlendiği bir testtir (Suhardi, 2020). Bilgisayar tabanlı sunulan testler isteğe bağılı olarak, yani sınava giren kişinin sınava girmeye hazır olduğu her yerde ve zamanda test yapılmasına olanak tanır. Ayrıca, bilgisayarda bireyselleştirilmiş testler kullanılarak test puanlarının istatistiksel doğruluğunu artırmak için bilgisayarlar kullanılabilir. Testi alan her bireye aynı sabit testi vermek yerine, BBT madde seçimi, bireysel sınava girenlerin yetenek düzeyine uyarlanır. Her yanıttan sonra, katılımcının yetenek kestirimi güncellenir ve sonraki madde, yeni kestirimde istenen özelliklere sahip olacak şekilde seçilir. BBT daha kısa test

uzunluğu ile sınava girenlerin belirli bir alandaki yeteneğini daha kesin bir şekilde kestirmek için bir araç sağlar.

Eğitimde ve psikolojide kullanılan testlerde aranan en önemli özelliklerden biri geçerliktir. Geçerlik “kanıt ve teorinin testlerin önerilen kullanımları için test puanlarının yorumlanmasını destekleme derecesini” ifade eder (AERA, APA & NCME, 2014). Geçerliğin, kabul gören bu tanımından yola çıkarak, geniş bir konu olduğu ve değerlendirilebilmesi için birden fazla kanıt türünün gerekli olduğu söylenebilir. Bu kanıt türlerinden biri de ölçülen yapıya ilişkin kanıtlardır. Ölçme Standartları’nda (AERA, APA & NCME, 2014) geçerlik kanıtı türleri; test kapsamı, yanıtlama süreçleri, içyapı, diğer değişkenlerle ilişkiler ve test sonuçları çıkarımları olarak belirtilmiştir. Eğitimde kullanılan testlerde ölçülen yapılar farklı konu ve içerik alanlarının birleşiminden oluşmaktadır. Testin kapsam geçerliğinin sağlanması, bu konu ve içerik alanlarının test içinde yeterli düzeyde temsil edilmesiyle mümkün olmaktadır. Test içeriğinde bir veya daha fazla gerekli alandan madde uygulanmazsa testin kapsam geçerliği sağlanamamış olacaktır ve bu aynı zamanda yapının yetersiz temsili ile sonuçlanacaktır. Bu durum ölçme sonuçlarının doğruluğunu, test puanlarının yorumlanmasını ve testin adilliğini etkileyecektir.

Geleneksel testlerde belirtke tabloları hazırlanarak test ile ölçülecek içerik alanları ve bu alanlara ilişkin madde sayıları belirlenebilmektedir. Böylece testi alan tüm bireylere tüm içerik alanlarından yeterli sayıda madde uygulanmış olmaktadır. Bu yönüyle geleneksel testler diğer test uygulama yöntemlerine kıyasla avantajlıdır. BBT uygulamalarında testi alanların her biri yetenek düzeylerine göre farklı maddeler alabilirler. Kullanılan madde seçim algoritmasına bağlı olarak, en ayırt edici bilgiyi sağlayan maddeler çok sayıda katılımcıya uygulanabilir. Madde kullanım sıklığını ve etkilerini sınırlamak için, madde seçim yönteminin havuz kullanımını göz önünde bulundurarak ayırt edici maddeleri seçmesi gerekir. Ölçme doğruluğunun, adaletinin ve test puanlarına dayalı yapılan yorumların geçerliğinin sağlanabilmesi için BBT uygulamalarında kapsam dengelemeye önem verilmelidir. BBT’nin doğası gereği, aynı testi alan katılımcılara farklı maddeler uygulanır fakat her birine içerik

alanına göre aynı madde dağılımı verilmelidir. Alan yazındaki madde seçimine yönelik çalışmalar göz önüne alındığında madde seçim yöntemlerinin kapsam dengeleme kullanıldığında bireylerin örtük puanları üzerinde nasıl bir etki gösterdiğinin incelendiği çalışmaların yetersiz olduğu görülmüştür.

Araştırmanın Amacı ve Önemi

BBT uygulamalarıyla elde edilen ölçme sonuçlarına dayanılarak alınan kararların tüm eğitim paydaşları üzerinde önemli etkileri mevcuttur. Bu nedenle BBT uygulamalarıyla geçerli ve güvenilir kestirimlerin yapılabilmesi oldukça önemlidir. Test planı veya içerik özelliklerine dayanan geleneksel kâğıt ve kalem testlerinin aksine, bilgisayarda bireyselleştirilmiş testler, madde seçimi sırasında içerik özelliklerini dikkate almaz. Bundan dolayı katılımcılara içerik alanlarına göre farklı madde dağılımı uygulanabilir. Örneğin, bir matematik testinde, katılımcılardan biri tamamen aritmetik maddelerinden oluşan bir test alabilirken bir başka katılımcı tamamen geometri maddelerinden oluşan bir test alabilir. Böyle bir durumda içerik karşılaştırılabilirliğinin olmaması, puanların geçerliği için bir tehdit oluşturabilmektedir. Katılımcılara uygulanan maddelerin içeriğinin dengelenip dengelenmeyeceği BBT uygulaması geliştirirken ele alınması gereken temel konulardan biridir. BBT'nin doğası gereği, sınava girenler aynı testte farklı maddeler alırlar bunun yanı sıra adil bir uygulama için her bir katılımcı her içerik alanından aynı sayıda madde almalıdır (Song, 2010).

Alanyazın incelendiğinde, kapsam dengeleme yapılan ve yapılmayan durumlarda; farklı örneklem büyüklükleri, madde seçme yöntemleri, yetenek kestirim yöntemleri ve sonlandırma kurallarından oluşana simülasyon koşullarının ölçme kesinliğine ve yetenek kestirimine etkisinin incelendiği bir çalışmaya rastlanmamıştır.

Yapılan araştırmada farklı koşullar altında, kapsam dengelemenin, ölçme kesinliğini nasıl etkilediğinin incelenmesi amaçlanmıştır. Böylelikle iki kategorili (1-0) puanlanan maddelerden oluşan BBT uygulamalarında kapsam dengelemenin yapılıp yapılmaması durumunda ortaya çıkan sonuçlar incelenmiştir. Farklı test uzunluklarında, farklı örneklem

büyükliklerinde, farklı madde seçme yöntemleri, yetenek kestirim yöntemleri ve sonlandırma kuralları ile kapsam dengelemenin incelenmesi literatüre getireceği katkı açısından oldukça önemlidir.

Araştırma Problemi

Bilgisayarda bireyselleştirilmiş test uygulamalarında kapsam dengelemesi yapılması; farklı örneklem büyüklüklerine, yetenek kestirim yöntemlerine, madde seçim yöntemlerine ve sonlandırma kurallarına, göre, ölçme kesinliğini ve yetenek kestirimini nasıl etkilemektedir?

Alt Problemler

1. Bilgisayarda bireyselleştirilmiş test uygulamalarında, farklı örneklem büyüklüklerinde (N=250, N=500) kapsam dengeleme yapılması ölçme kesinliği ve yetenek kestirimini nasıl etkilemektedir?
2. Bilgisayarda bireyselleştirilmiş test uygulamalarında, farklı yetenek kestirim yöntemlerinde (MOK, BSD) kapsam dengeleme yapılması ölçme kesinliği ve yetenek kestirimini nasıl etkilemektedir?
3. Bilgisayarda bireyselleştirilmiş test uygulamalarında, farklı madde seçim yöntemlerinde (MFB, KL) kapsam dengeleme yapılması ölçme kesinliği ve yetenek kestirimini nasıl etkilemektedir?
4. Bilgisayarda bireyselleştirilmiş test uygulamalarında, farklı sonlandırma kurallarında (20 madde, 60 madde, $SH \leq 30$, $SH \leq 40$, $SH \leq 50$) kapsam dengeleme yapılması ölçme kesinliği ve yetenek kestirimini nasıl etkilemektedir?

Sınırlılıklar

Araştırma,

1. İki kategorili 1-0 puanlanan maddelerden oluşan simülatif madde havuzu kullanımı,
2. Test başlatma kuralının $-0.5 \leq b \leq 0.5$ olarak belirlenmesi,

3. Madde kullanım sıklığı kontrolü yapılmaması,
4. “catR” paketinde yer alan “Kısıtlı Bilgisayarda Bireyselleştirilmiş Test” yöntemi kullanılarak kapsam dengeleme yapılması ile sınırlandırılacaktır.

Tanımlar

Kapsam Dengeleme: Bilgisayarda bireyselleştirilmiş testlerde her bir içerik alanının önceden tanımlanmış oranlarda temsil edilmesiyle testin geçerlik, güvenilirlik ve adilliğini sağlamaya yönelik bir madde seçimi stratejisidir.



Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde Bilgisayarda Bireyselleştirilmiş Test uygulamalarının temel aldığı ölçme kuramlarından Madde Tepki Kuramı, BBT uygulamalarının bileşenleri ve ilgili araştırmalara yer verilmiştir.

Madde Tepki Kuramı

Psikolojik yapılar genellikle davranışın altında yatan gizil değişkenler olarak tanımlanır. Gizil değişkenler, test puanları veya madde yanıtları gibi gözlemlenebilir değişkenleri etkileyen gözlemlenemeyen yapılardır. Bir maddeye verilen yanıt veya test puanı, bireyin gizil değişken üzerindeki konumunun bir göstergesidir ancak gizil değişkeni tam olarak tanımlamaz. Psikolojik yapıların ölçümleri genellikle dolaylıdır; gizil değişkenler, ilgili maddeler üzerindeki tepkiler gözlemlenerek ölçülür. Hem bireylerin hem de psikolojik boyuttaki maddelerin özellikleri tepkilerden yola çıkarak kestirilir. Bu nedenle psikolojideki bir ölçme kuramı, davranışları psikolojik yapıyla ilişkilendirmek için bir gerekçe sağlamalıdır. Özellik doğrudan ölçülemediğinden, gizil özellikler veya yetenekler olarak adlandırılır (Embretson & Reise, 2000). Gözlenen değişkenler ile gözlemlenmeyen gizli özellikler arasındaki ilişkileri tanımlamak için yaygın olarak Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) kullanılmaktadır. KTK'da madde güçlüğü, madde ayırt ediciliği gibi madde parametreleri, testi alan gruba bağlıdır ve farklı yetenek düzeyine sahip gruplarda değişkenlik gösterebilir. Madde, yetenek düzeyi yüksek olan bir gruba uygulandığında kolay bir madde olarak kabul edilir ve güçlük düzeyi düşüktür. Bunun yanında aynı madde düşük yetenek düzeyine sahip bireylerden oluşan bir gruba uygulandığında zor yani güçlük düzeyi yüksek bir madde olarak kabul edilmektedir. Ayrıca bireylerin yetenek düzeyi kestirimleri de uygulanan teste bağlıdır yani farklı maddelerden oluşan farklı testler uygulanan kişiler arasında karşılaştırma yapmak çok güçtür. Bilgisayarda bireyselleştirilmiş testler maddelere bağlı olduğundan KTK yerine, gruptan

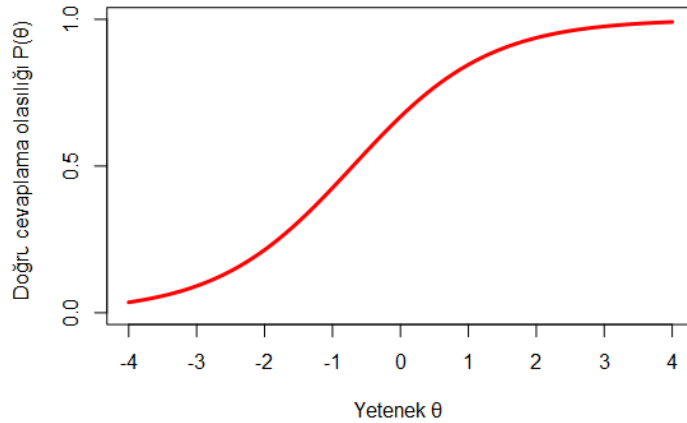
bağımsız madde parametrelerine, maddeden bağımsız yetenek kestirimlerine ve bireysel güvenilirlik tahminlerine imkân veren MTK kullanılmaktadır.

Madde tepki kuramı, gözlemlenebilir test performansı ile test performansının altında yatan gözlemlenemeyen özellikler veya yetenekler arasındaki ilişkiyi belirtir. Madde tepki kuramı bireyin test performansının, özellikler, gizli özellikler veya yetenekler olarak adlandırılan bir dizi faktör tarafından tahmin edilebileceğini, açıklanabileceğini varsayar. Bunun yanında incelenen madde performansı ile madde performansını etkilediği varsayılan bir dizi özellik arasındaki ilişki, madde karakteristik fonksiyonu adı verilen monoton artan bir fonksiyonla tanımlanır. Bu fonksiyon söz konusu gizli özelliğten yüksek puan alan bireylerin, daha düşük puan alanlara göre maddeyi doğru yanıtlama olasılıklarının daha yüksek olduğunu belirtir (Hambleton & Swaminathan, 1985). MTK'nın gizli yetenek düzeyleri ile herhangi bir maddeye verilen yanıtlar arasındaki ilişkiyi tanımlayan güçlü bir matematiksel modeldir. Bu ilişki madde karakteristik eğrileri (MKE) ile tanımlanmaktadır. Madde karakteristik eğrisi, madde puanlarının ölçülen gizli özellik üzerindeki doğrusal olmayan regresyondur (Crocker & Algina, 1986). Madde karakteristik eğrisi gizli yetenek düzeylerine bağlı olarak bir maddeye verilen belirli bir yanıtın beklenen olasılığını gösteren s şeklinde lojistik bir eğridir. MTK'da madde parametrelerinin testin uygulandığı gruba göre değişmezlik özelliği mevcuttur. Yani farklı bireylerden oluşan gruplardan elde edilen madde parametreleri aynıdır. MTK testlerde kullanılan maddelerden bağımsız yetenek kestirimleri sağlar. Bu, aynı bireye iki farklı madde seti uygulandığında bireyin kestirilen yetenek düzeyinin farklı olmayacağı anlamına gelir. Bunun yanında MTK, her yetenek düzeyi için bireysel standart hatayı hesaplamaya imkân verir (Embretson & Reise, 2000; Lord, 1980). Şekil 1'de iki kategorili puanlanan bir maddeye ait madde karakteristik eğrisi verilmiştir. Verilen şekilde, x eksenini bireylerin θ ölçeğinde gizli yetenek düzeyini, y eksenini ise her bir gizli yetenek düzeyi ile ilişkili bu maddeyi doğru yanıtlama olasılığını ($P(\theta)$) temsil etmektedir. Gizli yetenek düzeyi bireyin yeteneğini temsil eder ve teorik olarak $-\infty$ ile $+\infty$ arasında değişir ancak pratikte -3 ile 3 arasında değerler alabilmektedir (Baker,

1992). Şekil 1' de görüldüğü gibi gizil yeteneğin düzeyi arttıkça maddenin doğru cevaplanma olasılığı da artmaktadır.

Şekil 1

Madde Karakteristik Eğrisi Örneği



Madde karakteristik eğrisini madde güçlüğü (b), madde ayırt ediciliği (a) ve sözde tahmin, diğer deyişle şans başarısı (c) parametresi şekillendirmektedir. Madde ayırt edicilik parametresi eğim veya diklik parametresi olarak da adlandırılır. Maddenin testi alanları gizil yetenek düzeyinde ayırt etme yeteneğidir (Emberson & Reise, 2000). Madde ayırt edicilik parametresi genellikle 0.5 ile 2 arasında değerler alır (Baker, 1992). Ayırt edicilik parametresinin düşük olması madde karakteristik eğrisini daha düz hale getirir ve farklı gizil yetenek düzeyleri arasında maddeyi doğru cevaplama olasılıkların neredeyse eşdeğer olduğu anlamına gelir (Baker, 1992). Madde güçlük parametresi eşik ya da konum parametresi olarak da adlandırılır. Bireyin bir maddeyi doğru cevaplama olasılığı 0.5 olduğunda gizil yetenek ölçeğinde karşılık gelen değerdir. Madde güçlük parametresinin düşük olması maddenin kolay, yüksek olması ise zor olduğu anlamına gelmektedir. Bir maddenin güçlük düzeyi arttıkça madde karakteristik eğrisinin konumu gizil yetenek ölçeğinde soldan sağa; azaldıkça sağdan sola kayacaktır. Şans parametresi alt asimptot olarak da adlandırılır. Oldukça düşük yetenek düzeyine sahip bireylerin maddeyi doğru cevaplama oranına karşılık gelmektedir. Şans başarısı ile doğru cevaplanabilecek maddelerde MKE'nin alt asimptotu sıfırdan büyüktür ve yetenek düzeyi çok düşük olsa da maddeyi doğru cevaplama olasılığı hiçbir zaman sıfıra

düşmez (Emberson & Reise, 2000). Bir maddenin şans parametresinin .20 olması, yetenek düzeyi ne olursa olsun testi alan herhangi bir bireyin maddeyi doğru cevaplama olasılığının en az %20 olduğu anlamına gelir. Testi alan bireyin herhangi bir maddeyi doğru cevaplama olasılığı yalnızca madde karakteristik eğrisinin biçimine bağlı olduğundan, bu olasılık, testi alan grubun yetenek dağılımından bağımsızdır. Bu nedenle bireyin bir maddeyi doğru cevaplama olasılığı, aynı yetenek düzeyine sahip diğer bireylerin sayısına bağlı olmayacaktır. Maddelerin kalibre edildiği madde karakteristik eğrilerinin bu değişmezlik özelliği MTK modellerinin en cazip özelliklerinden biridir (Hambleton & Swaminathan, 1985).

Madde Tepki Kuramı Varsayımları

MTK'nın avantajlarından faydalanabilmek için tek boyutluluk, yerel bağımsızlık ve model veri uyumu varsayımlarının karşılanması gerekir (Hambleton & Swaminathan, 1985). Bireyin test performansını açıklamak için yalnızca bir yetenek veya özelliğin gerekli olduğu kabul edilir. Tek bir gizil yeteneği varsayan MTK modelleri tek boyutlu olarak adlandırılır (Crocker & Algina, 1986; Hambleton vd., 1991). Bu varsayımın tam olarak karşılanması güçtür çünkü her zaman test performansını bir dereceye kadar etkileyen başka bilişsel, kişisel faktörler mevcuttur. Bu faktörler arasında motivasyon düzeyi, sınav kaygısı, cevap kağıtlarının doğru kullanımına ilişkin bilgi, test madde seti tarafından ölçülen baskın olanın yanı sıra diğer bilişsel beceriler vb. sayılabilir. Bu varsayımın test verisi tarafından yeterince karşılanması için gerekli olan şey, test performansını etkileyen "baskın" bir bileşen veya faktördür. Bu baskın bileşen veya faktör, test tarafından ölçülen yetenek olarak adlandırılır. Tek boyutluluk varsayımın test edilmesinde kullanılan en yaygın yöntemler maddeler arası korelasyon matrisinin özdeğerlerinin analizi, Stout'un temel tek boyutluluk testi ve tek boyutlu bir modelden elde edilen artıkların analizine dayalı değerlerdir. Stout, yetenek düzeyine bağlı olarak ikili madde kovaryanslarının ortalama mutlak değeri yaklaşık 0 olduğunda tek boyutluluğun sağlanacağını öne sürerek yerel bağımsızlık ve tek boyutluluk arasındaki ilişkiyi vurgulamıştır (DeMars, 2010). Yerel bağımsızlık, test performansını etkileyen gizil yetenek düzeyi sabit tutulduğunda, testi alanların herhangi bir madde çiftine verdiği yanıtların istatistiksel olarak

bağımsız olduğu anlamına gelir. Bu varsayımın gerçekleşmesi için, bireyin bir madde üzerindeki performansı, testteki diğer maddelere verdiği yanıtları iyi ya da kötü yönde etkilememelidir. Yani, testi alanların farklı maddelere verdikleri yanıtlar arasında hiçbir ilişki yoktur. Yerel bağımsızlık varsayımı karşılandığında maddelerin testin tamamına katkısı ayrı ayrı değerlendirilebilir ve madde bazındaki özelliklerin toplanmasıyla test özellikleri elde edilebilir (De Ayala, 2009; Hambleton vd., 1991). Yerel bağımsızlık varsayımının ihlali, maddelerin bir başka madde için ipucu niteliği taşıması, ölçme aracının uzunluğu, aracın uygulanması için tanımlanan sürenin yetersiz olması veya MTK modelinde yetersiz sayıda gizil değişken tanımlanmış olmasından kaynaklı olabilir (De Ayala, 2009). Genel olarak, bu varsayıma göre, sınava girenlerin farklı maddelere verdiği yanıtlar arasında bir ilişki yoktur ve modelde belirtilen yetenekler, sınava girenlerin test maddelerine verdiği yanıtlar üzerinde etkisi olan tek faktördür. Tek boyutluluk varsayımı karşılandığında yerel bağımsızlık varsayımının da karşılandığı kabul edilir (Lord, 1980). Model veri uyumu varsayımı ise elde edilen veriler için en yüksek model veri uyumuna sahip MTK modelinin kullanılması anlamına gelmektedir. Model ile veriler arasındaki uyum, modeldeki hatalı belirlemeleri kontrol etmek için değerlendirilebilir. Örneğin, bir parametrelili lojistik modelin kullanıldığı bir durumda veriler değişen eğimlere veya sıfırdan farklı bir alt asimptot değerine sahipse maddelerin çoğu modele uyum göstermeyecektir. Madde uyumuna ek olarak modelin genel uyumu da dikkate alınabilir. Gözlemlenen toplam puan dağılımının model tarafından tahmin edilen puan dağılımıyla karşılaştırılmasını içerir. Dağılımlar görsel olarak karşılaştırılabileceği gibi aralarındaki fark χ^2 (ki-kare) testi gibi istatistiksel bir testle de değerlendirilebilir (DeMars, 2010).

Madde Tepki Kuramı Modelleri

MTK modelleri iki kategorili ve çok kategorili modeller olarak sınıflandırılabilir. Bireyin verdiği cevaplarla ilgili doğru-yanlış, evet-hayır gibi iki kategorili bir puanlama söz konusu olduğunda iki kategorili MTK modelleri kullanılır. Çok kategorili puanlama çoktan seçmeli test maddelerinde olası yanıtlara puanlama ağırlıkları eklendiğinde ortaya çıkar. Açık uçlu sorular ve Likert ölçekleri için puanlama sistemi genellikle çok kategorilidir. İki kategorili modellerden

bazıları; bir parametrelili (1PLM), iki parametrelili (2PLM), üç parametrelili (3PLM) lojistik modellerdir.

Bir Parametrelili Lojistik Model (1PLM)

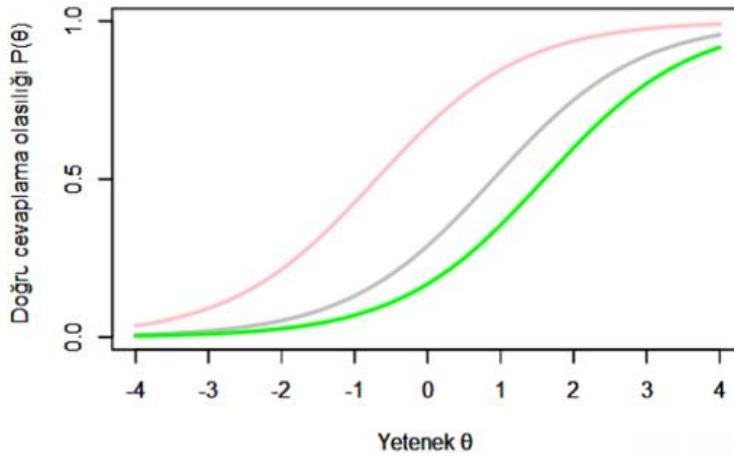
Bireylerin yetenekleri yalnızca madde güçlük parametresi (b_i) hesaplanarak kestirilir. Maddelere ait ayırt edicilik parametreleri (a_i) eşit kabul edilir ve hesaplama dahil edilmez. Maddelere ait ayırt edicilik parametrelerinin eşit ve 1'e eşit olduğunun varsayıldığı model ise Rasch Model olarak adlandırılır. Bir parametrelili lojistik model matematiksel olarak şu şekilde tanımlanmıştır:

$$P_i(\theta) = \frac{e^{D\bar{a}(\theta-b_i)}}{1 + e^{D\bar{a}(\theta-b_i)}}$$

$P_i(\theta)$, θ yetenek seviyesinde rastgele seçilen bir bireyin i maddesine doğru yanıt verme olasılığı, b_i , i maddesine ait güçlük parametresidir. Bu değer teorik olarak $-\infty$ ile $+\infty$ aralığındadır fakat hedeflenen test evreni için maddenin çok kolay ya da çok zor olmaması istendiğinden genellikle -2 ile +2 arasında değer alır. Madde güçlük parametresi -2'ye yakın maddeler çok kolay, +2'ye yakın maddeler çok zordur (Demars, 2010). e değeri 2,718 olan sabit bir sayıdır, D ise ölçekleme parametresidir ve değeri 1,7'dir. Şekil 2'de farklı güçlük parametrelerine sahip üç maddeye ait madde karakteristik eğrileri verilmiştir. ($b = -0.70$, $b = 0.90$, $b = 1.6$, $a = 1$)

Şekil 2

1 PLM'ye Ait Madde Karakteristik Eğrileri



Şekilde görüldüğü gibi 1PLM'de madde ayırt edicilik parametreleri eşit olduğundan eğriler keskişmemektedir. Madde güçlük parametresi en yüksek olan yani en zor olan maddeye ilişkin eğri en sağda konumlanmıştır.

İki Parametrelili Lojistik Model (2PLM)

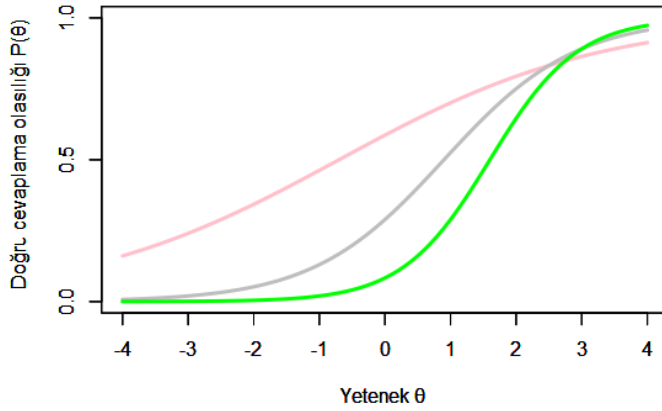
İki parametrelili lojistik modelin 1PLM'den farkı madde güçlük parametresine ek olarak madde ayırt edicilik parametresinin de hesaba katılmasıdır. 2PLM'de madde tepki fonksiyonlarının eğimleri birbirinden farklıdır. İki parametrelili lojistik model matematiksel olarak şu şekilde tanımlanmıştır:

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$$

1PLM'den farklı olarak hesaplamaya ayırt edicilik parametresinin de eklendiği a parametresine alt indis eklenmesiyle gösterilmiştir. Madde ayırt ediciliğini gösteren eğim ya da a parametresi teorik olarak $-\infty$ ile $+\infty$ aralığındadır fakat uygulamadaki aralık 0 ile 2 ya da 3 aralığında olabilmektedir. Negatif ayırt ediciliğe sahip maddeler testten çıkarılmalıdır (Baker, 1992; Demars, 2010). Şekil 3'te güçlük ve ayırt edicilik parametreleri farklı olan üç maddeye ait madde karakteristik eğrileri verilmiştir. ($b = -0.70, a = 0.50$; $b = 0.90, a = 1$; $b = 1.6, a = 1.5$)

Şekil 3

2 PLM'ye Ait Madde Karakteristik Eğrileri



1PLM'den farklı olarak 2PLM'de eğrilerin kesiştiği görülmektedir. Madde karakteristik eğrilerinin eğimi madde ayırt edicilik parametresinin değerini göstermektedir. Madde ayırt edicilik değeri yükseldikçe eğrinin dikleştiği görülmektedir.

Üç Parametrelili Lojistik Model (3PLM)

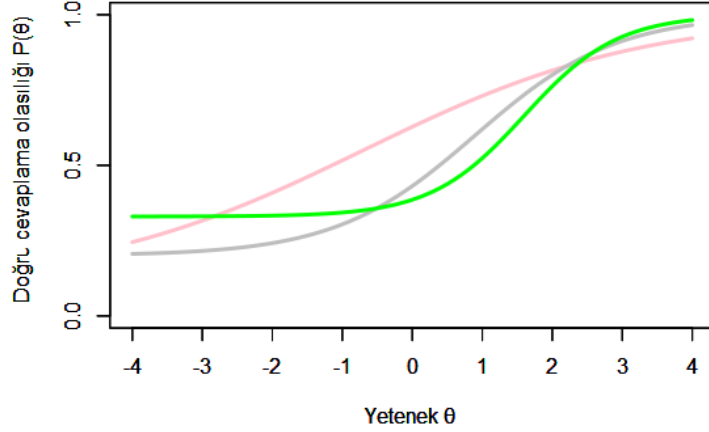
2PLM'ye, şans başarısının katkısını gösteren c parametresi eklenerek üç parametrelili lojistik model geliştirilmiştir. 3PLM, bir ve iki parametrelili lojistik modele ek olarak c_i (pseudo-chance-level) parametresini içermektedir. Bir ve iki parametrelili lojistik modellerde alt asimptot sıfır olarak kabul edilmekte ve şansla doğru yanıt verme olasılığı dikkate alınmamaktadır. 3PLM'de ise madde karakteristik eğrisinde sıfırdan farklı alt asimptot imkânı sunmaktadır. c_i parametresi, madde karakteristik eğrisinin alt asimptotudur ve düşük beceriye sahip adayların bir maddeyi doğru yanıtlama olasılığını temsil eder. Bu parametre, diğer parametrelerin yanında tahmin etmenin test performansında bir faktör olduğu, düşük yetenekli adaylardan gelen madde yanıt verilerini hesaba katmak için modele dahil edilmiştir (Hambleton & Swaminantan, 1985). Üç parametrelili lojistik model matematiksel olarak şu şekilde tanımlanmıştır:

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

Şekil 4'te güçlük, ayırt edicilik ve şans parametreleri farklı olan üç maddeye ait madde karakteristik eğrileri verilmiştir. ($b=-0.70$, $a=0.50$, $c=0.10$; $b=0.90$, $a=1$, $c=0.20$; $b=1.6$, $a=1.5$, $c=0.33$)

Şekil 4

3 PLM'ye Ait Madde Karakteristik Eğrileri



Şans parametresinin değerine göre en düşük yetenek düzeyindeki maddeyi doğru cevaplama olasılığı değişkenlik göstermektedir. Maddenin şans parametre değeri arttıkça en düşük yetenek düzeyinde dahi maddeyi doğru cevaplama olasılığı sıfırdan uzaklaşmaktadır.

Bahsi geçen modellerin en genel hali üç parametrelili lojistik modeldir. 3PLM, farklı test uygulamaları ve BBT uygulamalarında sıklıkla uygulanan MTK modelidir (Wainer & Mislevy, 2000). Bu modelin kendine has varsayımları bulunmamaktadır. Şans parametresinin sıfıra eşit olduğu varsayımı ile iki parametrelili lojistik model elde edilir. Bu modele de ayırt edicilik parametresi değerlerinin eşit olduğu varsayımı eklenerek bir parametrelili lojistik model elde edilmiş olur. MTK'ya dayalı analizlerde parametre ve yetenek kestirimlerinin hata miktarının mümkün oldukça azaltılabilmesi için kullanılan veri setinin tercih edilen modele ait varsayımları karşılaması ve model ile uyumlu olması gerekir.

Bilgisayarda Bireyselleştirilmiş Testler (BBT)

Eğitimde etkili rehberlik, her bir öğrencinin bireysel farklılıkları ve bu farklılıklara konu olan özellikleri hakkında ayrıntılı bilgilerin edinilmesini gerektirir. Çoğu durumda eğitimciler,

rehberlik ve eğitim alanında bilgi edinmek için veri elde etmek amacıyla test, envanter, anket gibi çok çeşitli psikolojik ölçme araçları kullanılmaktadır. Eğitimde kullanılan testler genellikle kâğıt kalem testleri ve performans değerlendirmelerine odaklanmıştır. Kişisel bilgisayarların eğitimde geniş ölçekte yaygınlaşmasının başladığı 1980'lerin sonlarından bu yana, bu test formatları hızla bilgisayar tarafından verilmeye uygun formatlara genişletilmiştir (van der Linden & Glas, 2000). Kâğıt-kalem testleri genel olarak Klasik Test Kuramı kullanılarak geliştirilmektedir. Geleneksel bir testte, her bir katılımcıya uygulanan sabit bir dizi madde söz konusudur. KTK'yi kullanarak, geleneksel test geliştiricileri genellikle bu maddeleri, madde analizine göre seçerek, ölçme aracını oluşturan maddeler kümesinin iç tutarlılık güvenilirliğini en üst düzeye çıkarmayı amaçlar. Bununla yanında, geleneksel testlerin en önemli sınırlılıklarından biri; KTK'ya dayanılarak madde seçimi yapıldığında, iç tutarlılık güvenilirliğini en üst düzeye çıkaran seçilmiş maddeler genellikle gruptaki ortalama yetenek düzeyindeki katılımcılar için uygun maddelerdir. Bu maddeler, sınava giren ortalama yetenek düzeyine sahip katılımcılar için en iyi ölçümü sağlayan maddelerdir fakat ölçülen özellikte ortalama yetenek düzeyinin altında olan katılımcılar için çok zor, üzerinde olan katılımcılar için ise çok kolay olabilmektedir. Testin sınava giren kişiye uyarlanması ilkesi, psikolojik ölçmenin ilk dönemlerinde, geleneksel kâğıt ve kalem testinin geliştirilmesinden önce, Alfred Binet tarafından Binet IQ testinin geliştirilmesinde tanınmıştır (Weiss, 1982). Binet'in test uygulama prosedürü; önceden kalibre edilmiş bir madde havuzunun olması, bireysel olarak uygulanıp tüm katılımcılar için en uygun güçlük düzeyini belirlemek için tasarlanması, değişken başlatma seçeneğinin olması, her bir katılımcı için uygulanacak bir sonraki madde seçiminin belli bir kuralının olması ve uygulanacak bir sonraki madde setinin, katılımcının önceki her bir madde setindeki performansına dayanması ile tamamen uyarlanabilir bir prosedürdür (Davey, 2011; De Ayala, 2009; Weiss, 1982; Weiss & Kingsbury, 1984).

BBT'nin temel varsayımı, yüksek yetenek düzeyine sahip bireyler için kolay maddelerin, benzer şekilde düşük düşük yetenek düzeyine sahip bireyler için zor maddelerin, bireyin yeteneği hakkında çok az veya hiç bilgi vermediğidir. MTK temelinde kalibre edilmiş bir madde

havuzu kullanan BBT algoritması her bireyin en uygun testi almasını sağlar (Chang, 2004; Hambleton vd.,1991) Bilgisayarda bireyselleştirilmiş testler, test maddelerini uygulamak için algoritma tabanlı bir yaklaşım kullanmaktadır. Spesifik olarak, seçilen ve uygulanan maddeler, test süreci sırasında testi alan kişinin kestirilen yetenek düzeyine uyarlanır, her bir madde uygulandıktan sonra kestirilen yetenek sürekli olarak güncellenir. Bu nedenle BBT madde düzeyinde uyarlanabilir bir testtir ve sabit veya değişken uzunlukta olabilir. Yetenek kestirimi yalnızca bir teste giren kişinin yetenek seviyesini temsil etmek için değil, aynı zamanda mevcut bir madde havuzundan sonraki maddelerin seçimini belirlemek için de kullanılır. Bireye uygulanacak bir sonraki madde bireyin yeteneği hakkında en fazla bilgi veren maddedir; bu da genellikle doğru cevaplardan sonra daha zor maddelerin ve yanlış cevaplardan sonra daha kolay maddelerin uygulanması anlamına gelir. Maddenin güçlüğündeki adım adım değişim genellikle başlarda fazla iken adayın yeteneği hakkında daha fazla bilgi edinildikçe azalır. Süreç, kişiyi belirli bir doğruluk düzeyiyle yetenek ölçeğine yerleştirmek için yeterli bilgi edinilene kadar devam eder (Green, vd., 1984). BBT, geleneksel doğrusal testten çok daha kullanışlı ve etkili olabilmektedir. Bu nedenle son yıllarda yaygın olarak kullanılmaktadır. BBT'lerin geleneksel doğrusal testlere göre birçok avantajı olduğu gösterilmiştir. BBT uygulamasının avantajları şunlardır (Economides & Roupas, 2007; Rudner, 1998; Tian, vd., 2007):

- Bilgisayarlı testler, test uygulamasının esnekliğini büyük ölçüde artırır.
- Testler talep üzerine verilir ve puanlar hemen hesaplanabilir.
- Cevap kağıtlarına ya da eğitimli test uygulayıcılarına ihtiyaç yoktur. Test uygulayıcısı farklılıkları, ölçme hatasında bir faktör olarak ortadan kaldırılmıştır. Ancak yine de denetime ihtiyaç vardır çünkü BBT'nin uygulandığı ortam test sonuçlarını etkileyebilir.
- Testlerin uygulanması bireysel olarak hızlandırılmıştır, böylece testi alan kişinin bir sonraki bölüme geçmeden önce diğerlerinin bitirmesini beklemesi gerekmez. Bireysel hıza göre uygulama aynı zamanda ihtiyacı olan bireyler için ekstra zaman sağlar.

- Aynı testi alan başka birinden cevapların kopyalanması gibi bir durum söz konusu olmadığından test güvenliği artırılmış olur.
- Bilgisayarlı test, zamanlama ve biçimlendirme için bir dizi seçenek sunar. Bu nedenle, daha geniş bir madde havuzuna sahiptir.
- Kabul edilebilir ölçüm hassasiyeti elde etmek için daha az maddeye ihtiyaç duyulduğundan, BBT'leri uygulamak için sabit madde testlerinden önemli ölçüde daha az zamana ihtiyaç duyulur. BBT'ler aynı güvenilirlik seviyesini korurken test süresini %50'den fazla azaltabilmektedir.
- Test süresinin kısalması, sınava giren kişinin test sonuçlarını önemli ölçüde etkileyebilecek bir faktör olan yorgunluğu da azaltmaktadır.
- BBT'ler, çok çeşitli yetenek düzeylerinde doğru ölçümler sağlayabilirken, geleneksel testler genellikle ortalama yetenek düzeyinde kullanışlıdır.

BBT uygulamalarının bu avantajlarının yanı sıra bazı sınırlılıkları da mevcuttur (Economides & Roupas, 2007; Rudner, 1998):

- Testin uygulanması esnasında bilgisayar kullanma ihtiyacı, bilgisayar kaygısı olan bireylerde rahatsızlık oluşturabilir.
- BBT uygulamaları testi alan bireylerin kısmen de olsa bilgisayar okur yazarlığına sahip olmasını gerektirir.
- Bilgisayarın donanım sınırlamaları ve maliyeti ekonomik açıdan sorun olabilir.
- Geniş bir madde havuzuna ihtiyaç duyulduğundan bu havuzun oluşturulması zahmetlidir.
- Belirli bir maddeyi atlamak mümkün olmamaktadır, cevap verilmedikçe maddeler arasında geçiş yapılmasına izin verilmez.
- BBT uygulamaları her madde türüne uygun olmayan Madde Tepki Kuramını temel almaktadır. Bu nedenle açık uçlu sorular ve kolayca kalibre edilemeyen maddelerde kullanılamamaktadır.

- Sonraki maddelerin daha önce cevaplanan maddelere göre seçilmesi nedeniyle test katılımcısının geri dönüp cevaplarını değiştirmesine izin verilmemektedir.
- BBT uygulama sürecinde bazı maddeler diğerlerine göre daha sık seçilebilir ve bu maddeler ezberlenerek diğer katılımcılara aktarılabilir.
- Testi alan her bireye farklı sorular uygulanması nedeniyle eşitsizlikler algılanabilir. Bu nedenle tüm katılımcıların nihai yetenek kestiriminin tartışılmaz derecede makul olmasını sağlamak için yeterli sayıda madde uygulanması gerekir.

Bilgisayarda bireyselleştirilmiş testlerin ilk uygulamaları farklı dallandırma yöntemlerine göre oluşturulmuş KTK'ya dayalı uygulamalardır. Fakat yetenek ve parametre değişmezliği özelliği sunduğundan Madde Tepki Kuramı bireye uyarlanmış testler için daha uygun olan psikometrik özellikleri sağlamaktadır. Bireyin aldığı maddelere göre kestirilen yetenek düzeyinin değişmemesi ve madde parametrelerinin uygulandığı grubun özelliklerine göre farklılaşmamasıyla bireyler farklı maddeler almış olsa da yetenek kestiriminin geçerliği olumsuz etkilenmemektedir. MTK'nın sunduğu destekleyici zemin sayesinde 1970'lerin sonlarından itibaren BBT ile ilgili araştırmalar yoğunlaştırılmıştır (Şenel, 2021).

BBT, bireylere geleneksel kâğıt-kalem formatından ziyade bilgisayar aracılığıyla uygulanan bir ölçme aracıdır. BBT'nin ana amacı, her bir katılımcının kendisi için "en uygun" ölçümü sağlayacak olan test algoritmasını oluşturmak için MTK değişmezlik özelliğini kullanmaktır. Yani test, maddelerin katılımcıya ne çok zor ne de çok kolay olmaması için bireysel olarak testi alan bireylere göre "uyarlanır". MTK tabanlı BBT uygulamaları genellikle geleneksel kâğıt kalem ölçümlerinden daha az madde içerir (Embretson & Reise, 2000).

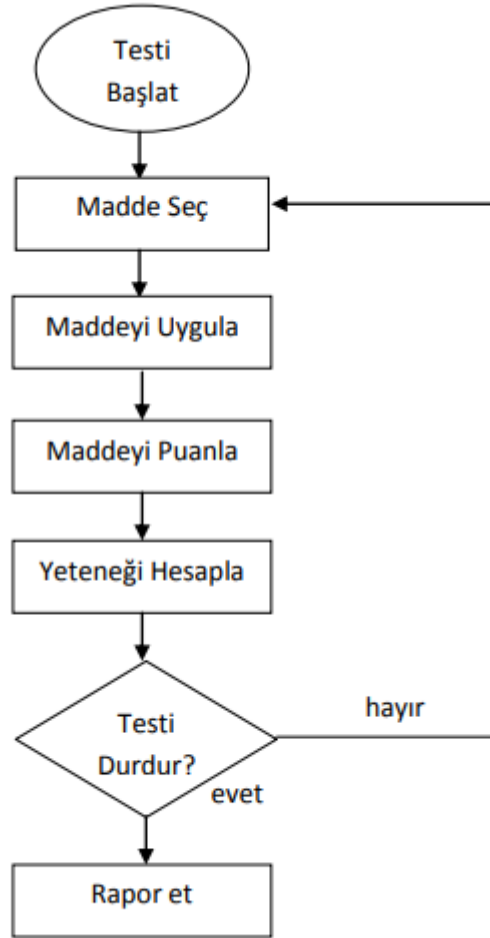
BBT uygulamaları için kullanılan matematiksel model MTK'yı temel alır. MTK metodolojileri, çeşitli BBT süreçlerinde kullanılır ve yetenek kestiriminin doğruluğunu ve verimliliğini artırmaya odaklanır. MTK testi alanın yeteneğinden yola çıkarak, madde üzerindeki performansın olasılığına odaklanmaktadır. MTK'nın en önemli özelliği bireylerin yeteneklerinin, maddelerden bağımsız olarak kestirilmesidir. İdeal bir BBT'de, maddeler her bir katılımcı için

optimal bir sırayla birer birer seçilir ve uygulanır, seçilen her madde, testin mevcut adımında en bilgilendirici maddedir. Ayrıca bir sonraki maddenin seçimi daha önce uygulanan maddelere, sınav katılımcısının yanıtlarına ve sınav katılımcısının yetenek düzeyinin geçici tahminine bağlıdır. BBT uygulamaları bir test algoritması tarafından yönetilir. Algoritma, BBT'lerin nasıl başlatıldığını, sürdürüldüğünü ve sonlandırıldığını belirleyen bir dizi kuralı içermektedir. Eggen (2004) bir BBT algoritmasını Şekil 5'teki gibi ifade etmiştir.

Bireyin yeteneğiyle ilgili önsel bilgiler genellikle mevcut olmadığından BBT, madde havuzundan rastgele seçilen bir maddeyi bireye sunmakla başlar. Bireyin test öncesi yeteneği ile ilgili herhangi bir bilgi varsa bu bilgi ilk madde seçiminde kullanılabilir. Uygulanan her maddeden sonra bir madde seçim prosedürü gerçekleştirilir. Madde havuzundan, bireyin o ana kadar verdiği cevaplara uygun bir madde seçilir, test bireyin yeteneğine göre uyarlanır. BBT temelli bir uygulamada, bireyin performansı yalnızca test sonunda değil, özellikle her bir maddenin uygulanmasından sonra, sınav süresi boyunca sürekli olarak değerlendirilir. Testi alan kişinin performansının ara değerlendirmesine dayanarak BBT algoritması, bir sonraki test maddesini, bireyin yeteneğiyle en alakalı olması beklenen zorluk seviyesinde seçer ve uygular. Bu şekildeki kişiselleştirilmiş uyarlanabilir test yapısı, ölçme aracının verimliliğini önemli ölçüde artırmaktadır (De Ayala, 2009; Han, 2018; Thissen & Mislevy, 2000).

Şekil 5

Uyarlanabilir Bir Testin Şematik Gösterimi



Bilgisayarda Bireyselleştirilmiş Testlerin Bileşenleri

BBT uygulaması süreci farklı bileşenden oluşur. İlk bileşen, kalibre edilmiş bir madde havuzudur. Madde havuzunun oluşturulmasıyla test içeriği geliştirilmiş olur. Diğer bileşenler içerikten ziyade psikometriktir ve BBT sürecindeki algoritmaları oluşturur. BBT süreci bileşenleri şunlardır (Thompson & Weiss, 2011; Weiss & Kingsbury, 1984).

1. Madde Havuzu
2. Başlangıç Kuralı
3. Madde Seçim Yöntemi

4. Kapsam Dengeleme

5. Madde Kullanım Sıklığı Kontrolü

6. Sonlandırma Kuralı

BBT uygulaması, ilk iki bileşeni verilen olarak alıp ardından sonlandırma kriteri sağlanana kadar 3, 4 ve 5. bileşen arasında döngü yaparak ilerler. Bilgisayara her maddeye ilişkin psikometrik verileri içeren madde havuzu önceden yüklenmiştir ve testi alan birey için belirli bir başlangıç noktası belirlenmiş olacaktır. Bu başlangıç noktası için testte uygulanacak ilk madde seçilir. Madde cevaplandıktan sonra puanlanır ve bireyin yeteneğinin (θ) bir kestirimi elde edilir. Daha sonra sonlandırma kriteri değerlendirilir; bu kriter karşılanmadıysa, bireyin cevaplayacağı başka bir madde seçilir ve tekrar bireyin örtük özelliğe ilişkin puanı (θ) güncellenir, sonlandırma kriteri bir kez daha değerlendirilir. Bu döngü sonlandırma kriteri sağlanana kadar devam eder.

Madde Havuzu

Madde havuzu, testi alan bireylere uygulanabilecek maddeler topluluğudur ve uyarlanabilir testler için temel araçtır. Madde havuzunun nitelik ve nicelik açısından gereksinimleri karşılayan maddelerden oluşması son derece önemlidir çünkü bu maddeler havuzun kalitesini belirler ve dolayısıyla BBT sonuçlarını etkiler (Davey, 2011; Wise vd., 2015). Nicel gereksinimler, havuzdaki maddelerin psikometrik özelliklerini temsil eder. Havuzdaki maddeler, gizil özellik ölçeğindeki farklı yetenek düzeylerinde bilgilendirici olmalıdır. BBT'nin temel özelliği, maddeleri testi alan bireylerin yetenek seviyelerine uyarlamak olduğundan, madde havuzunun farklı güçlük seviyelerinde (kolay, orta, zor) çok sayıda maddeyi içermesi gerekir. Ayrıca, testi alan her birey için seçilen madde kümesi, kapsam dengeleme gibi istatistiksel olmayan kısıtlamaları karşılamalıdır (Chang, 2004). Madde havuzunun kalitesi BBT uygulamalarının sonuçları üzerinde önemli bir etkiye sahiptir. BBT uygulamasına başlanmadan önce madde havuzu; madde yazma, gözden geçirme ve saha testi gibi bir dizi işlem yoluyla oluşturulur (Flaughner, 2000). Testin amacına (örn; yeterlilik testi veya sınıflandırma testi) göre

belirlenen test özelliklerine dayalı madde havuzu ve bu havuz için ne tür maddelere ihtiyaç duyulacağı tanımlanmalıdır. Madde yazma aşamasında madde yazarları ve içerik uzmanları, test özelliklerine dayalı olarak çok sayıda madde oluşturur. Madde kalitesini ve adaletini sağlamak için yazılan bu maddeler incelemeye tabi tutulur. Madde incelemesine ek olarak bu maddelerin, madde analizi ve madde kalibrasyonu ile madde özelliklerini incelemek için sahada test edilmesi de gereklidir. BBT uygulamasında, madde havuzunun BBT süreci başlamadan önce kalibre edildiği varsayılır. Yani, bireye uyarlanan testte madde seçimi için bir ön test veya pilot uygulamanın yapıldığı kalibre edilmiş geniş bir madde havuzu mevcuttur. Madde havuzu içerik alanı özelliklerini ve madde biçimlendirmesini temsil eden nitelik ve havuzdaki maddelerin psikometrik özelliklerini temsil eden nicelik bakımından, gereksinimleri karşılayan maddeler topluluğu olmalıdır. Okuma becerisini ölçmek üzere kullanılacak bir testin eleştirel okuma, anlama, paragraf yorumlama, kelime bilgisi ve dilbilgisi gibi nitelikleri temsil eden birçok madde içermesi gerekir. Test geliştiricilerinin test özelliklerini belirlemeleri, her içerik alanında yeterli sayıda madde oluşturmaları, testin amacına ve özelliklerine göre madde kalitesini, maddelerin adil olup olmadığını gözden geçirmeleri ve yeni yazılan maddelerin ön testini veya pilot uygulamasını gerçekleştirmeleri gerekir. Böylece madde havuzuna sadece seçilmiş nitelikli maddeler eklenmiş; ayrıca madde havuzu, boyutu, özellikleri ve içerik dengesi açısından periyodik olarak değerlendirilerek yenilenmiş olur.

Üzerinde durulması gereken bir diğer konu da madde havuzunun genişliğidir. Madde havuzu ne kadar genişse BBT süreci için o kadar iyidir. Ancak son derece geniş bir madde havuzunu uygulamak her zaman mümkün değildir. İşlevsel bir uyarlanabilir test için madde havuzunun ne kadar geniş olması gerektiğini söylemek zordur. Dengeli bir madde havuzu, kolaydan zora kadar tüm zorluk seviyeleri için maddeler içermelidir. Havuzdaki maddeler gizil özellik ölçüğünde farklı yetenek seviyelerinde bilgilendirici olmalıdır. Bilgisayarda bireyselleştirilmiş testlerin temel özelliği maddelerin testi alanların yetenek seviyelerine uyarlanması olduğundan havuzda farklı güçlük düzeylerinde birçok madde yer almalıdır. Böylelikle tüm yetenek seviyeleri aralığının doğru bir şekilde tahmin edilmesini sağlanır. Çünkü

kolay maddeler düşük yetenek seviyeleri için en bilgilendirici iken zor maddeler yüksek yetenek seviyeleri için en bilgilendiricidir (Magis, vd., 2017). Flaugher (2000), uyarlanabilir test için tatmin edici madde havuzunun, yüksek ayırt ediciliğe, ($a > 1$) dikdörtgen şeklinde bir güçlük dağılımına ve düşük şans parametresine ($c \leq .2$) sahip olması gerektiğini belirtmiştir. Urry (1977) ise madde ayırt edicilik değerlerinin .80'den büyük ($a > .80$), tahmin parametresinin .30'dan küçük ($c \leq .30$) olması gerektiğini ve madde güçlük parametresinin eşit ve geniş bir şekilde dağılması gerektiğini önermiştir.

Başlangıç Kuralı

Başlangıç adımı BBT'de uygulanacak ilk maddenin seçimini içerir. Bu adımın amacı, madde havuzundaki en az bir maddeyi seçmek ve onu test katılımcısına uygulamaktır. Testin başlangıcında kestirilen bir yetenek düzeyi (θ) olmadığından, maddeler yetenek tahminiyle seçilemediğinden, uyarlanabilir testin nasıl başlatılacağına ilişkin bazı kuralların belirlenmesi gerekir (De Ayala, 2009). Sıklıkla kullanılan başlama kuralı testi alan grubun ortalama yetenek düzeyine ($\theta = 0$) yönelik maddenin seçilmesidir (Magis, vd., 2017). Başlangıçta bireylerin yetenek düzeyi hakkında herhangi bir bilgi yoksa bu yöntemin tercih edilmesi uygun olacaktır. Başlangıçta orta güçlükte bir maddenin ($-0.5 \leq b \leq 0.5$) seçilmesi de kullanılabilir alternatif bir giriş kuralıdır. Bireyin yetenek düzeyiyle ilgili ön bilgiler mevcut ise uygulanacak başlangıç maddesinin güçlük düzeyi bu bilgiler kullanılarak belirlenebilir (Thissen & Mislevy, 2000). Tüm test katılımcılarına aynı maddeler uygulandığında maddeler tüm bireylere görüneceğinden test güvenliği düşecektir. Bu sebeple ilk maddenin madde havuzundan ortalama güçlükte maddelerden seçilmesi başlangıçta uygulanacak madde sayısını artırmış olacaktır.

Önerilen yaklaşımlar birçok yönden geliştirilebilir. İlk olarak, sınava giren kişinin yetenek düzeyi hakkında bazı ön bilgiler mevcutsa, ilk adıma dahil edilebilir. İkinci olarak, başlangıç maddesi farklı bir uygunluk kriteri ile seçilebilir. Urry (1970), başlangıç yetenek düzeyi için en bilgilendirici olan maddeyi belirlemek yerine, zorluk seviyesi bu ilk yetenek düzeyine en yakın olan bir maddenin seçilmesini önermiştir. Bu yöntem bOpt veya Urry kuralı

olarak adlandırılır. Zorluk seviyesinin çok kategorili MTK modellerinde doğrudan bir eşdeğeri olmadığından, bu yöntem iki kategorili MTK modelleri için uygulanabilir (Magis, vd., 2017). İlk maddenin uygulanmasının ardından yetenek kestirimi yapılarak test süreci sonlandırılana kadar bu döngü devam ettirilir.

Yetenek Kestirim Yöntemleri

Yeteneğin kestiriminde farklı kestirim yöntemleri mevcuttur. BBT'lerde kullanılan en yaygın yetenek tahmin yöntemleri Maksimum Olabilirlik Kestirimi (MOK), Maksimum Sonsal Kestirim (MSK), Beklenen Sonsal Dağılım (BSD) ve OWEN yöntemidir. Bunlardan BSD, MSK ve OWEN Bayes tabanlı tahmin yöntemleridir. Bayes tahmin yöntemlerinin ortak özelliği, yetenek tahminlerini türetmek için önsel bilgileri verilere dahil etmeleridir oysa MOK yalnızca verilere dayanır.

Maksimum Olabilirlik Kestirimi (MOK)

MOK, testi alan bireyin yeteneği için bazı önsel değerlerle başlar ve bir yetenek düzeyi için verilen yanıt modeliyle olabilirlik fonksiyonunu hesaplar. Olabilirlik fonksiyonu, bu yanıt vektörüne sahip olma olasılığını tahmin eder ve gözlemlenen modelle en olası sonuçlanan θ değerini bulur. Daha sonra MOK, sınava giren kişinin madde yanıt modelinin olasılığını en üst düzeye çıkaran noktayı bulur ve gizil özellik ölçeğinde karşılık gelen puana gider. Bu puan, sınava giren kişinin yetenek tahminidir. Bu yöntemin ilk avantajı matematiksel olarak kullanımının kolay olmasıdır. Bir diğer avantajı ise BBT uygulamalarında madde parametreleri önceden bilindiğinden, MOK ile yapılan tahminler madde parametrelerinin bilinmediği doğrusal testlere göre tarafsızdır (Wang & Vispoel, 1998). MOK yönteminin tutarlılık, verimlilik ve asimptotik normallik gibi çeşitli işlevsel özellikleri mevcuttur. Yeterli sayıda madde ile yetenek tahminlerinin standart hataları bilgi fonksiyonu kullanılarak elde edilebilir ve yetenek tahminleri için asimptotik güven aralıkları oluşturulabilir (Hambleton & Swaminathan, 1985). MOK birçok istatistiksel uygulamada parametre tahmini için yaygın olarak kullanılmaktadır. Madde tepki

kuramı temelinde yetenek kestirimi yapılırken, bilinen parametrelere sahip bir dizi maddeye bir yanıt vektörü tanımlandığında olasılık fonksiyonu aşağıdaki gibidir:

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

Burada $P_i(u_i | \theta)$, sınava giren kişinin gerçek yeteneği θ verildiğinde, i maddesinde u_i yanıtı alma olasılığıdır ve n , madde sayısıdır. Testi alan bireyin gerçek yeteneği θ 'nın maksimum olasılık tahmini, bu olabirlik fonksiyonunu veya eşdeğer olarak log olabirlik fonksiyonunu maksimuma çıkaran değer olan $\hat{\theta}$ 'dir. MOK ile ilgili büyük bir dezavantaj, tüm maddeleri doğru veya tüm maddeleri yanlış cevaplayan bireyler için yetenek kestirimi yapılamamasıdır. Testi alanlar tüm maddeleri doğru veya yanlış cevapladığında olasılık denkleminin çözümünde sorunlar ortaya çıkar çünkü bu tür yanıt kalıpları sonsuz yetenek tahminleri sağlar. Bu durum olabirlik fonksiyonunun sonsuza kadar artmasına neden olur ve olabirlik fonksiyonunda en yüksek değeri bulmak imkânsız hale gelir (Bock & Mislevy, 1982).

Bayes Tabanlı Yöntemler

Yetenek kestirimine yönelik Bayes tabanlı yaklaşımlar örneklemin yetenek dağılımının doğası hakkındaki varsayımların yetenek kestirimine dahil edilmesi açısından MOK'tan farklılık gösterir. Maddelerin doğru ve yanlış cevaplanma olasılığı hakkındaki bilgileri örneklemin yetenek dağılımının biçimi, ortalaması ve varyansı hakkındaki varsayımlarla birleştiren yetenek kestirimleri sağlar. Başlangıçta örnekleme ait yetenek dağılımının ortalaması 0 ve varyansı 1 olacak şekilde normal dağıldığı varsayılır ve önsel dağılım adını alır. Testi alan birey ilk maddeyi yanıtladıktan sonra bu yanıtla ilişkili olasılık sonsal dağılım adı verilen düzeltilmiş bir yetenek dağılımı oluşturmak için önceki yetenek dağılımı hakkındaki bilgilerle birleştirilir. Bu sonsal dağılım daha sonra bireyin ikinci maddeye verdiği yanıtla ilişkili olasılıkla birleştirilecek olan önsel dağılımı oluşturur (Wang & Vispoel, 1998).

Maksimum Sonsal Kestirim (MSK)

Maksimum Sonsal Kestirim, MOK'a benzer, ancak MSK bir önceki dağılımı belirtir ve ardından bu önceki dağılımı olasılık fonksiyonuyla çarpıp sonrasında aynı şeyi MOK ile yapar. Çoğu zaman önsel dağılım normal dağılımdan seçilir. MSK'nın en önemli avantajı, maddelerin tamamını doğru ya da tamamını yanlış cevaplayan bireyler için yetenek tahmini sağlaması ve MOK'dan daha iyi performans göstermesidir (Wang & Vispoel, 1998). MSK ile ilgili diğer bir avantaj, yetenek tahminini, ilk maddenin uygulanmasından sonra bile yapılabilmesi ve normallik varsayılmamasıdır.

Beklenen Sonsal Dağılım (BSD)

BSD ayrıca bir önsel dağılım belirtir ancak MSK'da bulunan olabilirlik fonksiyonu üzerindeki en yüksek noktadan farklı olarak, gizil özelliğin tahminini temsil eden sonsal dağılımın ortalamasını bulur. MOK ve MSK'dan farklı olarak iteratif bir süreç gerektirmez ve uygulanması oldukça kolaydır. Ayrıca, BSD ön dağılım varsaymaz ve MOK ve MSK'dan daha iyi performans gösterir. Bu durum BSD kestirim yönteminin diğer BBT yetenek kestirim yöntemlerine göre daha düşük standart hata ve yanlışlık ürettiği anlamına gelir. Uygun olmayan bir önsel dağılım belirtilirse, bu durum BSD kestirim sonuçlarının doğruluğunu etkileyebilir (Bock & Mislevy, 1982). Ayrıca, daha doğru sonuçlar elde etmek için BSD kestirim yöntemi BBT uygulamasında en az 20 madde gerektirir, aksi halde diğer yetenek kestirim yöntemlerinden daha yanlış sonuçlar üretir (Wainer & Thissen, 1987).

Kareleme yoluyla hesaplanan BSD kestirimleri, MSK veya MOK yöntemlerinden önemli ölçüde daha az işlem gerektirir. Bu hesaplamalarda kullanılan log olasılıkları, ardışık maddeler sunuldukça basit toplamlar halinde birikir. Ayrıca belirlenen kareleme noktalarındaki yanıt olasılıkları önceden değerlendirilerek ilgili maddeyle birlikte madde havuzunda saklanabilmektedir. Dolayısıyla, BSD tahminlerinin hesaplanması ve hata varyansı olarak hizmet eden sonsal dağılımın varyansı, sabit büyüklüklerin çarpımlarının basit toplamlarıdır. Yanıt fonksiyonu karmaşık olduğunda, çoklu kategori puanlamasında olduğu gibi, BSD tahmincisi, yetenek değişkenine göre birinci türevin hesaplanmasını gerektirmeme avantajına

sahiptir. Her farklı kategori için yanıt olasılıklarının her kareleme noktasında önceden hesaplanması ve madde havuzundaki ilgili maddeyle birlikte saklanması ihtiyacının dışında, hesaplamalar ikili puanlanan maddelerle aynıdır (Bock & Mislevy, 1982).

Madde Seçim Yöntemi

Başlangıç maddelerinin uygulamasının ardından test adımı; mevcut madde yanıtları seti kullanılarak yeteneğin tahmin edilmesi, belirlenen madde seçim yöntemine göre bir sonraki maddenin seçilmesi, bireyin yanıtının kaydedilerek yanıt modelinin güncellenmesi ve tekrar tahmininin yapılmasıyla devam eder. Maksimum olabilirlik ve Bayes kestirim yöntemleri BBT’de en çok kullanılan yetenek kestirim yöntemleridir. Test adımında, uygulanan belli sayıda maddeden sonra veya belli bir koşul sağlandığında bir yetenek kestirim yöntemi ile başlayıp daha sonra başka bir kestirim yöntemine geçilerek melez bir kural kullanılması da tercih edilebilir (Magis, vd., 2017). Madde seçim yöntemi BBT uygulamalarının önemli bileşenlerinden biridir. Test maddelerinin, BBT algoritmasının kapsam dengeleme bileşenindeki içerik özelliklerine dayalı olarak uygun olduğu belirlendikten sonra, bu maddeler uygulanacak bir sonraki madde seçimi için değerlendirilir. Madde seçim algoritması, madde havuzundan en iyi maddeyi seçmeye yönelik bir işlemdir. Madde seçimini etkileyen diğer koşulların getireceği sınırlılıklardan dolayı en uygun madde her zaman testi alan bireyin yeteneği hakkında en fazla bilgi veren madde olmayabilir. Örneğin, BBT uygulama sürecinde en iyi maddenin seçimi, yetenek hakkında en fazla bilgiyi sağlayan madde yerine belirli bir içerik alanındaki bilgiyi ölçmek için gerekli olan maddenin seçilmesiyle mümkün olabilir. Test ölçüm alanında kapsamlı bir dizi madde seçim yöntemi geliştirilmiştir ancak gerçek BBT uygulamasında bu yöntemlerden çok azı kullanılmaktadır (Han, 2018). Maksimum Fisher bilgisi, Urry kuralı olarak adlandırılan bOpt yöntemi, olabilirlikle ağırlıklandırılmış bilgi yöntemi, sonsal ağırlıklı bilgi yöntemi, Kullback-Leibler yöntemi, Sonsal Kullback-Leibler yöntemi, beklenen maksimum bilgi yöntemi, minimum beklenen sonsal varyans yöntemi, rastgele seçim yöntemi, thOpt yöntemi gibi birçok farklı madde seçim kuralı mevcuttur (Magis, vd., 2017). Sıklıkla kullanılan madde seçim yöntemleri aşağıda tanıtılmıştır.

Maksimum Fisher Bilgisi (MFB)

Testler, bireyin yetenekleri hakkında bilgi sağlayan ölçme araçlarıdır. Bireyler uygun şekilde seçilmiş sorulara yanıt verdikçe daha fazla bilgi birikir ve yetenek tahmininin ölçme kesinliği artar. Her yanıtta ne kadar fazla bilgi sağlanırsa ölçme kesinliği o kadar hızlı artar. Belirli bir yetenek tahmini için bir yanıtta bilgiyi en üst düzeye çıkarmak, BBT uygulamalarında madde seçiminde kullanılan en yaygın ve en eski yöntemlerden biridir. Bu yöntem, daha önce bir test katılımcısına uygulanan test maddelerine dayalı olarak belirli bir θ 'da maksimum Fisher bilgisine (MFB) sahip bir maddenin seçilmesini içerir (Weiss & Kingsbury, 1984). i maddesi için Fisher bilgisi şu şekilde hesaplanabilir (Hambleton & Swaminathan, 1985; van der Linden, 2010):

$$I(\theta) = \sum_{i=1}^n I_i(\theta) = \sum_{i=1}^n \frac{((P'(\theta))^2)}{P(\theta)[1 - P(\theta)]}$$

Burada $P(\theta)$, belirli bir θ düzeyine sahip bireyin maddeye doğru yanıt verme olasılığıdır, ve $P'(\theta)$, $P(\theta)$ 'nin birinci türevidir. Bu kural, uygulanan n maddeye dayanan belirli bir geçici yetenek tahmini (θ_n) için, seçilen maddenin maksimum $I(\theta_n)$ değerini veren madde olmasını şart koşar. BBT uygulamasını oluşturan maddeler MFB kuralı kullanılarak seçildiğinde, test bilgisi ve yetenek tahmininin kesinliği önemli ölçüde artacaktır (Hambleton, vd., 1991). Ancak bu durum yalnızca yetenek tahmini gerçek yetenek düzeyine yakın olduğunda geçerlidir. BBT'nin ilk aşamalarında geçici yetenek tahmini, gerçek yetenek düzeyine yakın olmayabilir (Han, 2012). MFB kuralı kullanılarak seçilen maddeler gerçek yetenek düzeyinde $I(\theta)$ 'yı en üst düzeye çıkaran maddeler olmayabilir. Bu durum verimlilik ve ölçme kesinliğinde azalmaya neden olacaktır (Chen & Ankenmann, 2004). Madde seçiminde, kapsam dengeleme gibi istatistiksel olmayan değerlendirmelerin yanı sıra, son otuz yılda en yaygın strateji madde bilgilerinin maksimizasyonu olmuştur. Özellikle, bir madde testi alan kişinin mevcut yanıtlarından hesaplanan, o anda kestirilen θ düzeyinde maksimum bilgiye sahipse tercih edilebilir. MFB yöntemi, madde seçimindeki 'hırslı' eğilimi nedeniyle, gerçek operasyonel BBT

uygulamalarında nadiren kullanılır (Han, 2018). MFB yönteminde, ayırt ediciliği yüksek olan maddelerin seçilmesi olasıdır. Bunun bir sonucu olarak BBT uygulamalarında madde havuzu yanlı kullanılabilen ve madde havuzunun kullanımında problemler ortaya çıkmaktadır (Chang, 2004; Han, 2012).

Kullback-Leibler (KL) Bilgisi

Bir diğer madde seçme yöntemi ise Kullback-Leibler bilgisidir. Bu madde seçim yönteminde, gizil özellik puanı üzerindeki iki yetenek düzeyi arasındaki mesafe olan Kullback-Leibler (KL) bilgisi, ilk olarak testi alan bireye hâlihazırda uygulanan madde sayısı üzerinden tahmin edilir. Daha sonra KL bilgisini en aza indiren, başka bir deyişle mevcut gizil puanla en iyi eşleşen bir sonraki madde seçilir. Herhangi bir olası yetenek seviyesi çifti arasındaki madde ayırıcılık düzeyini değerlendiren küresel bir bilgi ölçüsüdür (Magis, vd., 2017). Global bilgi yaklaşımına dayanan Kullback Leibler bilgisi şu şekilde tanımlanır (Chang & Ying, 1996):

$$K_i(\theta \parallel \theta_0) = P_i(\theta_0) \log \left[\frac{P_i(\theta_0)}{P_i(\theta)} \right] + [1 - P_i(\theta_0)] \log \left[\frac{1 - P_i(\theta_0)}{1 - P_i(\theta)} \right]$$

KL bilgisi θ ve θ_0 değişkenlerinin bir fonksiyonu olmakla birlikte üç boyutlu uzayda bir yüzeye karşılık gelmektedir. KL bilgisi tanımlanan θ ve θ_0 yetenek düzeyleri arasında bir maddenin değişme kapasitesini ortaya koyar. Genel olarak, Kullback-Leibler bilgisi, iki olasılık arasındaki mesafenin bir ölçüsüdür. Bu bilgi ne kadar büyük olursa, iki olasılığı ayırt etmek o kadar kolay olmaktadır (Chang, 2004). Kullback-Leibler bilgisi, testin başında oluşabilecek madde parametresi kestirim hatası problemiyle başa çıkmak için MTK'da madde seçimi için bir kriter olarak kullanılmıştır. Test uzunluğunun kısa olduğu veya θ 'nın kestirim hatasının yüksek olduğu test uygulamasının erken aşamasında, MFB yöntemi yerine KL bilgisi yönteminin kullanılması yetenek tahminindeki yanlılığı ve ortalama karesel hataları azaltabilecektir (Chang & Ying, 1996).

Olabilirlikle Ağırlıklandırılmış Bilgi (OAB)

Bu yöntem, madde seçiminde olabilirlik fonksiyonu ile ağırlıklandırılmış maksimum bilgiyi kullanmaktadır. OAB yönteminde, bilgi fonksiyonunun kendisini maksimize etmek yerine o anda uygulanan yanıt modelinin olabilirlik fonksiyonu ile ağırlıklandırılması amaçlanır (Magis, vd., 2017). OAB yönteminde bilgi fonksiyonu yetenek (θ) ölçeği boyunca toplanır ve o ana kadar uygulanan maddelerden sonra olasılık fonksiyonu ile ağırlıklandırılır (Han, 2018). En yüksek bilgiyi sağlayan madde belirlenirken MFB ve olabilirlik fonksiyonunun birleşimi dikkate alınır. Bu yöntemde seçilecek i maddesi aşağıda verilen eşitliğin değeri maksimum yapacak şekilde belirlenir:

$$\int_{\theta=-\infty}^{\infty} L(\theta; x_{m-1}) I_i[\theta] d\theta$$

Verilen eşitlikte $L(\theta; x_{m-1})$; $(m-1)$ 'inci maddenin uygulanmasının ardından yanıtlayıcı vektörü x_{m-1} 'in olabilirlik fonksiyonuna karşılık gelmektedir. Burada $L(\theta; x_{m-1})$, yanıt vektörü x_{m-1} 'den sonraki olasılık fonksiyonudur. Bu seçim, θ ölçeğine karşılık gelen θ değerinin bireyin gerçek yetenek değeri olma olasılığı daha yüksek olduğunda bilgi fonksiyonu değeri $I(\theta)$ 'ya daha fazla ağırlık vermek anlamına gelir. Yeteneğin maksimum olabilirlik tahmini kullanılıyorsa, en fazla mevcut yetenek tahminine yakın değerler ağırlıklandırılır (Veerkamp & Berger, 1997). Testin başlangıcında, olabilirlik fonksiyonu düzdür ve testin sonuna doğru olabilirlik fonksiyonu daha fazla tepe noktasına gelme eğilimindedir (van der Linden & Glas, 2000).

Kapsam Dengeleme

Test gereksinimlerine göre, BBT uygulamalarında dengeli bir kapsam temsiline sahip olmak için madde seçiminin belirlenen kapsamın gereksinimlerini karşılaması gerekir. BBT uygulaması, her konu alanından gelen maddelerin önceden tanımlanmış yüzdelerine göre dengelenmelidir. Uyarlanabilir testlerde, testi alan bireylerin test puanlarının karşılaştırılabilir olması için içerik alanına göre aynı madde dağılımını almaları gerekir. Testi alan bireyler arasında, testlerde yeterli alan temsiline sağlamak için uygun bir belirtke tablosuna dayalı

kapsam dengelemesi gereklidir. Bir test ile ölçtüğü içerik alanı arasındaki uyum, test puanı çıkarımlarının geçerliği için temel kanıtı temsil eder. Doğrusal testlerin içerik temsilini değerlendirmek için farklı yöntemler geliştirilmiş olsa da bu yöntemler büyük madde havuzları gerektiren ve sabit test formları kullanmayan BBT'lere kolayca uygulanamaz (Wise vd., 2015). Bireye uyarlanmış testlerde, testi alanlara farklı maddeler uygulanır, ancak her birine içerik alanına göre aynı madde dağılımı verilmelidir. Örneğin, 20 maddelik bir matematik testinde katılımcılardan birine yalnızca aritmetik, diğerine yalnızca geometri maddelerinin uygulanması ölçme sonuçlarının geçerliğini olumsuz etkiler. Geçerli ölçümler elde edebilmek için ölçülen içerik alanları arasında bir denge sağlanmalıdır. Test verimliliğini korurken kapsam dengelemeye ulaşmak için çeşitli yöntemler geliştirilmiştir. Bu yöntemlerden en sık kullanılan Kingsbury ve Zara (1989) tarafından önerilen *Kısıtlı Bilgisayarda Bireyselleştirilmiş Test* (KBBT) yöntemidir. Bu yöntem, iki aşamalı kapsam dengeleme kontrolü sağlayan bir algoritma kullanır. Hedeflenen uygulama yüzdesinin altında kalan içerik alanlarından madde seçilerek her bir içerik alanının dengeli bir şekilde temsil edilmesi sağlanır. Bu yöntemde, BBT sürecinin her adımında her bir kategori için deneysel yüzdeler hesaplanır. Daha sonra teorik oran ile deneysel değeri arasında en fazla fark olan kategori belirlenir ve bu kategoriden madde seçilir. Madde havuzundaki her bir içerik alanında yeterince madde bulunduğu sürece, istenen herhangi bir içerik dağılımı karşılanabilir. Ancak KBBT'nin bir sınırlılığı, testin erken aşamalarında öngörülebilir bir içerik dizisi oluşturabilmesidir. Bu durum, katılımcıların hangi içerik alanından madde geleceğini tahmin ederek strateji geliştirmelerine yol açabilir. Ayrıca, test uzunluğu değiştiğinde veya her bir içerik alanındaki madde sayısına önceden karar verilemediğinde KBBT yönteminin etkinliği azalabilir. Bu sorunlar madde takımları kullanılarak aşılabilir. Chen, Ankenmann ve Spray (1999), içerik alanlarının sırası tahmin edilebilir olduğundan, KBBT'nin istenmeyen sıra etkileri oluşturabileceğini savunarak bu yöntem yerine dengeli içerik gereksinimini karşılamak için *Modifiye Edilmiş Çok Terimli Model* (MÇM) geliştirmişlerdir. Bu yöntemde toplamları 1.0 olan içerik alanlarının hedef yüzdelere dayalı olarak önce kümülatif bir dağılım oluşturulur. Daha sonra, bir sonraki en uygun maddenin seçileceği kümülatif dağılımdaki karşılık gelen içerik alanını belirlemek için tek biçimli

dağılımdan $U(0,1)$ rastgele bir sayı kullanılır. Bir içerik alanı hedef yüzdesine ulaştığında, kalan içerik alanlarının karşılanmayan yüzdeleri ayarlanarak yeni birçok terimli dağılım oluşturulur. Bu yöntem seçkisiz uygulandığından, içerik alanlarının sırası değişir. Leung ve arkadaşları (2000) ise *Modifiye Edilmiş Kısıtlı Bilgisayarda Bireyselleştirilmiş Test* (MKBBT) yöntemini geliştirmiştir. Bu yöntemde hedef yüzdesinin en altında mevcut madde kullanım sıklığı oranına sahip içerik alanıyla sınırlandırılmak yerine, hedef yüzdesi tamamlanmayan tüm içerik alanlarından en uygun madde seçilebilir. Böylelikle KBBT'nin istenmeyen sıra etkisi önlenmiş olmaktadır.

Bu üç yöntemin yanı sıra *Ağırlıklı Ortalama Dengeleme* (van der Linden, 2005) ve *Uyarlanabilir Kapsam Dengeleme* (Luecht, 2003) gibi yaklaşımlar da kapsam dengeleme bağlamında etkili yöntemler olarak öne çıkmaktadır. Ağırlıklı ortalama dengeleme, içerik alanlarına tanımlanmış ağırlıklar atanarak, madde seçiminin bu ağırlıklara uygun yapılmasını sağlar. Bu yöntem, büyük madde havuzlarında ve uzun testlerde kapsam dengeleme hedeflerine ulaşmak için uygun bir seçenek sunar. Uyarlanabilir kapsam dengeleme ise, bireylerin yanıtlarına bağlı olarak içerik dağılımını dinamik şekilde düzenleyerek kapsam dengeleme doğruluğunu artırır. Bu yöntemlerin tümü, bireyselleştirilmiş testlerin geçerliğini ve güvenilirliğini artırmanın yanı sıra, test adillliğini sağlamak için önem taşımaktadır. Doğru seçilen bir kapsam dengeleme yöntemi, testin hedeflerine ulaşmasını sağlarken, içerik alanlarının dengeli bir şekilde temsil edilmesine katkıda bulunur.

Madde Kullanım Sıklığı Kontrolü

Sayıları artan büyük ölçekli BBT uygulamaları ile birlikte teori pratiğe çevrilirken ölçme uzmanları birçok sorunla karşılaşmaktadır. Bu sorunlardan biri bazı maddelerin kısa bir süre içinde aynı bölgede tekrar tekrar bireylere sunulmasıdır. Mevcut madde seçim prosedürlerine göre, genellikle belirli bir geçici yetenek tahmini için en yüksek bilgi değerlerini veren maddeler seçilir. Böylelikle yüksek ölçme hassasiyeti ve uygulama verimliliği sağlanmış olur. Bununla birlikte bir madde çok sık sunulursa, sınava girenlerin büyük bir kısmı bu maddeyi görecektir.

Sınav katılımcılarından bazıları sınav uygulamalarından önce sorulara eriştiğinde bu durum sorun yaratacaktır (Chang, 2004). Yüksek oranda madde teşhiri büyük bir test güvenlik sorununa yol açar. Adayların yakın zamanda aynı madde havuzunu kullanan BBT uygulamasını almış diğer adaylarla iletişim kurmaları olağandır. BBT'nin geleneksel kâğıt kalem testine rakip olabilmesi, geleneksel testlerde olduğu gibi tüm katılımcılar için adaleti sağlamak adına maddelerin teşhirini sınırlayacak yöntemlerin geliştirilmesi ile mümkündür (Chang & Ansley, 2003). Madde Kullanım sıklığı oranı, bir test maddesinin farklı test uygulamalarında katılımcılara sunulma sıklığını ifade eder. Bir maddeyi gören katılımcıların yüzdesini, mutlak sayısını içerir. Yüzde 20'lik bir madde kullanım sıklığı oranı, bir maddenin beş katılımcıdan birine sunulacağı anlamına gelir.

Bazı maddelerin çok sık kullanılmasının nedeni bu maddelerin ortalama yetenek düzeyinde çok bilgilendirici olması, dolayısıyla diğer maddelerden daha sık seçilmesidir. Madde kullanım sıklığı oranının çok yüksek olduğu durumda test güvenliği test için ciddi bir endişe haline gelir. Bir madde çok sık uygulandığında bu madde hakkında ön bilgi edinilebilir ve bu durum bazı bireyler için adaletsizlik yaratabilir. Bu maddeler yanıltıcı bilgi verebilir ve yetenek tahmini için sorunlara neden olabilir. Madde kullanım sıklığı oldukça fazla olan maddeleri ortadan kaldırmak madde havuzu için geliştirme ve kalibre etme konusundaki birçok süreçten dolayı maliyetli olabilir (Magis, vd., 2017). Bu nedenle maddelerin çok sık uygulanmasının önüne geçmek önemlidir. Madde kullanım sıklığı oranını kontrol etmek, bir maddeyi hedeflenen maksimum istenen görülme oranıyla uygulama sıklığını sınırlamaktır. Madde kullanım sıklığının kontrolünde farklı yöntemler önerilmiştir. Bunlardan bazıları aşağıda tanıtılmıştır.

Sympson-Hetter Yöntemi (SH)

Bu yöntemde, bir maddenin testi alanlardan rastgele örneklenmiş bir katılımcı için en bilgilendirici madde olarak seçilme olasılığı $P(S)$ ile bu maddenin seçildiğinde uygulanma olasılığı $P(A/S)$ arasında ayırım yapar. Yöntem, $P(A)=P(A/S)* P(S)$ olmak üzere, bir maddenin $P(A)$ uygulanma olasılığının genelini kontrol etmeye ve havuzdaki tüm maddeler için

maksimum $P(A)$ deęerinin, istenen maksimum madde kullanım oranı olan r deęerinden küçük olmasını saęlamaya alıřmaktadır. Madde kullanım sıklığı kontrol parametreleri $P(A/S)=K$, yetenek daęılımından alınan bir rnekle, test tasarımının bir dizi yinelemeli simlasyonu zerinden belirlenir (Georgiadou, vd., 2007). ok uygulanan maddeler iin, kullanım sıklığı kontrol parametreleri nceden belirlenmiř istenen kullanım sıklığı oranı kadar dřk olabilir, bu da bu maddelerin seildięinde serbeste uygulanamayacaęını gstermektedir. Nadiren uygulanan maddeler iin, iliřkili kullanım sıklığı kontrol parametreleri 1'e yakın olabilir, yani bu maddeler bir kez seildięinde, neredeyse her zaman sunulurlar (Chang & Ansley, 2003). Bu yntem madde kullanımını olasılıksal bir tarzda kontrol etmek iin tasarlanmıřtır. oęu madde nceden belirlenmiř bir maksimum madde kullanım sıklığı oranından daha az sıklıkta uygulanır. Sympson-Hetter yntemi, rastgele madde seiminden daha iyi performans gsteriyor gibi grnse de kararlı madde kullanım parametrelerini bulmak iin gereken yinelemeli simlasyonlar zaman alıcıdır. Ayrıca yinelemeli simlasyonların, madde havuzu, madde seim kuralları gibi, BBT ayarlarında deęiřiklik olduęunda tekrarlanması gerekir (Chan & Ankenmann, 2004).

Azalarak Kaybolma (Fade-away) Yntemi

Bilgisayar aęı teknolojisi, test merkezlerindeki ana bilgisayar sunucularının ve istemci bilgisayarların, madde kullanımını dahil olmak zere eřitli test bilgilerini yeniden yapılandırmak iin BBT uygulaması ncesinde, sırasında ve/veya sonrasında iletiřim kurmasını mmkn kılmaktadır. Ana sunucuda tutulan eksiksiz madde kullanım bilgileri, evrimii aę aracılıęıyla her BBT uygulaması sırasında veya sonrasında bilgisayarlar tarafından dzenli olarak gncellenebilir. Ayrıca, her bilgisayar, bir sonraki test uygulamasının bařlamasından hemen nce sunucudan gncellenen madde kullanım bilgilerine eriřebilir. Bu tr aę teknolojisi, tekrarlamalı simlasyonları ieren SH yntemini kullanmak gibi bařka yollarla madde kullanım sıklığını tahmin etme ihtiyacını ortadan kaldırarak, BBT sisteminin kullanım sıklığı kontrol iin neredeyse gerek zamanlı madde kullanım sıklığı bilgilerini kullanmasını saęlar. Azalarak kaybolma ynteminde, madde havuzundaki her uygun madde iin madde seim lt,

güncellenmiş gerçek kullanım sıklığı oranı ile hedeflenen kullanım sıklığı oranı arasındaki oran ile ters orantılı olarak ağırlıklandırılır. Bu yöntemle, nadiren kullanılan maddelerin daha sık uygulanması beklenirken, aşırı kullanılan maddelerin de madde seçiminden “kaybolması” olasıdır (Han, 2018).

Rastgele Seçim Yöntemi

Kingsbury ve Zara (1989), madde kullanım sıklığını kontrol etmek için rastgele bir yöntem önermiştir. Rastgele seçim yöntemi, en bilgilendirici tek bir maddeyi seçmek yerine, en bilgilendirici maddelerden oluşan bir gruptan uygulanacak bir sonraki maddeyi tesadüfi olarak seçmektedir. Bununla birlikte rastgele yöntem, ilk birkaç maddeden sonra en bilgilendirici madde seçimine dönmek yerine, test uygulaması boyunca tesadüfi bir bileşen kullanmaya devam eder. Bir maddenin rastgele uygulandığı grubun büyüklüğü, her bir test programı için özelleştirilebilir. Bu yöntemin amacı, yalnızca testin başındaki madde çakışmasını azaltmak değil, aynı zamanda testin ilerleyen aşamalarında benzer yetenek düzeylerine sahip katılımcılara maddelerin kullanım sıklığını kontrol etmektir (Davis, 2002). Rastgele madde seçim yönteminde seçilen ancak uygulanmayan maddeler sonraki kullanım için madde havuzuna iade edildiğinden çoğu bilgilendirici maddenin kullanım sıklığı oranları kontrol edilemeyebilir. Bu nedenle madde kullanım sıklığı, rastgele madde seçimi yaklaşımı altında yeterince kontrol edilmez (Chan & Ankenmann, 2004).

Sonlandırma Kuralı

Test geliştiricileri, durdurma kuralı seçiminin genellikle test amacına, madde havuzu özelliklerine ve operasyonel kısıtlamalara büyük ölçüde bağlı olduğunu bulmuşlardır (Segall, 2004). Bir BBT uygulamasının ne zaman durdurulacağına dair karar en önemli unsurdur. Test çok kısaysa, yetenek tahmini hatalı olabilir. Test çok uzunsa, zaman ve kaynaklar boşa harcanır ve madde gereksiz yere uygulanmış olur. Test katılımcısı yorulabilir ve performans seviyesi düşebilir, bu durum geçersiz test sonuçlarına yol açabilir. BBT uygulaması şu durumlarda durur (Tian, vd., 2007):

- Madde havuzu tükendiğinde sonlanır. Bu durum genellikle küçük madde havuzlarında, testi alan bireye tüm maddeler uygulandığında meydana gelir.
- Maksimum test uzunluğuna ulaşıldığında sonlanır. Testi alan kişiye uygulanmasına izin verilen önceden belirlenmiş bir maksimum madde sayısı vardır.
- Yetenek ölçüsü yeterli hassasiyetle kestirildiğinde test sonlanır. Her yanıt, standart ölçme hatasını azaltarak, ölçme kesinliğini artırarak, yetenek ölçümü hakkında daha fazla istatistiksel bilgi sağlar. Ölçüm belirlenen yeterli kesinliğe ulaştığında test durur.
- Yetenek ölçüsü geçti-kaldı kriterinden yeterince uzak olduğunda test sonlanır. Test katılımcılarını geçti-kaldı kriter düzeyine göre değerlendiren BBT uygulamaları için, geçti-kaldı kararı istatistiksel olarak kesin olduğunda test durabilir. Bu durum, yetenek tahmininin ölçüt seviyesinden en az iki standart hata uzakta olduğu veya adayın geçerli geçti-kaldı kararını değiştirmesi için testte yeterli madde kalmadığında meydana gelebilir.
- Test katılımcısı test dışı davranış sergilediğinde test sonlandırılır. BBT programı, çok hızlı yanıt veren ve çok yavaş yanıt veren yanıt kümelerini algılayabilir. Test katılımcısına, testi durdurma veya erteleme konusunda nihai karar vermesi için test gözetmenini araması talimatı verilebilir.

Sonlandırma kuralı, maddelerin uyarlanabilir uygulamasını durdurmak için parametreler tanımlamaktadır. Genel olarak (a) uzunluk kriteri (b) kesinlik kriteri (c) sınıflandırma kriteri ve (d) bilgi kriteri olarak adlandırılan dört ana durdurma kuralı tanımlanmaktadır (van der Linden, 2010; Wainer, 2000). Uzunluk kriterinde, uygulanacak toplam madde sayısı belirlenir ve belirlenen sayıda madde uygulandığında BBT süreci durur (Babcock & Weiss, 2009; Stafford vd., 2019). Nispeten daha uzun testler, yetenek tahminlerinin kesinliğini artırır, ancak daha kısa testler, BBT'nin erken aşamalarında belirli konuları araştırmak için de faydalı olabilir. Sabit BBT uzunluğunun kullanıldığında, tüm test katılımcıları aynı sayıda madde alır ancak bu durumda yetenek kestirimleri için farklı kesinlik değerleri elde edilmiş olur. Kesinlik kriteri, yetenek seviyesi önceden tanımlanmış bir kesinlik

seviyesine ulařtıđında BBT uygulamasını durdurur. Yani BBT uygulaması, geici yetenek dzeyi nceden belirlenmiř eřik deđerine eřit ya da daha kk standart hataya sahip olduđunda durur. Daha uzun testler daha fazla kesinlik (ve dolayısıyla daha kk standart hatalar) sađladıđından, bu kriter genellikle yeterince uzun testler uygulandıđında karřılanır. Tm testi alanlar iin minimum kesinlik sađlamak istendiđinde nerilir ancak madde havuzunun kestirilen yetenek seviyeleri iin ne kadar bilgilendirici olduđuna bađlı olarak, katılımcılara farklı uzunluklarda BBT uygulamaları yapılması anlamına gelecektir (Stafford vd., 2019). Kesinlik ile sonlandırma kriterinin, belirlenen standart hata deđerı yeterince dřkse hem uygulanan az sayıda madde hem de yksek yetenek tahmini dođruluđu aısından olduka iyi performans gsterdiđi bulunmuřtur (Babcock & Weiss, 2009).

Beceride yeterlilik test edilirken sınıflandırma kriteri kullanılır. Temel ama, katılımcının beceri yeterliliđini gsteren yetenek dzeyinden daha yksek veya daha dřk bir yetenek dzeyine sahip olup olmadıđını belirlemektir. Uygulamada, bu yeterlilik dzeyi bazı yetenek eřiklerinde sabitlenir ve kesinlik kriteri, mevcut yetenek kestiriminin geici gven aralıđının bu eřik ile karřılařtırılmasından oluşur. Bu gven aralıđı eřik ile rtřyorsa, sınava giren kiřinin nihai sınıflandırması (uzmanlařma veya uzmanlařmama) hakkında yeterli kesinlik yoktur. te yandan, gven aralıđı sınıflandırma eřiđini kapsamıyorsa (eřik alt gven sınırından dřkse) sınav katılımcısı beceriye sahip veya deđil (eřik st gven sınırından yksekse) olarak sınıflandırılabilir. Bilgi kriteri ise geici yetenek tahmininde her bir maddenin sahip olduđu bilgi miktarına odaklanır (Magis, vd., 2017). BBT uygulamasının devam etmesi iin gerekli kořul, kalan uygun maddelerin toplam bilgiyi nemli lde artırmak iin yeterli bilgiyi paylařmasıdır. Bu kural iin eřik, uygun maddelerden en az biri tarafından tařınan minimum bilgidir. Geici yetenek kestiriminde, tm uygun maddelerin nceden tanımlanmıř eřikten daha kk bilgi deđerleri varsa BBT sonlandırılır. Bu kriter, yetenek tahmini ve ilgili kesinlik iin yeterli bilgi tařımayan maddelerin uygulanmasından kaınmak iin kullanılabilir. BBT programlarının ođu tek bir durdurma kuralına odaklansa da birka durdurma kuralını aynı anda dřnmek ve

kurallardan en az biri yerine getirilir getirilmez BBT sürecini durmaya zorlamak mümkündür (Tian, vd., 2007).

İlgili Araştırmalar

Leung, vd. (2003a) yapmış oldukları çalışmada, çeşitli test uzunluğu (16,28 ve 40 madde) ve belirlenen maksimum madde kullanım sıklığı oranı (.1 ve .2) koşulları altında, kısıtlı BBT (KBBT), modifiye edilmiş kısıtlı BBT (MKBBT) ve modifiye edilmiş çok terimli model (MÇM) olmak üzere üç kapsam dengeleme yöntemini karşılaştırmıştır. Yapılan simülasyon çalışmasının sonuçları, kapsam dengeleme yönteminin ölçme verimliliği ve havuz kullanımında sistematik bir etkisinin olmadığını göstermiştir. Bu üç yöntem arasında MÇM, madde seçim yöntemi, test uzunluğu veya hedeflenen maksimum madde kullanım sıklığı oranından bağımsız olarak, madde kullanım sıklığı açısından daha kullanışlı bulunmuştur. Çalışma yalnızca bir madde havuzunu içerdiğinden, MÇM'nin diğer iki yöntemle göre avantajlarının farklı madde havuzları kullanılarak ve ek test koşulları altında çapraz incelenmesi gerektiği önerilmiştir.

Leung, vd. (2003b) yapmış oldukları başka bir çalışmada BBT uygulamalarında tabakalaşma tasarımları ile kapsam dengeleme gereksinimlerinin birleştirilmesinin sonuçlarını incelemişlerdir. KBBT, MÇM ve MKBBT olmak üzere üç içerik dengeleme yöntemi ile a-tabakalı (ASTR), b-bloke etme yöntemiyle a-tabakalı (BASTR) ve BASTR yönteminin çoklu tabakalaşmaya genişletilmiş hali olan (CBASTR) tabakalama yöntemlerinden oluşan dokuz koşulun her birinin performansı madde güvenliği, ölçme verimliliği ve havuz kullanımı açısından değerlendirilmiştir. Üç kapsam dengeleme yöntemi arasında KBBT, tüm madde havuzunu kullanma ve madde çakışmasını kontrol etmede en az etkili yöntem olarak bulunmuştur. Bu durumun her seçim noktasında daha küçük bir aday maddeler kümesinden uygun maddenin seçimini tekrar tekrar kısıtlayan içerik dizisinin öngörülebilirliğinden kaynaklandığı belirtilmiştir. Buna karşılık, MÇM'deki randomizasyon mekanizması tüm havuzun daha dengeli bir şekilde kullanılmasını sağlamıştır. CBASTR ve MÇM

kombinasyonunun, aynı anda madde örtüşmesini kontrol eden, tüm havuzu kullanan ve pratik kapsam dengeleme gereksinimini karşılayan BBT için en uygun madde seçim yöntemi olduğu önerilmiştir.

Yi ve Chang (2003) çalışmalarında, kapsam kısıtlamalı, modifiye edilmiş a-katmanlı yöntemi önermiştir. Tabakalı BBT tasarımlarının iyileştireceği önerilen yeni yöntem içerik özelliklerini madde havuzu katmanlaştırmasına dahil etmektedir. Yeni yöntemin mevcut olan üç madde seçim yöntemiyle (a tabakalı yöntem; a-tabakalı b-blokaj yöntemiyle ve Sympson-Hetter madde kullanım sıklığı kontrolü ile maksimum Fisher bilgi yöntemi) karşılaştırılması amacıyla simülasyon çalışmaları yapılmıştır. Dört yöntemin tümünün kapsamı dengelemeye zorlandığı bir durumda, a-tabakalı tasarımın, madde kullanım sıklığı oranlarını azaltmada, havuz içindeki madde kullanımını dengelemede ve ölçme kesinliğini korumada iyi performans gösterdiği bulunmuştur.

Cheng ve Chang (2007) çalışmalarında modifiye edilmiş çok terimli model (MÇM)'nin bir uzantısı olan MÇM+; modifiye edilmiş kısıtlı BBT (MKBBT)'in bir uzantısı olan MKBBT+; her iki aşamada da MÇM'yi kullanan iki aşamalı bir kapsam dengeleme yöntemi olan TPM ve ilk aşamada MÇM'yi ve ikinci aşamada MKBBT'yi kullanan iki aşamalı bir kapsam dengeleme yöntemi olan TPF yöntemini karşılaştırmıştır. Elde ettikleri simülasyon sonuçları, tüm yöntemlerinin kapsam dengelemede iyi çalıştığını ve TPF'nin ölçme duyarlılığını korurken madde kullanım sıklığı kontrolü ve madde havuzu kullanımında en iyi performansa sahip yöntem olduğunu göstermiştir.

Song (2010) içerik özelliklerinin kontrolü ile tek boyutlu bir MTK modelinin bilgisayarda bireyselleştirilmiş testteki çok boyutlu verilere uygulanmasının etkisini araştırmıştır. Maksimum bilgi maddesi seçim yöntemine sahip kısıtsız BBT temel olarak kabul edilmiş ve üç kapsam dengeleme yönteminden; kısıtlı BBT, modifiye edilmiş çok boyutlu model ve modifiye edilmiş kısıtlı BBT performansları değerlendirilmiştir. Çalışmada test verileri çok boyutlu olduğunda, kapsam dengeleme yöntemlerinin maksimum bilgi yöntemine benzer hatta daha yüksek doğrulukla sonuçlandığı görülmüştür. Bunun nedeni, içerik alanlarındaki madde yüzdelerinin

kontrol edilmesinin verilerin her boyutunun yeterli temsilini sağlaması ve dolayısıyla ölçme kesinliğini geliştirmesi olarak açıklanmıştır. İçerik alanlarının zorluk seviyesinin dört yöntemin performansını etkileyen önemli bir faktör olduğu öne sürülmüştür. İçerik alanları aynı zorluk seviyesine sahip olduğunda kapsam dengeleme yöntemleri, maksimum bilgi yöntemiyle karşılaştırılabilir ölçme kesinliği sağlamıştır. Çalışmada kapsam dengeleme yöntemleri, daha verimli madde havuzu kullanımı ve daha düşük madde kullanım sıklığı yüzdeleri sağlamıştır.

Shin, vd. (2012) sabit ve değişken uzunlukta BBT'ler için üç kapsam dengeleme yöntemini karşılaştırmıştır. Çalışmada hem sabit hem de değişken uzunluklu BBT'ler için basit ve karmaşık içerik kısıtlama yapısı, farklı madde havuzları, madde kullanım sıklığı kontrol özelliği ve yetenek kestirim seçeneklerini içeren çeşitli koşullar altında ağırlıklı penalty modeli ile ağırlıklı sapma yöntemini karşılaştırmak amaçlanmıştır. Basit içerik kısıtlama yapısına sahip testler için karşılaştırmaya kısıtlı BBT (KBBT) yöntemi de dahil edilmiştir. Ağırlıklı penalty modeli, kapsam dengeleme ve ölçme kesinliğinde biraz daha iyi performans göstermiştir fakat bunun yanında yüksek madde kullanım sıklığı ve hiç kullanılmamış madde oranına yol açmıştır. Ölçme kesinliği ve madde kullanım sıklığı kontrol sonuçları, her üç yöntemde de benzer bulunmuştur.

Zheng, vd. (2013) yapmış oldukları çalışmada, iki faktörlü modellenmiş bir BBT madde seçiminde kapsam dengeleme stratejisi kullanılmış ve bunun ölçme doğruluğu ve madde havuzu kullanımı üzerindeki etkisi incelenmiştir. Kapsam dengeleme stratejisi, iki faktörlü modellenmiş BBT'nin madde seçim prosedüründe benimsenmiştir. BBT maddelerinin ölçme doğruluğu ve kullanımı, kapsam dengeleme stratejisi olan ve olmayan testler arasında karşılaştırılmıştır. Kapsam dengeleme stratejisiyle uygulanan BBT'nin hem genel sağlık durumu hem de SF-36'nın iki sağlık alanı (fiziksel ve zihinsel) için daha iyi genel ölçme doğruluğu sunduğu bulunmuştur.

Özdemir (2015) maddeler-arası boyutluluk modeline dayalı çok boyutlu BBT yöntemlerinin performanslarını karşılaştırmış ve kapsam dengelemenin çok boyutlu BBT yöntemleri üzerindeki etkisini incelemiştir. Hacettepe Üniversitesi İngilizce Yeterlik Sınavında

her bir testte yer alan dinleme, dilbilgisi ve okuduğunu anlamaya ilişkin maddeler ile üç boyutlu gerçek madde havuzu oluşturulmuştur. En uygun çok boyutlu BBT yöntemi belirlenirken; iki farklı yetenek kestirim yöntemi (Bayesyen MAP ve Fisher'in puanlama yöntemi), üç farklı madde seçim yöntemi (A-optimality, D-optimality ve seçkisiz) ve hata varyansı durdurma kuralına dayalı üç farklı ölçüt kullanılmıştır. Kapsam dengelemenin yapıldığı ve yapılmadığı koşullara ilişkin sonuçlar boyutlara ilişkin güvenilirlik katsayıları, ölçmenin standart hatası ve RMSD değerlerine göre karşılaştırılmıştır. Fisher'in puanlama yönteminin hem madde seçim yöntemlerinden hem de kapsam dengelemeden etkilendiği ayrıca kapsam dengeleme uygulandığında her bir koşul için testteki ortalama madde sayısı artarken, güvenilirlik katsayılarının, RMSD ve standart hataların azaldığı bulunmuştur.

Sarı ve Manley (2017) çeşitli test uzunluklarında farklı içerik alanlarının sayısı değiştiğinde, bilgisayarda bireyselleştirilmiş test (BBT) ve çok aşamalı test (ÇAT) sonuçlarının kesinliğini araştırdıkları bir simülasyon çalışması yapmıştır. Bir BBT ve iki ÇAT tasarımı (1-3 ve 1-3-3 panel tasarımları), toplam test uzunluğu (24 maddelik ve 48 maddelik) ve kontrollü içerik alanlarının sayısı dahil olmak üzere çeşitli manipüle edilmiş koşullarda karşılaştırılmıştır. İçerik alanı koşulu, sıfır (içerik kontrolü yok), iki, dört, altı ve sekiz içerik alanını içerecek şekilde manipüle edilmiştir. Test uzunluğunun ve test uygulama modeli türünün, sonuçları içerik alanı sayısından daha fazla etkilediği görülmüştür. BBT, iki ÇAT tasarımından daha iyi performans göstermiştir. Çalışmada içerik alanı sayısının hem BBT hem de ÇAT üzerindeki etkisine dair herhangi bir kanıt bulunmamış ve kontrol edilen içerik alanlarının sayısını artırılması veya içerik kontrolü olmaması, çalışma sonuçlarının hiçbirini anlamlı bir şekilde etkilememiştir.

Şahin ve Özbaşı (2017) tarafından yapılan çalışmada BBT uygulamalarında kapsam dengeleme kullanıldığında madde seçim yöntemindeki değişikliğin yetenek kestirimine etkisini MFB ve onun bir alternatifi olan OAB yöntemlerini kullanarak belirlemek amaçlanmıştır. Monte Carlo simülasyonları ile dört grup (250, 500, 750, 1000) ve 10 farklı içerik alanına sahip 500 maddelik bir banka oluşturulmuştur. Gerçek ve kestirilen yetenek (θ) düzeyleri, katılımcılar altı alt gruba ayrılarak karşılaştırılmıştır. Bunun yanında kullanılan ortalama madde sayısı da

karşılaştırılmıştır. OAB kullanıldığında tüm sınava giren gruplar arasında θ düzeyi arttıkça korelasyonlar istikrarlı bir şekilde azalmıştır. Kapsam dengeleme uygulandığında, yetenek düzeyi -3 ile 0 arasındaki katılımcılar için OAB'nin θ tahmininde daha uygun olduğunu ve yetenek düzeyi θ 0'ın üzerinde olduğunda MFB'nin daha kararlı olduğu bulunmuştur. Bu iki madde seçim yöntemini birleştiren bir madde seçim algoritması önerilmiştir.

Demir (2019), Bilgisayarda Bireyselleştirilmiş Sınıflama Testlerinde (BBST) iki kategorili puanlanan maddelerden oluşan tek boyutlu madde havuzu üzerinden sınıflamanın iki, üç veya dört kategoride yapıldığı durumlarda ortalama test uzunluğu, ortalama sınıflama doğruluğu ve ölçme kesinliğinin; madde seçme yöntemlerine göre kapsam dengeleme ve madde kullanım sıklığı kontrolü gibi pratik kısıtlamalar altında nasıl değiştiğini incelemiştir. Çalışmada Kısıtlanmış Bilgisayarda Bireyselleştirilmiş Test (KBBT) ve Modifiye Edilmiş Multinomial Model (MÇM) ile kapsam dengeleme; Sympson-Hetter Yöntemi (SH) ve Madde Uygunluk Yöntemi ile madde kullanım sıklığı kontrol edilmiştir. Kapsam dengelemeyi sağlamada MÇM yöntemi daha başarılıyken, test etkililiği (ortalama test uzunluğu ve ortalama sınıflama doğruluğu) bakımından KBBT'nin daha avantajlı olduğu; madde kullanım sıklığı kontrolünde ise Madde Uygunluk Yöntemi daha iyi performans gösterirken, test etkililiğinde SH'nin daha avantajlı olduğu belirtilmiştir. Kategori sayısının artışı ortalama test uzunluğunu artırırken ortalama sınıflama doğruluğunu azaltmış; korelasyon, yanlılık, RMSE değerlerinde ise iyileşme sağlamıştır.

Yasuda ve Hull (2021), daha önceki çalışmalarında bilgisayarda bireyselleştirilmiş test formunu geliştirmiş oldukları Kuvvet Kavramı Envanteri'nin uygulanmasında kapsam dengeleme prosedürünü kullanmanın standartlaştırılmış ortalama farkın doğruluğu ve kesinliği üzerindeki etkisini analiz etmişlerdir. Envanter maddeleri 6 alt gruba ayrılmıştır. Kapsam dengelemeyi kontrol etmek için kısıtlı kapsam dengeleme yöntemi kullanılmıştır. Envanterin BBT formunun optimal uzunluğunu analiz etmek için standartlaştırılmış ortalama farkın, Cohen'in d'sinin doğruluğu ve kesinliği hesaplanmıştır. Yapılan monte carlo simülasyon çalışmasında kapsam dengeleme uygulandığında RMSE artışının 2-5 maddelik test

uzunlukları için %6 - 7 olduğu ve test uzunluğu 13 maddeden fazla ise %1'den az olduğu bulunmuştur. Yeterince büyük test uzunluğu için (örneğin, 13 öğeden daha büyük) kapsam dengelemenin Kuvvet Kavramı Envanterinin ölçme kesinliğini tehlikeye atmadığından bu koşulda kapsam dengelemenin kullanılması önerilmiştir.

Özer Taymur (2023) kapsam dengelemenin ölçme kesinliği üzerindeki etkisini; madde havuzu genişliği, madde seçme yöntemi, yetenek kestirim yöntemi ve madde kullanım sıklığı kontrolüne ait farklı koşullar altında araştırmıştır. Elde edilen RMSE, yanlılık ve korelasyon değerlerine göre en iyi performansın geniş madde havuzunda, madde seçiminde Maksimum Fisher Bilgisi, yetenek kestiriminde BSD yönteminin kullanıldığı koşullara ait olduğu bulunmuştur. Kapsam dengeleme yapıldığı durumlarda, yapılmadığı durumlara göre RMSE değerlerinde bir miktar artış olduğu belirtilmiştir. BBT uygulamalarında kapsam dengeleme ve madde kullanım sıklığı kontrol yöntemleri kullanılarak madde havuzunda yapılan sınırlandırmaların, bireyleri yetenek düzeylerine uygun olmayan güçlükteki maddeleri almaya zorlayabileceğinden, ölçme kesinliği ve test verimliliğini etkileyebileceği belirtilmiştir.

İlgili Araştırmalar Özeti

Alan yazın incelendiğinde yurt içinde yapılan çalışmalarda sıklıkla BBT'nin başlama kuralı, madde seçim yöntemi, yetenek kestirim yöntemi, madde kullanım sıklığı kontrolü ve sonlandırma kurallarından bazılarının manipüle edilmesiyle oluşturulan koşullarda elde edilen sonuçlar hata göstergeleri olan RMSE, yanlılık ve bazılarında uygulanan madde sayısına göre yorumlanmıştır. Yurt içinde yapılan çalışmalarda manipüle edilen bileşen olarak kapsam dengelemenin dahil edildiği çalışma sayısı oldukça sınırlıdır. Yurt dışında yapılan çalışmalarda ise farklı test uzunluklarında ve farklı sonlandırma kuralları kullanıldığında kapsam dengeleme yöntemlerinin etkililiğini inceleyen çalışmalar mevcuttur. Bazı çalışmalarda içerik alanı sayısının da manipüle edildiği görülmektedir. Sonuçlar madde kullanım sıklığı oranları, ölçme kesinliği ve madde havuzu kullanımının verimliliği açısından yorumlanmıştır. Yapılan çalışmaların çoğunda kapsam dengelemenin verimli madde havuzu kullanımında, daha düşük

madde kullanım sıklığı yüzdeleri elde etmede ve ölçme kesinliğini korumada etkili olduğu bulunmuştur.

Bu arařtırmada farklı test uzunluklarında ve farklı örneklem büyüklüklerinde, farklı madde seçme yöntemleri, yetenek kestirim yöntemleri ve sonlandırma kuralları ile kapsam dengelemenin ölçme kesinliđi üzerindeki etkisinin incelenmesi özellikle yurt içi literatüre getireceđi katkı açısından önemlidir.



Bölüm 3

Yöntem

Bu bölümde araştırmanın yöntemi, simülasyon koşulları, verilerin üretilmesi, simülasyon uygulamaları ve verilerin analizi ile ilgili bilgilere yer verilmiştir.

Araştırma Yöntemi

Çalışma ile simülatif bilgisayarda bireyselleştirilmiş test uygulamalarında, iki kategorili puanlanan maddelerde kapsam dengelemenin, farklı koşullarda ölçme kesinliğini nasıl etkilediğinin incelenmesi amaçlanmaktadır. Araştırmada kullanılan madde ve yetenek parametreleri simülatif olarak üretilmiş ve BBT uygulaması simülasyonlar yardımıyla gerçekleştirilmiştir. Bu nedenle araştırma verilerin simülatif olarak elde edildiği Monte Carlo simülasyon çalışmasıdır.

Verilerin Üretilmesi

Araştırmanın amacına yönelik farklı ölçme koşullarının sağlanabileceği veri setlerine ulaşmak oldukça güç olduğundan simülatif veriler kullanılmıştır. Bilgisayarda bireyselleştirilmiş test uygulaması katılımcıları araştırmacı tarafından R programlama dili kullanılarak simülatif olarak oluşturulmuştur (R Core Team, 2022). İlk olarak bireylere ait yetenek parametre değerleri (gerçek θ), daha sonra madde parametre değerleri üretilmiştir. BBT uygulamaları için farklı simülasyon koşulları belirlenmiş; belirlenen koşullar aynı madde havuzları ve aynı bireyler üzerinden karşılaştırmalı olarak incelenmiştir.

Yetenek Parametrelerinin Üretilmesi

BBT uygulamasına katılan bireyler araştırmacı tarafından R programlama dilinde yer alan “catR” (Magis, vd., 2018) ve “MCMCglmm” (Hadfield, 2010) paketlerinden faydalanılarak simüle edilmiştir (R Core Team, 2022). 250 ve 500 kişilik iki farklı büyüklükte örneklem oluşturulmuştur. Wise ve DeMars (2005), BBT uygulamalarında örneklem büyüklüğünün test sonuçlarının geçerliği ve güvenilirliği üzerindeki etkilerini inceledikleri çalışmalarında, 250 kişilik

bir örneklemin küçük ölçekli simülasyonlar için yeterli olduğunu, 500 kişilik örneklemin ise daha genelleştirilebilir sonuçlar sağladığını belirtmiştir. Araştırma kapsamında bireylerin yetenek parametreleri her bir örneklem büyüklüğü koşulu için θ değeri -3 ve +3 aralığında normal dağılım $\theta \sim N(0, 1)$ gösterecek şekilde üretilmiştir.

Madde Havuzunun Oluşturulması

Madde parametreleri tek biçimli dağılım gösterecek şekilde araştırmacı tarafından belirlenmiştir. Feinberg ve Rubright (2016), üç parametrelili lojistik model kullanılırken madde parametrelerinin genellikle tek biçimli (uniform) dağılıma uygun olarak simüle edildiğini belirtmiştir. Kapsam dengeleme yapılırken, üç parametrelili lojistik (3PL) madde tepki modeli kullanılarak iki kategorili puanlanan beş farklı içerik alanı, eşit oranlarda ağırlıklandırılarak 750 maddelik madde havuzu oluşturulmuştur. BBT uygulamaları, madde havuzunda yer alan maddeler, farklı yetenek düzeylerine hitap edecek şekilde eşit dağılıma sahip yeterli sayıda ve ayırt ediciliği yüksek olduğunda daha iyi sonuçlar vermektedir (DeMars, 2010; Flaugher, 2000). Bu nedenle madde ayırt edicilik parametreleri “a” (0.5 ile 2); madde güçlük parametreleri “b” (-3 ile 3) ve şans parametreleri “c” (.05 ile 0.2) aralığında tek biçimli dağılım gösterecek şekilde üretilmiştir (Ree & Jensen, 1983; Thompson, 2009). Madde havuzu beş farklı içerik alanının her birinden 150 madde içerecek şekilde 750 maddeden oluşturulmuştur. Oluşturulan madde havuzundaki maddelere ait betimsel istatistikler Tablo 1’de verilmiştir.

Tablo 1

Madde Havuzundaki Maddelerin Betimsel İstatistikleri

Madde parametreleri	Minimum Değeri	Maksimum Değeri	Ortalama	Standart Sapma
a	0.502	1.999	1.225	0.446
b	-2.989	2.976	0.020	1.691
c	0.050	0.199	0.123	0.045

BBT Simülasyon Koşulları

BBT uygulaması kapsamında ele alınan koşullar; kapsam dengeleme yapıldığı ve yapılmadığı durumlarda kullanılan farklı örneklem büyüklüğü, farklı madde seçme yöntemi, farklı yetenek kestirim yöntemi ve farklı sonlandırma kurallarını içermektedir. Araştırmada farklı büyüklükte 2 örneklem, kapsam dengelemenin yapıldığı ve yapılmadığı 2 farklı durum, 2 madde seçme yöntemi, 2 yetenek kestirim yöntemi ve 5 sonlandırma kuralı olmak üzere toplam 80 koşul belirlenmiştir. Mevcut çalışmada en sık kullanılan yetenek kestirim yöntemlerinden; MOK ve BSD yöntemi koşul olarak ele alınmıştır (Chen vd., 1998; Segall, 2004). Belirlenen simülasyon koşulları Tablo 2'de yer almaktadır:

Tablo 2

BBT Simülasyon Koşulları

BBT Bileşenleri	Koşullar	Koşul Sayıları
Sonlandırma Kuralı	20 madde	5
	60 madde	
	SH≤.30	
	SH≤.40	
	SH≤.50	
Örneklem Büyüklüğü	250 birey	2
	500 birey	
Yetenek Kestirim Yöntemi	MOK	2
	BSD	
Test Başlatma Kuralı	-0.5≤b≤0.5	1
Madde Seçim Yöntemi	MFB	2
	KL	
Madde Kullanım Sıklığı Kontrolü	Yok	1
Kapsam Dengeleme	Yapıldı	2
	Yapılmadı	

Bireyin yeteneği hakkında ön bilgiye sahip olunmadığı durumda, ortalama yetenek düzeyine sahip olduğunu kabul etmek en uygun tahmindir. Başlangıç maddesi ortalama güçlük düzeyinde olduğunda BBT uygulaması psikometrik açıdan daha etkili olacaktır (Mills ve Stocking, 1996). Bu nedenle simülatif BBT uygulaması için test başlatma kurallarından -

.50≤b≤.50 aralığı yöntemi kullanılmıştır. BBT simülasyonu için “catR” paketi kullanılmıştır (Magis, vd., 2018). Belirlenen 80 simülasyon koşulunun tamamı için 50 tekrarın ortalaması alınarak değerler hesaplanmıştır. Kapsam dengelemenin kısa ve uzun testlerdeki etkisinin incelenmesi amacıyla 20 ve 60 maddede testin sonlandırılması koşullara eklenmiştir. Çalışmada 60 maddelik test uzunluğunun seçilmesi benzer araştırmalarla desteklenmekte ve çeşitli avantajlar sunmaktadır. Kingsbury vd. (2009), 60 maddelik bir testin yeterli içerik temsili sağlamanın yanı sıra iki katı uzunluktaki geleneksel testlerle eşdeğer geçerli ve güvenilir puanlar sağladığını göstermiştir. Dahası, Sarı (2019), daha uzun testlerin test güvenliği ve güvenilirliğiyle ilgili olumsuz etkileri azalttığını göstermiştir. Bu nedenle, 60 maddelik test uzunluğunun yeterli içerik temsiliyi garanti etmesi ve yüksek test güvenilirliğini ve geçerliğini koruması, mevcut çalışmanın hedefleriyle örtüşmektedir.

Test verimliliği korunarak kapsam dengeleme yapılabilmesi için önerilen yöntemlerden en sık kullanılan, basit ve anlaşılır bir yöntem olan Kısıtlı Bilgisayarda Bireyselleştirilmiş Test (KBBT) yöntemidir (Kingsbury & Zara, 1989). Mevcut simülasyon çalışmasında verilerin analizi için kullanılan “catR” paketinde yer alan KBBT yöntemi kapsam dengeleme yöntemi olarak kullanılmış, diğer kapsam dengeleme yöntemleri çalışmanın kapsamı dışında tutulmuştur. BBT uygulamasında madde kullanım sıklığı kontrolü yapılmamıştır.

Verilerin Analizi

Araştırma kapsamında ele alınacak her bir koşul için ölçme kesinliği değerlendirilirken yaygın olarak kullanılan hata kareleri ortalamasının karekökü RMSE (Root Mean Squared Error), tahmin edilen yetenek düzeyinin gerçek yetenek düzeyinden sapmasının ölçüsü olan yanlılık ve tahmini ve gerçek yetenek düzeyleri arasındaki ilişkiye bağlı olarak elde edilen korelasyon değerleri anlamına gelen uyum değerleri kullanılmıştır (Babcock & Weiss, 2012). Çoğu MTK çalışması için, Harwell vd. (1996) örneklem yanlılığını azaltmak, istikrarlı ve güvenilir sonuçlar elde etmek için en az 25 replikasyon önermektedir ancak bazı çalışmalarda

bu sayının çok daha yüksek olabileceğini de belirtmişlerdir. Çalışmada bu değerler 50 replikasyonun her biri için ayrı ayrı hesaplanmış ve ardından ortalaması alınmıştır.

Uyum katsayısı; bireylerin başlangıçta simülatif olarak üretilmiş olan gerçek θ seviyeleri ile her bir araştırma koşulunda ve replikasyonda tahmin edilen θ seviyeleri arasındaki korelasyon hesaplanarak değerlendirilmiştir. Her bir katılımcı için kestirilen θ değerlerinin ortalama korelasyonunu elde etmek için bu korelasyonların ortalaması alınmıştır. Uyum katsayısını hesaplamak için Pearson Momentler Çarpımı Korelasyonu kullanılmıştır. Uyum katsayısı aşağıdaki formül kullanılarak hesaplanmaktadır:

$$r = \frac{\text{cov}(\hat{\theta}, \theta)}{ss(\hat{\theta})ss(\theta)}$$

Her bir madde için her bir replikasyonda kestirilen parametre değeri ile gerçek parametre değeri arasındaki farkın karesinin ortalamasının karekökü olan RMSE, kestirimlerin doğruluğunu değerlendirmek için en sık kullanılan ölçülerden biridir. Öklid mesafesini kullanarak kestirimlerin gerçek değerlerden ne kadar uzaklaştığını göstermektedir. Kestirime ilişkin sistematik hatayı ifade eden yanlılık, her bir replikasyonda her bir madde için kestirilen parametre değerinin ortalaması ile gerçek parametre değeri arasındaki farka eşit olup, ölçme kesinliğinin göstergesi olan bir başka ölçüdür. RMSE ve yanlılık değerleri aşağıdaki eşitliklerle hesaplanmaktadır (Zheng ve Chang, 2014):

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

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

Eşitliklerde n birey sayısını, θ_i bireyin gerçek yetenek düzeyini ve $\hat{\theta}_i$ ise bireyin kestirilen yetenek düzeyini ifade etmektedir. Uyum katsayısının yüksek olması, yanlılık ve RMSE değerlerinin düşük olması gerçek yetenek düzeyi ile kestirilen yetenek düzeyi arasında farkın olmadığına işaret etmektedir. Sonlandırma koşulu olarak $SH \leq .30$, $SH \leq .40$ ve $SH \leq .50$ alınan durumlardaki ortalama test uzunluğu da incelenmiştir.

Bölüm 4

Bulgular, Yorumlar ve Tartışma

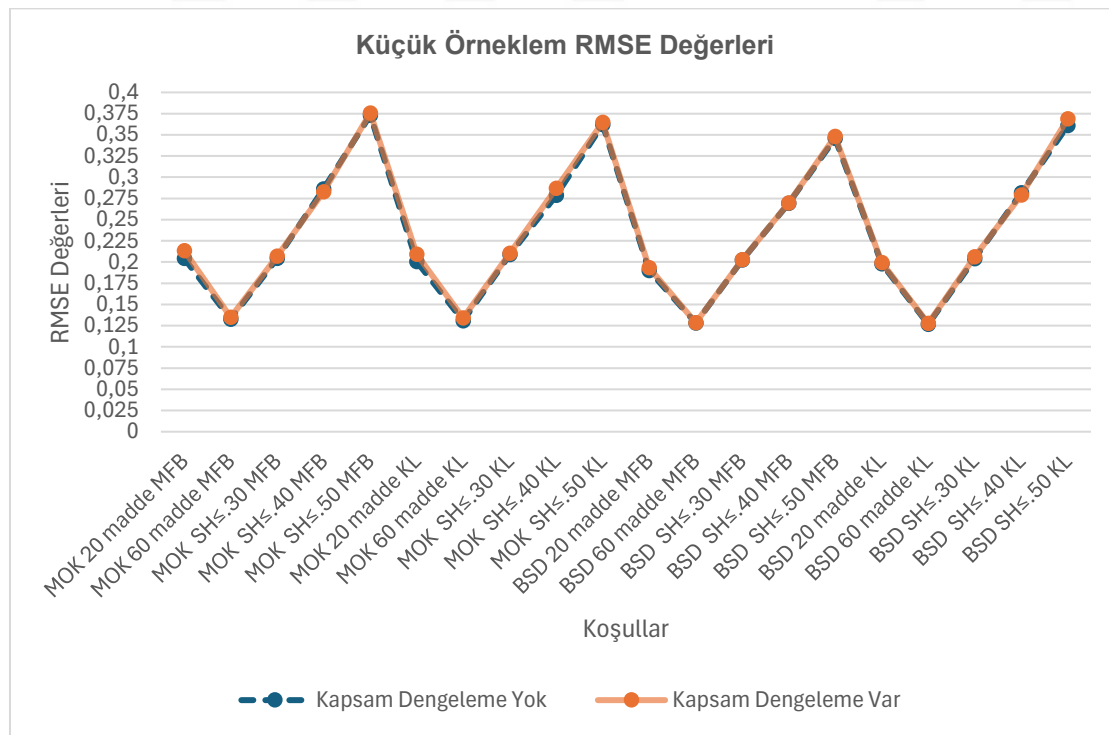
Bu bölümde araştırmanın alt problemlerine ait bulgulara ve yorumlara yer verilmiştir. Araştırma kapsamında belirlenen 80 koşula ait ölçme kesinliğinin göstergesi olan RMSE, yanlılık ve uyum katsayıları ve bu değerlere ait standart hatalar Eklerde sunulmuştur.

Birinci Alt Probleme Ait Bulgular ve Yorumlar

Araştırmanın birinci alt probleminde bilgisayarda bireyselleştirilmiş test uygulamalarında, farklı örneklem büyüklüklerinde (N=250, N=500) kapsam dengeleme yapılmasının ölçme kesinliği ve yetenek kestirimini nasıl etkilediği araştırılmıştır. Çalışmada ele alınan koşullar simülatif olarak oluşturulmuş 250 ve 500 kişilik örneklem üzerinde analiz edilmiştir. Farklı örneklem büyüklüklerine ait RMSE değerleri Şekil 6 ve 7’de verilmiştir.

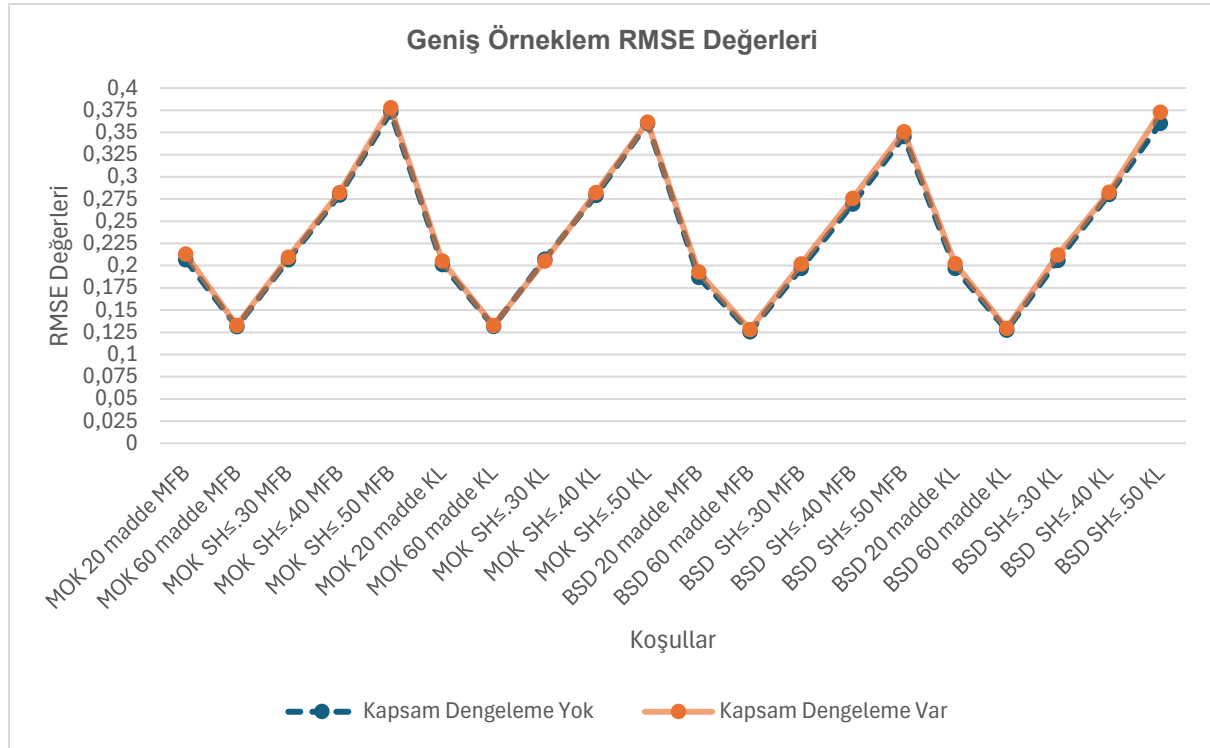
Şekil 6

Küçük Örneklemden Elde Edilen RMSE Değerleri



Şekil 7

Geniş Örneklemden Elde Edilen RMSE Değerleri



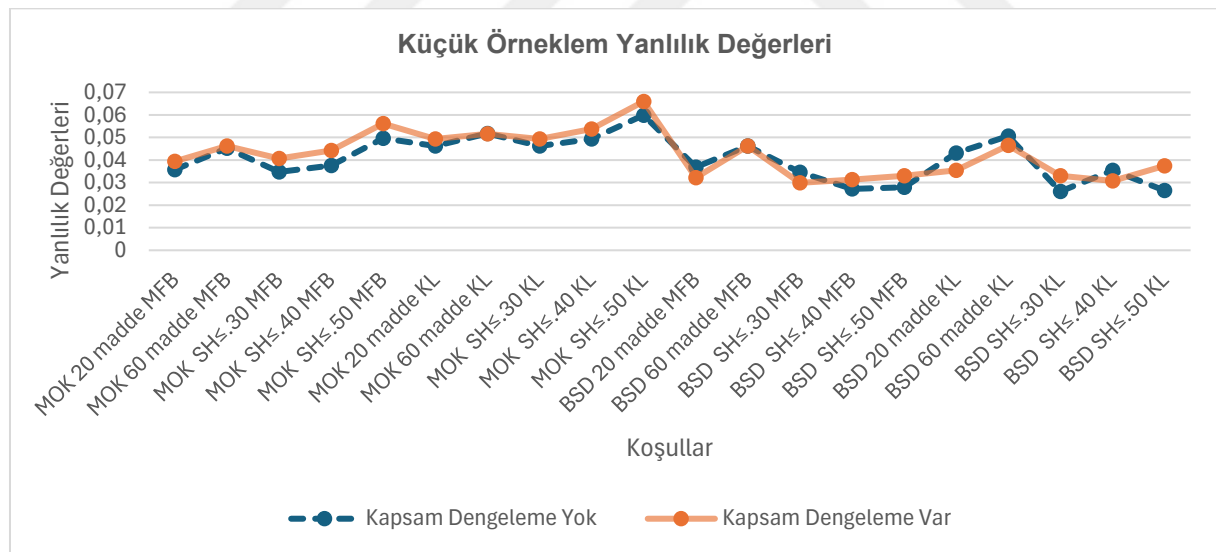
Farklı örneklem büyüklüklerinden elde edilen RMSE değerleri incelendiğinde Şekil-6 ve 7'de de görüldüğü gibi değerler birbirine oldukça yakındır. Her iki örneklem büyüklüğünde de en düşük RMSE değerleri testin 60 maddede sonlandırıldığı durumda elde edilmiştir. Farklı örneklem büyüklüklerinde, hemen hemen tüm koşullarda kapsam dengeleme yapıldığında elde edilen RMSE değeri yapılmadığı duruma göre az miktarda artış göstermiştir. Tüm koşullar örneklem büyüklüklerine göre incelendiğinde her iki örneklem büyüklüğünde de elde edilen RMSE değerlerinin tamamına yakınında BSD yetenek kestirim yöntemi kullanıldığında ML yetenek kestirim yöntemine göre düşüş görülmüştür. Şekil-7'de de görüldüğü gibi tüm koşullarda örneklem büyüklüğü RMSE değerlerini etkilememiştir. Hem 250 hem de 500 kişilik örneklemelerde kullanılan madde seçim yöntemi genel olarak RMSE değerlerinde farklılık yaratmamıştır Bunun yanı sıra BSD kestirim yöntemi kullanıldığında, SH≤.40 ve SH≤.50 sonlandırma kurallarında her iki örneklem büyüklüğünde de kapsam dengeleme yapıp yapılmaması fark etmeksizin, KL madde seçim yöntemi nispeten daha yüksek RMSE değerleri vermiştir.

Küçük örnekleme en düşük RMSE değeri (0.127); BSD kestirim yöntemi ve KL madde seçim yöntemi kullanıldığında, test 60 maddede sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek RMSE değeri (0.376) ise MOK yöntemi ve MFB madde seçim yöntemi kullanıldığında, test $SH \leq 50$ kuralına göre sonlandırılıp kapsam dengelemenin yapıldığı koşulda elde edilmiştir. Benzer şekilde geniş örnekleme en düşük RMSE değeri (0.126); BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanıldığında, test 60 maddede sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek RMSE değeri (0.378) ise küçük örnekleme olduğu gibi; MOK yöntemi ve MFB madde seçim yöntemi kullanıldığında, test $SH \leq 50$ kuralına göre sonlandırılıp kapsam dengelemenin yapıldığı koşulda elde edilmiştir.

Farklı örneklem büyüklüklerine ait yanlılık değerleri Şekil 8 ve 9'da verilmiştir.

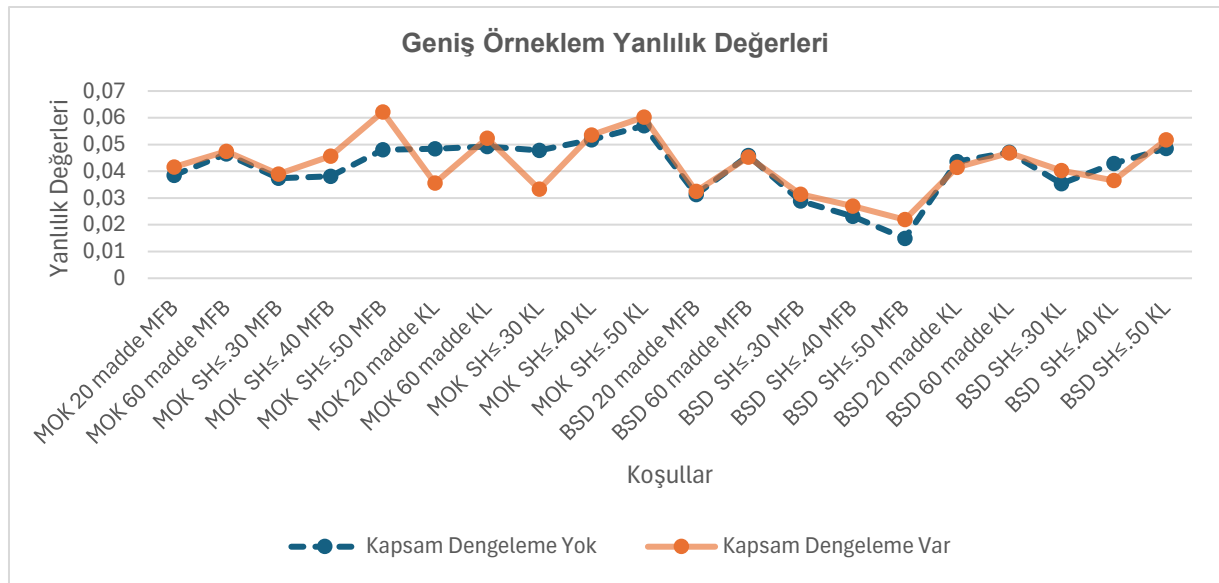
Şekil 8

Küçük Örneklemden Elde Edilen Yanlılık Değerleri



Şekil 9

Geniş Örneklemden Elde Edilen Yanlılık Değerleri



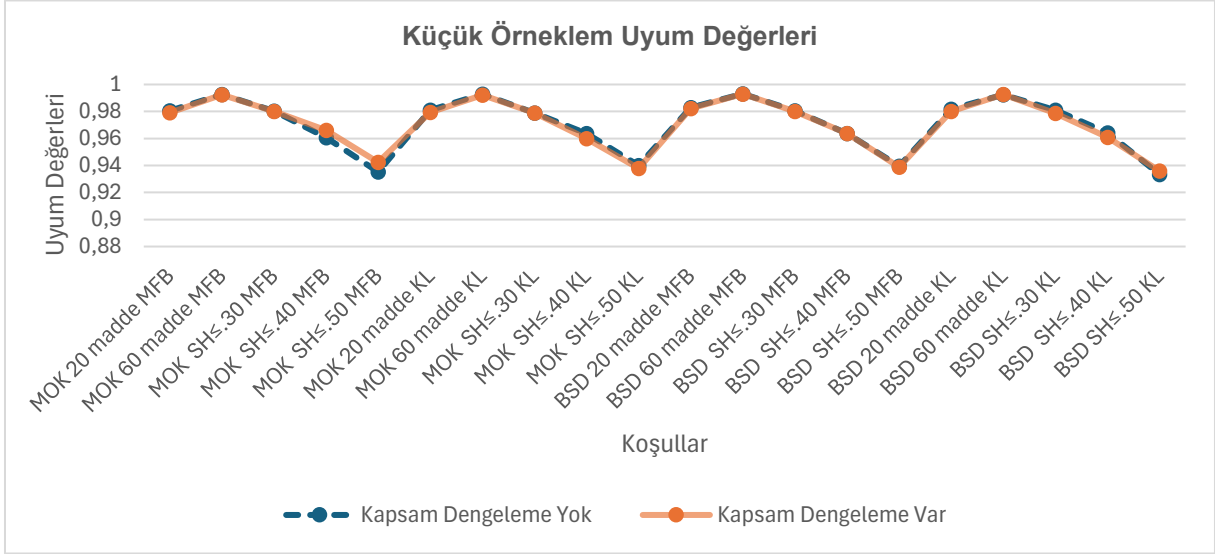
Örneklem büyüklüğüne göre yanlılık değerleri incelendiğinde kapsam dengeleme yapıldığı ve yapılmadığı koşullarda küçük örnekleme değerlerin birbirine oldukça yakın olduğu görülmektedir. Geniş örnekleme ise çoğu koşulda yanlılık değerleri birbirine oldukça yakinken MOK kestirim yöntemi ve KL madde seçim yöntemi kullanılan koşullarda, test SH≤.30 ve 20 madde kuralına göre sonlandırıldığında, kapsam dengeleme yapılan koşullarda az da olsa düşüş gösterdiği göze çarpmaktadır. Geniş örnekleme MOK yöntemi ve MFB madde seçim yöntemi kullanılan koşullarda, test SH≤.40 ve SH≤.50 kuralına göre sonlandırıldığında kapsam dengeleme yapılmayan koşullarda yanlılık değeri daha düşük bulunmuştur. Küçük örnekleme en düşük yanlılık değeri (0.026); BSD kestirim yöntemi ve KL madde seçim yöntemi kullanıldığında, test SH≤.30 kuralıyla sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek yanlılık değeri (0.066) ise MOK yöntemi ve KL madde seçim yöntemi kullanıldığında, test SH≤.50 kuralına göre sonlandırılıp kapsam dengelemenin yapıldığı koşulda elde edilmiştir. Geniş örnekleme en düşük yanlılık değeri (0.015); BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanıldığında, test SH≤.50 kuralıyla sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek yanlılık değeri

(0.062) ise; MOK yöntemi ve MFB madde seçim yöntemi kullanıldığında, test $SH \leq 50$ kuralına göre sonlandırılıp kapsam dengelemenin yapıldığı koşulda elde edilmiştir.

Farklı örneklem büyüklüklerine ait uyum değerleri Şekil 10 ve 11’de verilmiştir.

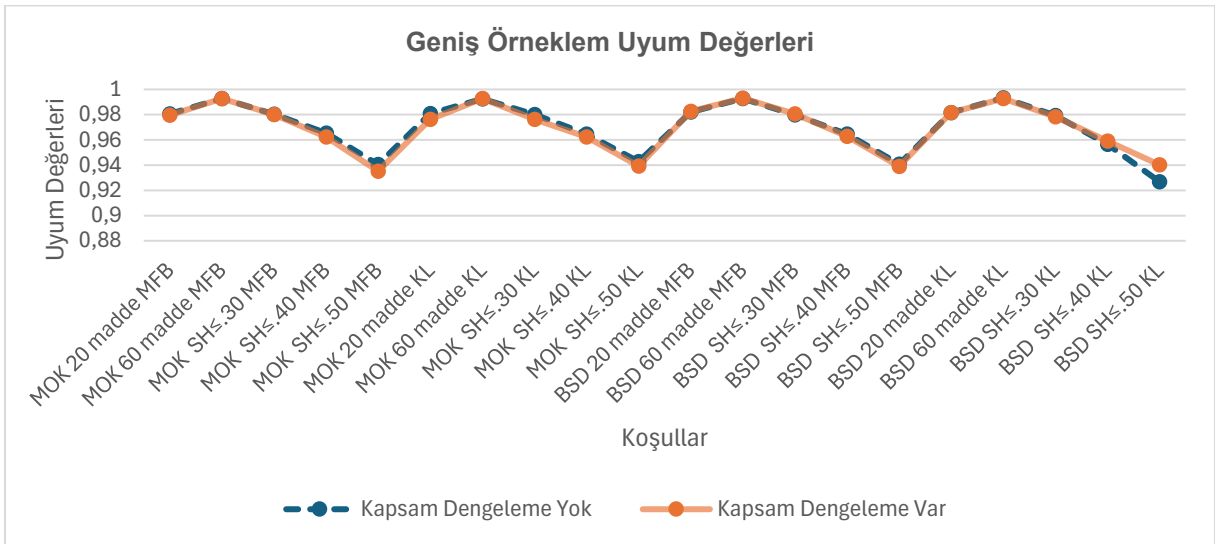
Şekil 10

Küçük Örneklemden Elde Edilen Uyum Değerleri



Şekil 11

Geniş Örneklemden Elde Edilen Uyum Değerleri



Örneklem büyüklüğüne göre uyum katsayıları incelendiğinde hem küçük hem de geniş örnekte kapsam dengeleme yapılan ve yapılmayan koşullarda uyum katsayılarının

birbirine oldukça yakın ve yüksek olduğu görülmüştür. Verilen şekillerde de görüldüğü gibi uyum katsayısının örneklem büyüklüğünden etkilenmediği söylenebilir. Her iki örnekleme de kullanılan madde seçim yöntemine ve yetenek kestirim yöntemine bakılmaksızın en düşük uyum değerleri testin $SH \leq 50$ kuralına göre sonlandırıldığı koşullardan elde edilmiştir.

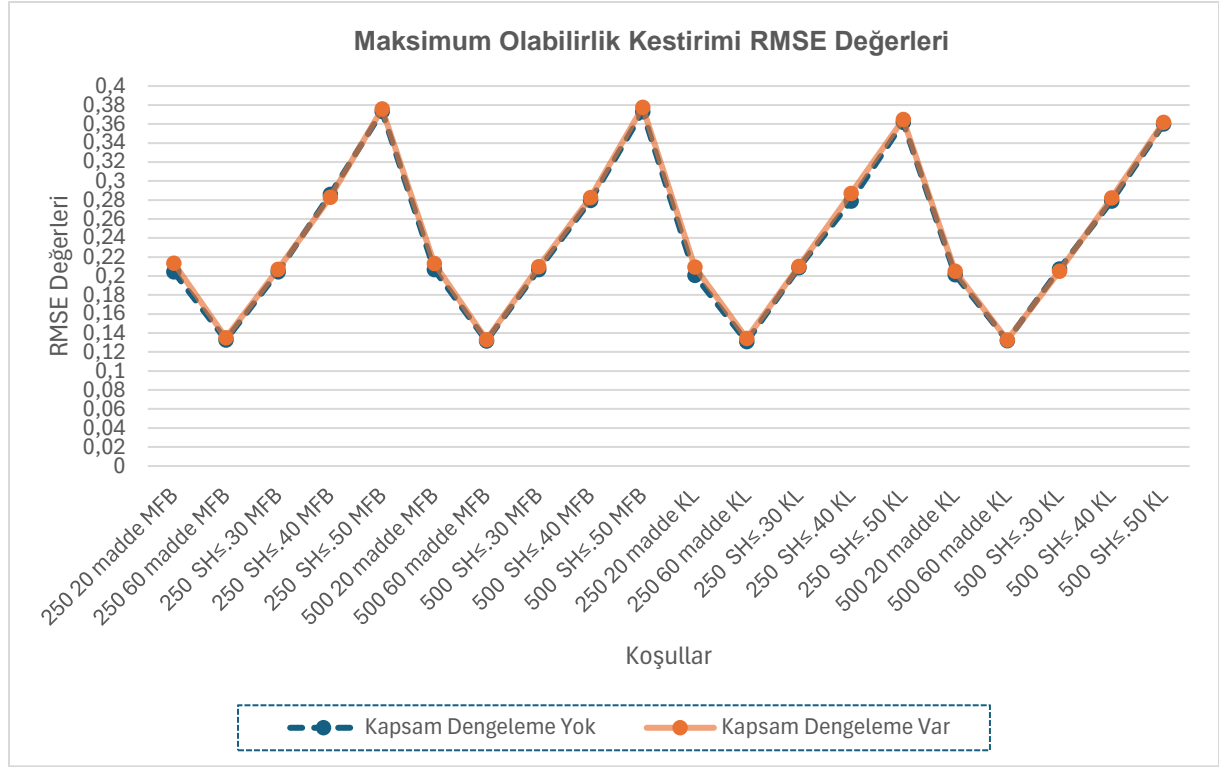
Küçük örnekleme en düşük uyum değeri (0.933); BSD kestirim yöntemi ve KL madde seçim yöntemi kullanıldığında, test $SH \leq 50$ kuralıyla sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek uyum değeri (0.993) ise BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanıldığında, test 60 maddede sonlandırıldığı koşullarda elde edilmiştir. Geniş örnekleme en düşük uyum değeri (0.927); BSD kestirim yöntemi ve KL madde seçim yöntemi kullanıldığında, test $SH \leq 50$ kuralıyla sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek uyum değeri (0.993) ise; BSD kestirim yöntemi ve KL madde seçim yöntemi kullanıldığında, test 60 maddede sonlandırılıp kapsam dengelemenin yapılmadığı koşulda elde edilmiştir.

İkinci Alt Probleme Ait Bulgular ve Yorumlar

Araştırmanın ikinci alt probleminde bilgisayarda bireyselleştirilmiş test uygulamalarında, farklı yetenek kestirim yöntemlerinde (MOK, BSD) kapsam dengeleme yapılmasının ölçme kesinliği ve yetenek kestirimini nasıl etkilediği araştırılmıştır. Farklı yetenek kestirim yöntemlerine ait RMSE değerleri Şekil 12 ve 13'te verilmiştir.

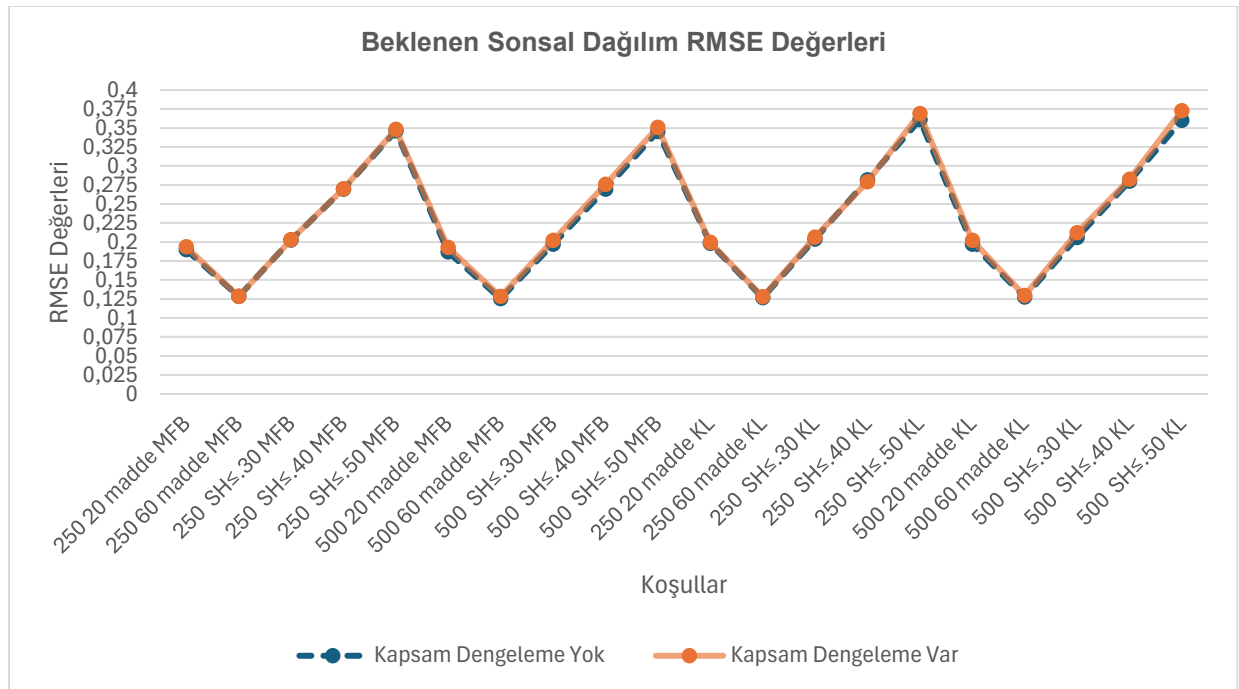
Şekil 12

Maksimum Olabilirlik Kestiriminden Elde Edilen RMSE Değerleri



Şekil 13

Beklenen Sonsal Dağılımdan Elde Edilen RMSE Değerleri

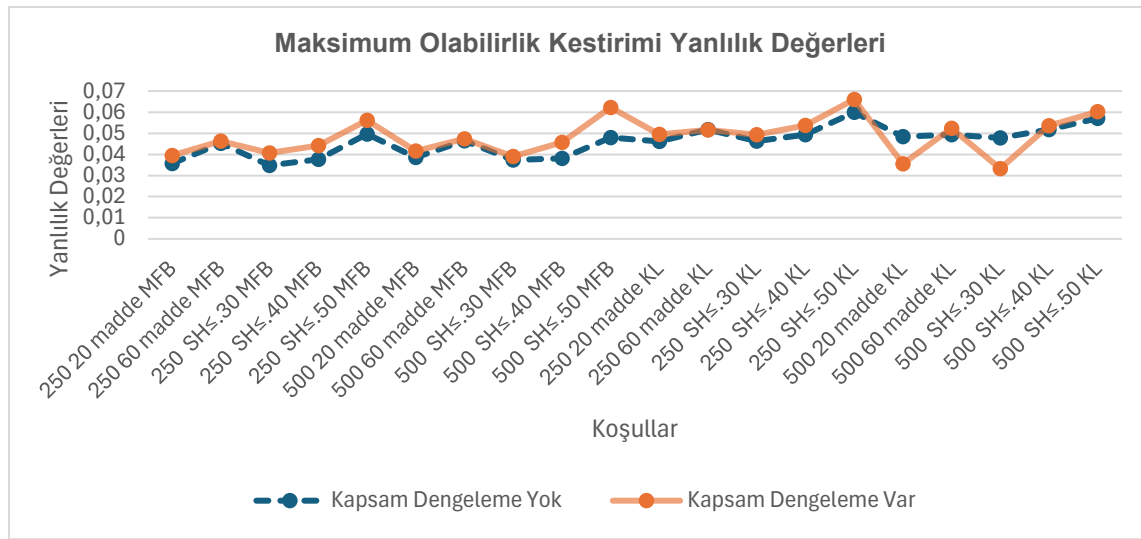


Farklı yetenek kestirim yöntemlerinden elde edilen RMSE değerleri incelendiğinde Şekil 12 ve 13'te de görüldüğü gibi değerler birbirine oldukça yakındır. Her iki kestirim yönteminde de en düşük RMSE değerleri testin 60 maddede sonlandırıldığı durumda elde edilmiştir. Kapsam dengeleme yapılan koşullarda yapılmayan koşullara nispeten RMSE değerlerinde oldukça az miktarda bir artış mevcuttur. Grafiklerde de görüldüğü gibi kapsam dengeleme yapılması koşullardan elde edilen RMSE değerlerini etkilememiştir. MOK yönteminde en düşük RMSE değeri (0.131); 250 kişilik örnekleme ve KL madde seçim yöntemi kullanıldığında, test 60 maddede sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek RMSE değeri (0.378) ise 500 kişilik örneklem ve MFB madde seçim yöntemi kullanıldığında, test $SH \leq 50$ kuralına göre sonlandırılıp kapsam dengelemenin yapıldığı koşulda elde edilmiştir. BSD kestirim yönteminde de benzer şekilde, en düşük RMSE değeri (0.127); 250 kişilik örnekleme ve KL madde seçim yöntemi kullanıldığında, test 60 maddede sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek RMSE değeri (0.373) ise 500 kişilik örneklem ve KL madde seçim yöntemi kullanıldığında, test $SH \leq 50$ kuralına göre sonlandırılıp kapsam dengelemenin yapıldığı koşulda elde edilmiştir. Kapsam dengeleme yapılan ve yapılmayan MFB madde seçim yöntemi kullanılan koşullarda RMSE değeri, MOK kestirim yöntemi tercih edildiğinde BSD'a göre daha yüksek bulunmuştur. Fakat KL madde seçim yöntemi kullanıldığında elde edilen değerler birbirine oldukça yakındır.

Farklı yetenek kestirim yöntemlerine ait yanlılık değerleri Şekil 14 ve 15'te verilmiştir.

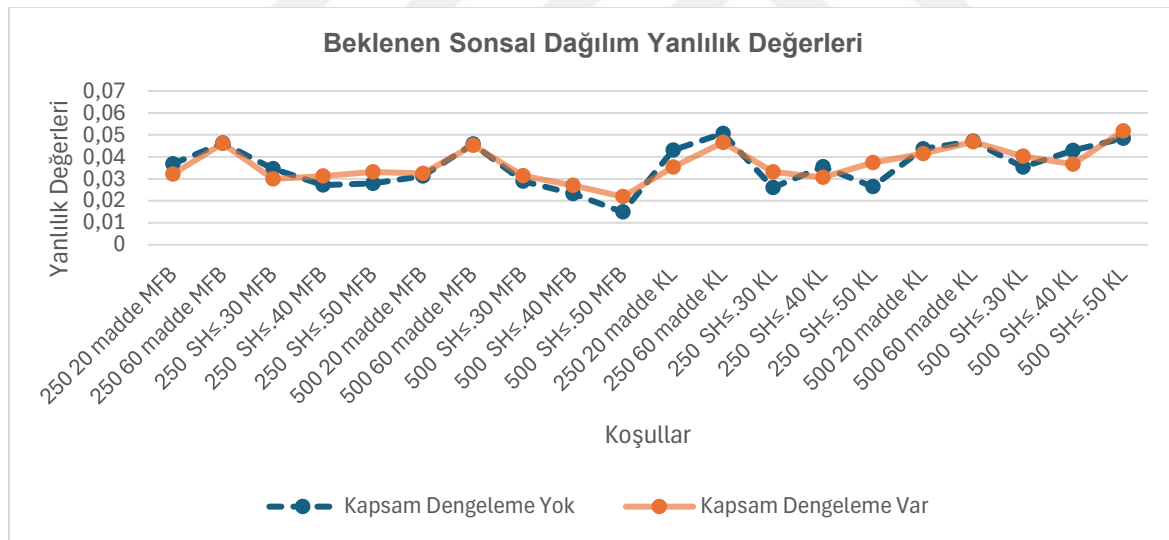
Şekil 14

Maksimum Olabilirlik Kestiriminden Elde Edilen Yanlılık Değerleri



Şekil 15

Beklenen Sonsal Dağılımdan Elde Edilen Yanlılık Değerleri



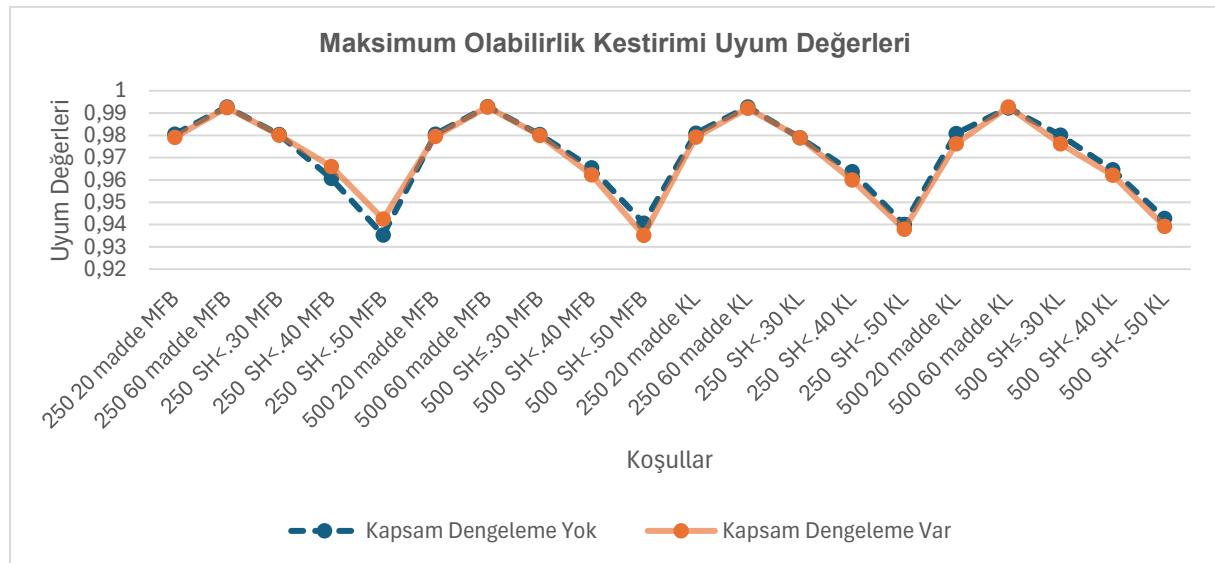
Farklı yetenek kestirim yöntemlerinden elde edilen yanlılık değerleri incelendiğinde genel olarak hem kapsam dengeleme yapılan hem de yapılmayan koşullarda, BSD kestirim yöntemi kullanılan koşullarda elde edilen yanlılık değeri MOK yöntemine göre daha düşük bulunmuştur. MOK yönteminde en düşük yanlılık değeri (0.033); 500 kişilik örnekleme ve KL madde seçim yöntemi kullanıldığında, test $SH \leq .30$ kuralıyla sonlandırılıp kapsam dengeleme yapıldığı koşulda elde edilmiştir. En yüksek yanlılık değeri (0.066) ise 250 kişilik örnekleme ve

KL madde seçim yöntemi kullanıldığında, test $SH \leq 50$ kuralına göre sonlandırılıp kapsam dengelemenin yapıldığı koşulda elde edilmiştir. MOK yönteminde koşulların çoğunda kapsam dengeleme yapılan durumlarda yanlılık daha yüksek bulunmuştur. Yalnızca 500 kişilik örnekleme KL madde seçim yöntemi kullanılıp testin 20 maddede ve $SH \leq 30$ kuralıyla sonlandırıldığı koşullarda kapsam dengeleme yapıldığında yanlılık daha düşük bulunmuştur. BSD kestirim yönteminde ise 250 kişilik örnekleme test 20 maddede sonlandırıldığında, madde seçim yöntemi fark etmeksizin, kapsam dengeleme yapıldığı durumda yanlılık değeri daha düşük bulunmuştur. Benzer şekilde örneklem büyüklüğü fark etmeksizin KL madde seçim yöntemi kullanılıp test $SH \leq 40$ kuralıyla sonlandırıldığında kapsam dengeleme yapılan koşullarda yanlılık değeri daha düşük bulunmuştur. BSD kestirim yönteminde en düşük RMSE değeri (0.015); 500 kişilik örnekleme ve MFB madde seçim yöntemi kullanıldığında, test $SH \leq 50$ kuralıyla sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek yanlılık değeri (0.052) ise MOK'a benzer şekilde, 500 kişilik örneklem ve KL madde seçim yöntemi kullanıldığında, test $SH \leq 50$ kuralına göre sonlandırılıp kapsam dengelemenin yapıldığı koşulda elde edilmiştir.

Farklı yetenek kestirim yöntemlerine ait uyum değerleri Şekil 16 ve 17'de verilmiştir.

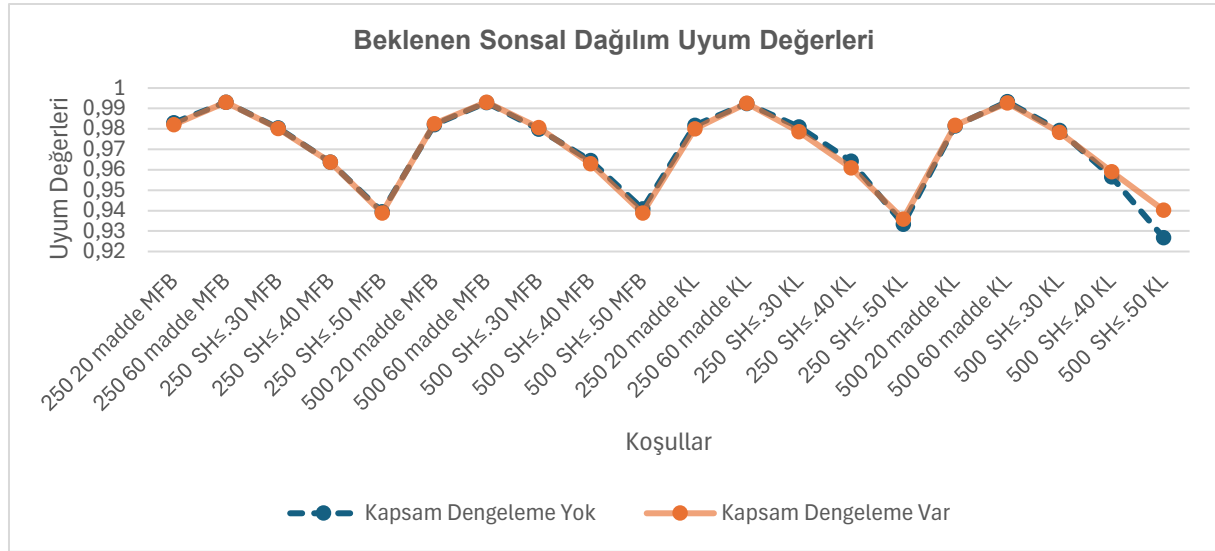
Şekil 16

Maksimum Olabilirlik Kestiriminden Elde Edilen Uyum Değerleri



Şekil 17

Beklenen Sonsal Dağılımdan Elde Edilen Uyum Değerleri



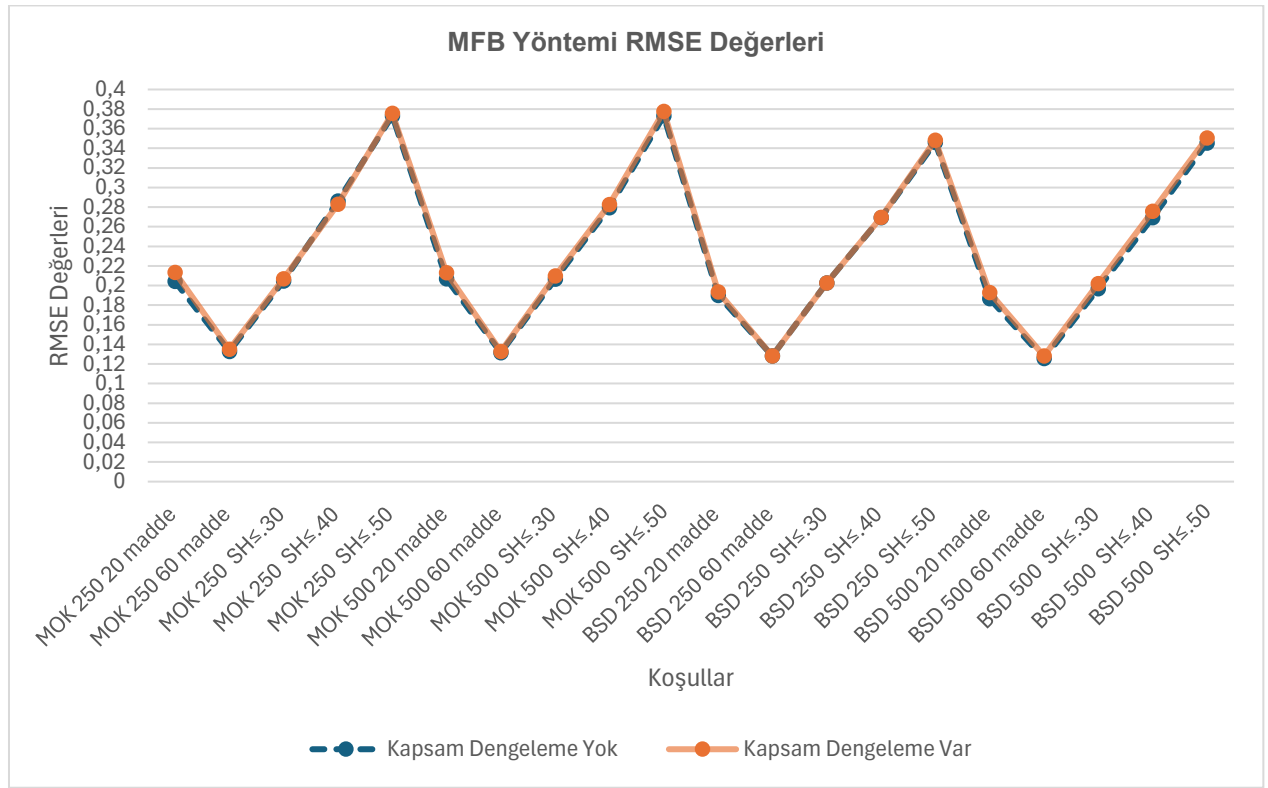
Farklı yetenek kestirim yöntemlerinden elde edilen uyum değerleri incelendiğinde Şekil 16 ve 17’de de görüldüğü gibi değerler birbirine oldukça yakındır. Her iki kestirim yönteminde de en düşük uyum değerleri testin SH≤.50 kuralı ile sonlandırıldığı koşullarda elde edilmiştir. Hem kapsam dengeleme yapılan hem de yapılmayan koşullarda her iki yetenek kestirim yönteminden elde edilen uyum değerleri birbirine oldukça yakındır. MOK yönteminde en düşük uyum değeri (0.935); 500 kişilik örnekleme ve MFB madde seçim yöntemi kullanıldığında, test SH≤.50 kuralıyla sonlandırılıp kapsam dengeleme yapıldığı koşulda elde edilmiştir. En yüksek uyum değeri (0.993) ise 500 kişilik örneklem ve MFB madde seçim yöntemi kullanıldığında, test 60 maddede sonlandırılıp kapsam dengelemenin yapılmadığı koşulda elde edilmiştir. BSD kestirim yönteminde ise en düşük uyum değeri (0.927); 500 kişilik örnekleme ve KL madde seçim yöntemi kullanıldığında, test SH≤.50 kuralıyla sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek uyum değeri (0.993) ise 500 kişilik örneklem ve KL madde seçim yöntemi kullanıldığında, test 60 maddede sonlandırılıp kapsam dengelemenin yapılmadığı koşulda elde edilmiştir.

Üçüncü Alt Probleme Ait Bulgular ve Yorumlar

Araştırmanın üçüncü alt probleminde bilgisayarda bireyselleştirilmiş test uygulamalarında, farklı madde seçim yöntemlerinde (MFB, KL) kapsam dengeleme yapılmasının ölçme kesinliği ve yetenek kestirimini nasıl etkilediği araştırılmıştır. Farklı madde seçim yöntemlerine ait RMSE değerleri Şekil 18 ve 19'da verilmiştir.

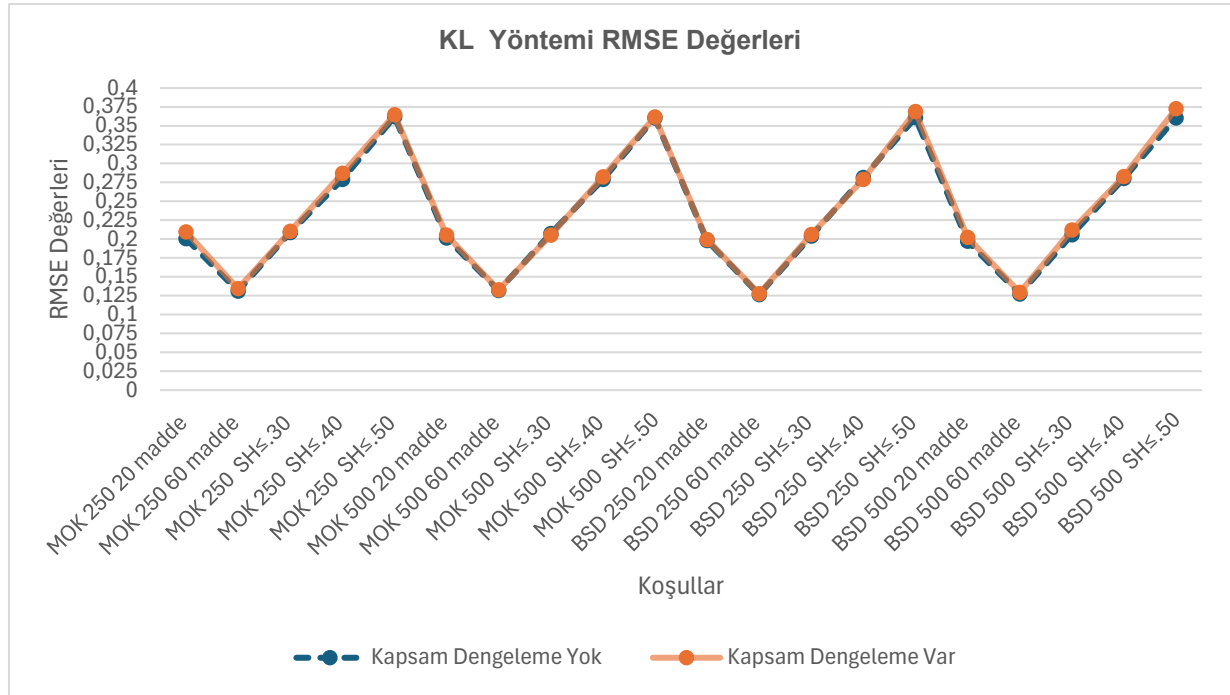
Şekil 18

Maksimum Fisher Bilgisinden Elde Edilen RMSE Değerleri



Şekil 19

Kullback Leibler Bilgisinden Elde Edilen RMSE Değerleri



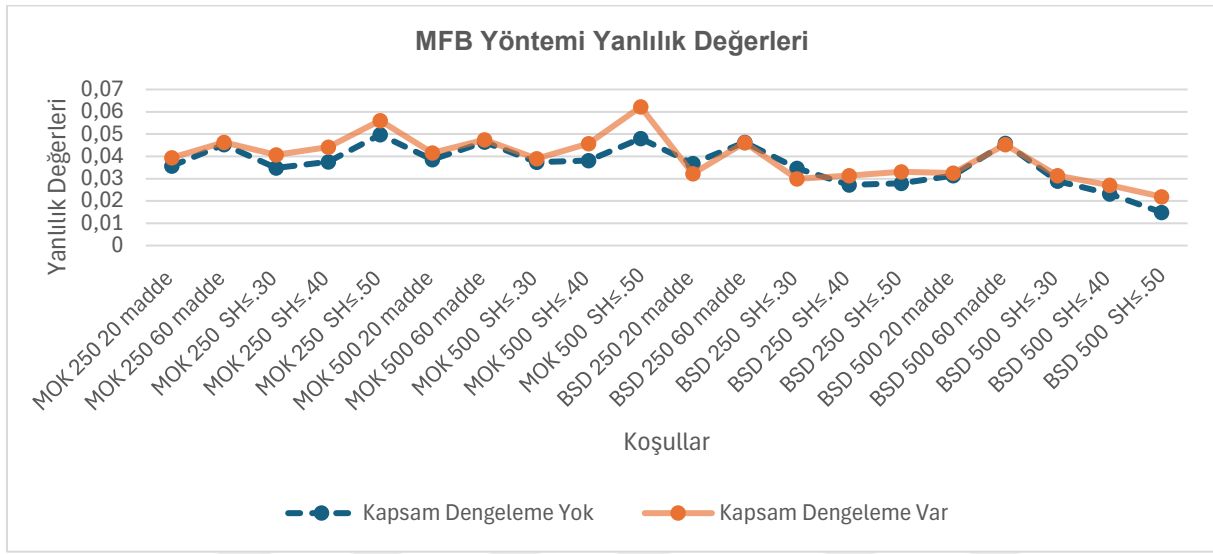
Farklı madde seçim yöntemlerinden elde edilen RMSE değerleri incelendiğinde Şekil 18 ve 19'da da görüldüğü gibi değerler birbirine oldukça yakındır. Her iki madde seçim yönteminde de en düşük RMSE değerleri testin 60 maddede sonlandırıldığı koşullarda elde edilmiştir. MFB yönteminde en düşük RMSE değeri (0.126); 500 kişilik örnekleme ve BSD yetenek kestirim yöntemi kullanıldığında, test 60 maddede sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek RMSE değeri (0.378) ise 500 kişilik örnekleme ve MOK yöntemi kullanıldığında, test SH≤.50 kuralına göre sonlandırılıp kapsam dengelemenin yapıldığı koşulda elde edilmiştir. KL madde seçim yönteminde de benzer şekilde, en düşük RMSE değeri (0.127); 250 kişilik örnekleme ve BSD kestirim yöntemi kullanıldığında, test 60 maddede sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek RMSE değeri (0.373) ise 500 kişilik örnekleme ve BSD kestirim yöntemi kullanıldığında, test SH≤.50 kuralına göre sonlandırılıp kapsam dengelemenin yapıldığı koşulda elde edilmiştir. Kapsam dengelemenin yapıldığı ve yapılmadığı koşullarda 500 kişilik örnekleme BSD kestirim yöntemi kullanılan tüm koşullarda KL madde seçim yönteminden

elde edilen RMSE değerleri MFB yönteminden elde edilen değerlere göre daha yüksek bulunmuştur.

Farklı madde seçim yöntemlerine ait yanlılık değerleri Şekil 20 ve 21'de verilmiştir.

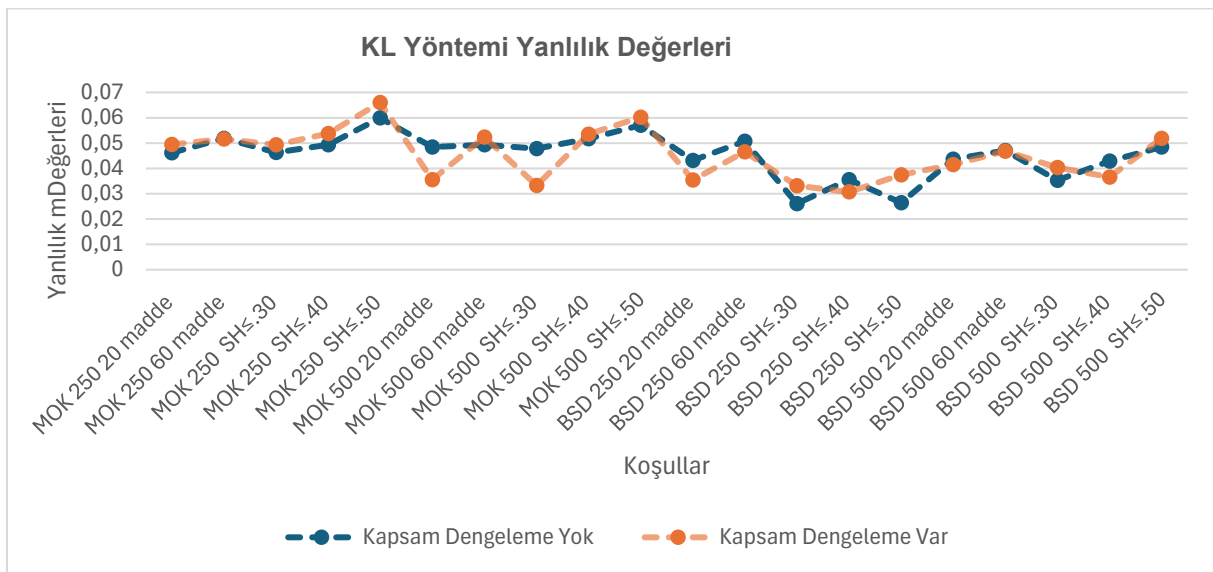
Şekil 20

Maksimum Fisher Bilgisinden Elde Edilen Yanlılık Değerleri



Şekil 21

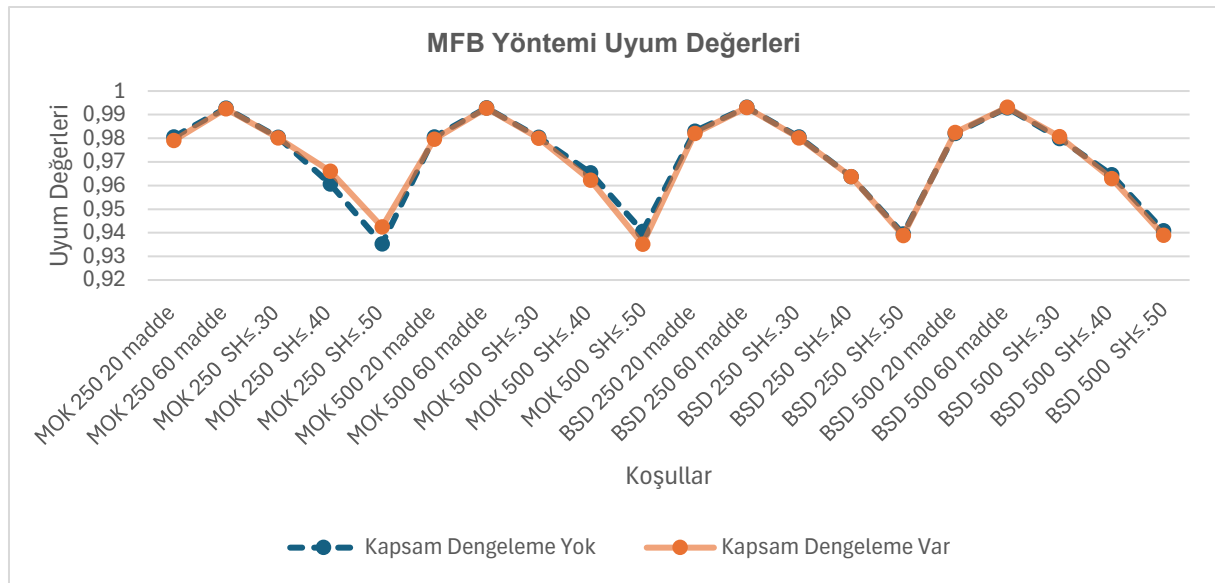
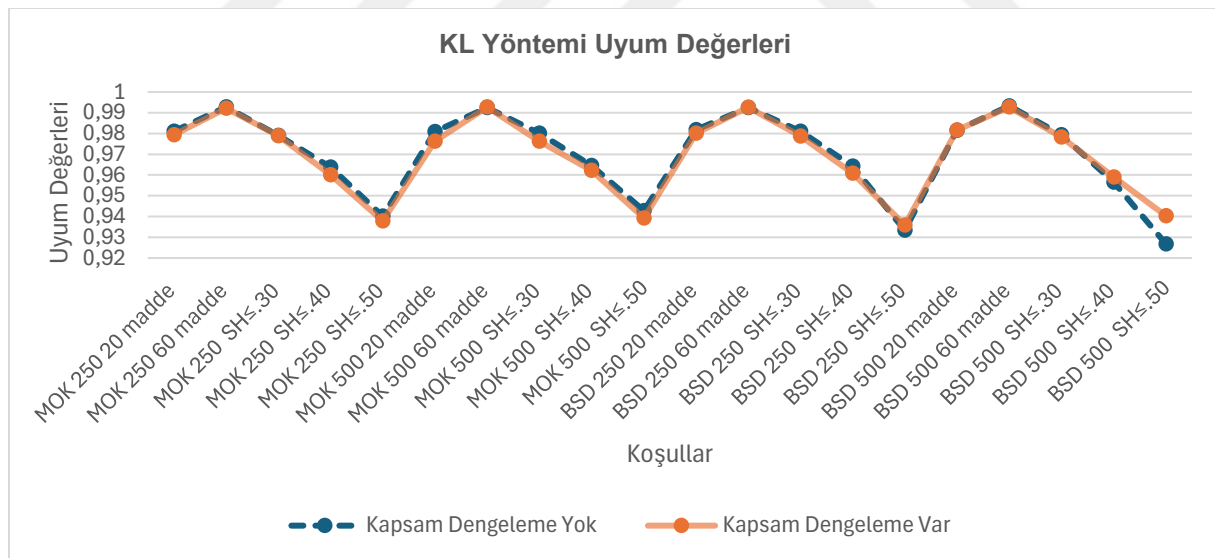
Kullback Leibler Bilgisinden Elde Edilen Yanlılık Değerleri



Farklı madde seçim yöntemlerinden elde edilen yanlılık değerleri incelendiğinde genel olarak hem kapsam dengeleme yapılan hem de yapılmayan koşullarda, MFB madde seçim

yöntemi kullanılan koşullarda elde edilen yanlılık değeri KL kestirim yöntemine göre daha düşük bulunmuştur. MFB madde seçim yönteminde en düşük yanlılık değeri (0.015); 500 kişilik örnekleme ve BSD kestirim yöntemi kullanıldığında, test $SH \leq 50$ kuralıyla sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek yanlılık değeri (0.062) ise yine 500 kişilik örneklem ve MOK yöntemi kullanıldığında, test $SH \leq 50$ kuralına göre sonlandırılıp kapsam dengelemenin yapıldığı koşulda elde edilmiştir. MFB madde seçim yönteminde MOK yöntemi kullanılan koşullarda kapsam dengelemenin yapılması yanlılığı az miktarda artırmışken BSD kullanılan koşulların neredeyse tamamında etkilememiştir. KL madde seçim yönteminde en düşük yanlılık değeri (0.026); 250 kişilik örnekleme ve BSD kestirim yöntemi kullanıldığında, test $SH \leq 30$ kuralıyla sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek yanlılık değeri (0.066) ise yine 250 kişilik örneklem ve MOK yöntemi kullanıldığında, test $SH \leq 50$ kuralına göre sonlandırılıp kapsam dengelemenin yapıldığı koşulda elde edilmiştir. KL madde seçim yöntemi ve ML kestirim yöntemi kullanıldığında kapsam dengeleme yapılması çoğu koşulda yanlılık değerini etkilemezken 500 kişilik örnekleme, 20 maddede ve $SH \leq 30$ durdurma kurallarının kullanıldığı koşullarda yanlılığı düşürmüştür. BSD kestirim yönteminin kullanıldığı koşullarda ise kapsam dengeleme yapılmasının yanlılık değerlerini etkilemediği söylenebilir, değerler birbirine oldukça yakındır.

Farklı madde seçim yöntemlerine ait uyum değerleri Şekil 22 ve 23'te verilmiştir.

Şekil 22**Maksimum Fisher Bilgisinden Elde Edilen Uyum Değerleri****Şekil 23****Kullback Leibler Bilgisinden Elde Edilen Uyum Değerleri**

Farklı madde seçim yöntemlerinden elde edilen uyum değerleri incelendiğinde Şekil 22 ve 23'te de görüldüğü gibi değerler birbirine oldukça yakındır. Her iki madde seçim yönteminde de en düşük uyum değerleri testin SH≤.50 kuralı ile sonlandırıldığı koşullarda elde edilmiştir. MFB yönteminde en düşük uyum değeri (0.935); 250 kişilik örnekleme ve MOK yetenek kestirim yöntemi kullanıldığında, test SH≤.50 kuralıyla sonlandırılıp kapsam dengelemenin

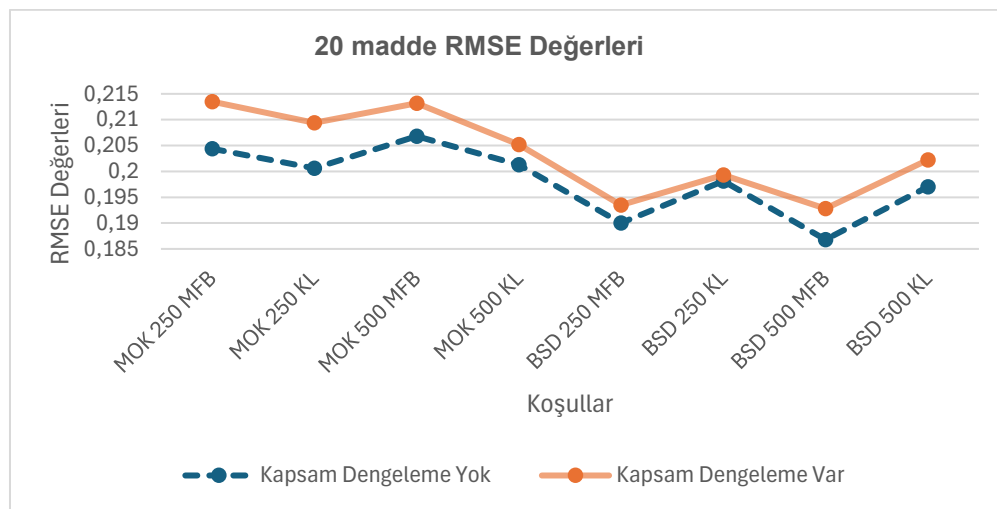
yapılmadığı koşulda elde edilmiştir. En yüksek uyum değeri (0.993) ise, kapsam dengeleme durumu fark etmeksizin, 250 kişilik örneklem ve BSD kestirim yöntemi kullanıldığında ve test 60 maddede sonlandırıldığı koşullarda elde edilmiştir. KL madde seçim yönteminde de benzer şekilde, en düşük uyum değeri (0.927); 500 kişilik örneklemde ve BSD kestirim yöntemi kullanıldığında, test $SH \leq 50$ kuralına göre sonlandırılıp kapsam dengeleme yapılmadığı koşulda elde edilmiştir. En yüksek uyum değeri (0.993) ise 500 kişilik örneklem ve BSD kestirim yöntemi kullanıldığında, test 60 madde ile sonlandırılıp kapsam dengelemenin yapılmadığı koşulda elde edilmiştir. Hem MFB hem de KL madde seçim yönteminde kapsam dengeleme yapılmasının uyum değerlerini etkilemediği söylenebilir uyum değerleri her iki durumda da oldukça yakın değerler almıştır.

Dördüncü Alt Probleme Ait Bulgular ve Yorumlar

Araştırmanın dördüncü alt probleminde bilgisayarda bireyselleştirilmiş test uygulamalarında, farklı sonlandırma kurallarında (20 maddede, 60 maddede, $SH \leq 30$, $SH \leq 40$ ve $SH \leq 50$) kapsam dengeleme yapılmasının ölçme kesinliği ve yetenek kestirimini nasıl etkilediği araştırılmıştır. Farklı test uzunluklarına ait RMSE değerleri Şekil 24 ve 25'te verilmiştir.

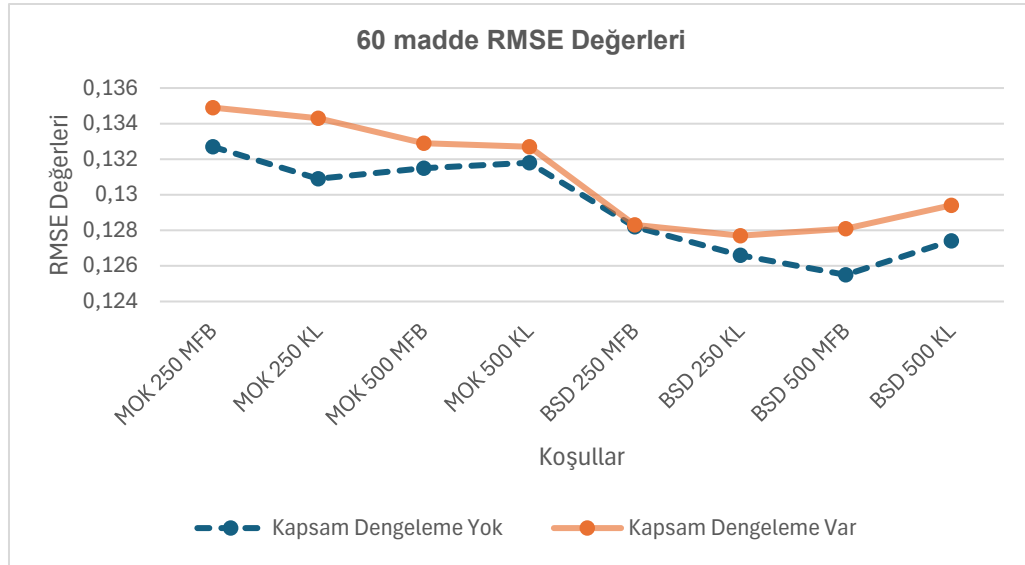
Şekil 24

Kısa Testten Elde Edilen RMSE Değerleri



Şekil 25

Uzun Testten Elde Edilen RMSE Değerleri

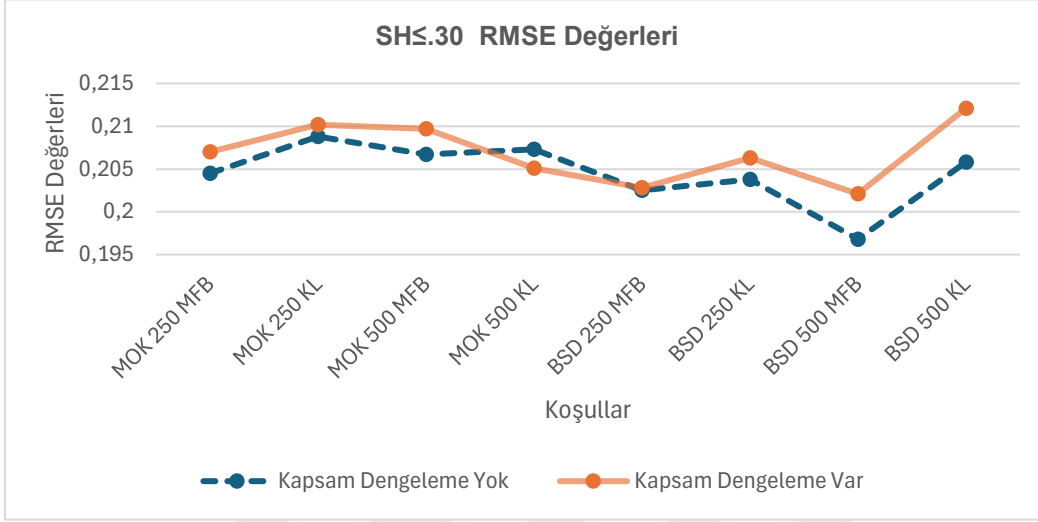


Testin 20 ve 60 maddede sonlandırıldığı koşullarda RMSE değeri kapsam dengeleme yapılan koşullarda, yapılmayan koşullara göre az da olsa yüksek bulunmuştur. Testin 20 maddede sonlandırıldığı koşullarda en yüksek RMSE değeri (0.214) 250 kişilik örnekleme, MOK yöntemi ve MFB madde seçim yöntemi kullanılan kapsam dengeleme yapılan durumda elde edilmiştir. En düşük RMSE değeri ise (0.187) 500 kişilik örnekleme BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan durumda elde edilmiştir. Kapsam dengeleme yapılan ve yapılmayan RMSE değerleri arasındaki fark oldukça düşük olmakla birlikte 0.001 – 0.009 arasındadır. Benzer şekilde, testin 60 maddede sonlandırıldığı koşullarda en yüksek RMSE değeri (0.135) 250 kişilik örnekleme, MOK yöntemi ve MFB madde seçim yöntemi kullanılan kapsam dengeleme yapılan durumda elde edilmiştir. En düşük RMSE değeri ise (0.126) 500 kişilik örnekleme BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan durumda elde edilmiştir. Kapsam dengeleme yapılan ve yapılmayan RMSE değerleri arasındaki fark, testin 20 maddede sonlandırıldığı koşullarda olduğu gibi, oldukça düşük olmakla birlikte 0.0001 – 0.003 arasındadır. Tüm koşullarda testin 20 madde ile sonlandırıldığı durumlarda RMSE değeri testin 60 maddede sonlandırıldığı koşullara göre daha yüksek bulunmuştur.

Farklı standart hatalara ait RMSE değerleri Şekil 26, 27 ve 28'de verilmiştir.

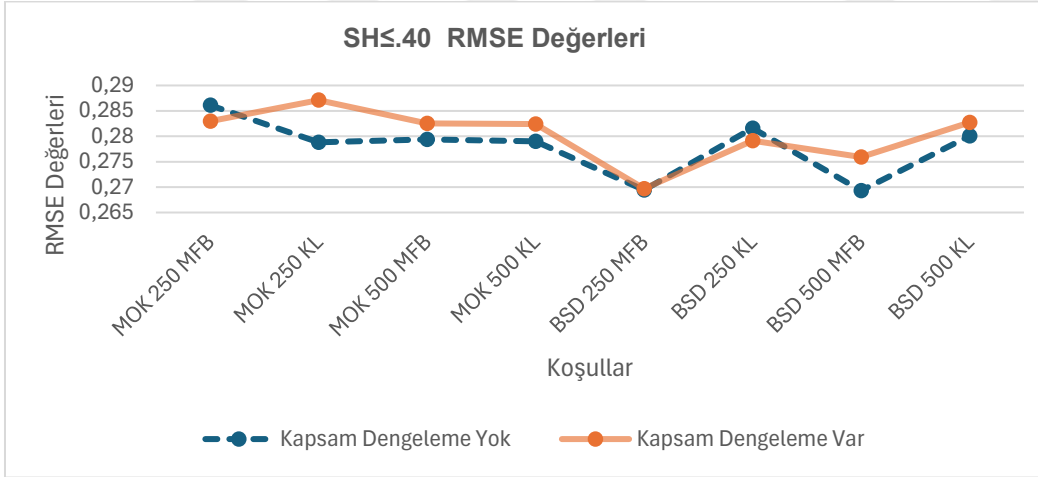
Şekil 26

SHs.30 Kuralından Elde Edilen RMSE Değerleri



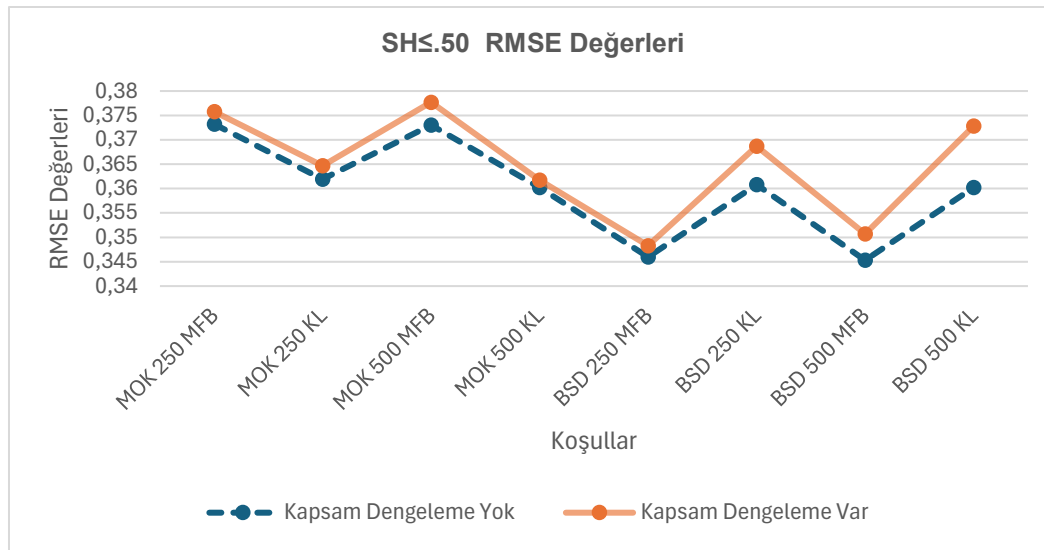
Şekil 27

SHs.40 Kuralından Elde Edilen RMSE Değerleri



Şekil 28

SH \leq .50 Kuralından Elde Edilen RMSE Değerleri



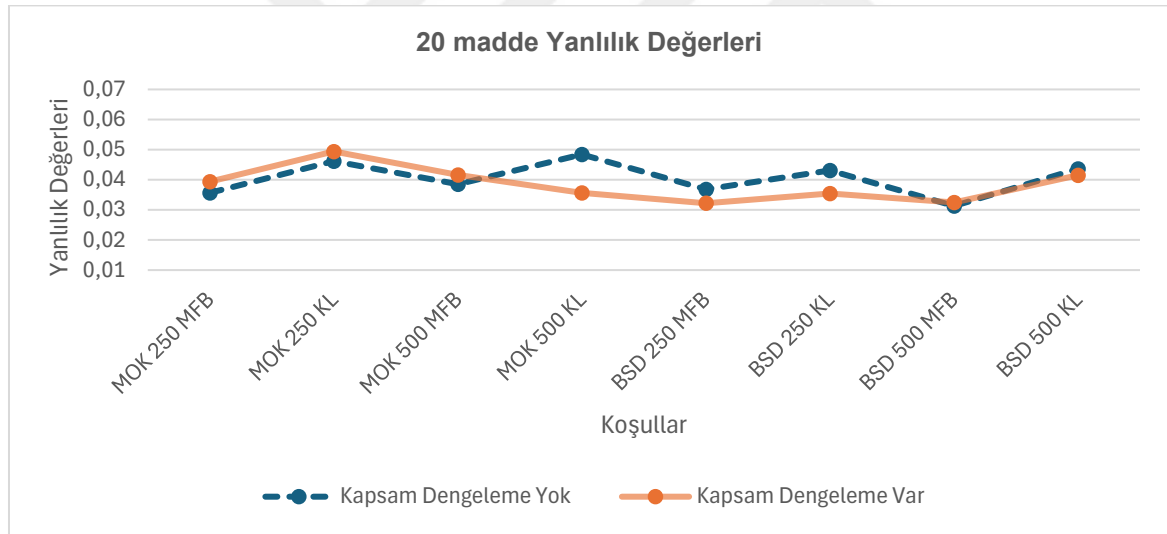
Çalışma kapsamında ele alınan üç ayrı standart hataya göre sonlandırılma koşulu göz önüne alındığında en düşük RMSE değerleri SH \leq .30; en yüksek RMSE değerleri ise SH \leq .50 sonlandırma kuralına ait koşullardan elde edilmiştir. SH \leq .30 sonlandırma kuralından elde edilen en yüksek RMSE değeri (0.212), 500 kişilik örnekleme, BSD kestirim yöntemi ve KL madde seçim yöntemi kullanılan kapsam dengeleme yapılan durumda elde edilmiştir. En düşük RMSE değeri ise (0.197) 500 kişilik örnekleme BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan durumda elde edilmiştir. SH \leq .30 kuralı ile test sonlandırıldığında, 500 kişilik örnekleme MOK yöntemi ve KL madde seçim yöntemi kullanıldığı koşul dışındaki tüm koşullarda, kapsam dengelemenin yapılmadığı durumda RMSE değeri daha düşük bulunmuştur. RMSE değerleri arasındaki fark oldukça düşük olmakla birlikte 0.0003 – 0.006 arasındadır. SH \leq .40 sonlandırma kuralından elde edilen en yüksek RMSE değeri (0.287), 250 kişilik örnekleme, MOK yöntemi ve KL madde seçim yöntemi kullanılan kapsam dengeleme yapılan durumda elde edilmiştir. En düşük RMSE değeri ise (0.269), 500 kişilik örnekleme BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan durumda elde edilmiştir. Genel olarak testin SH \leq .40 kuralı ile sonlandırıldığı koşulların çoğunda kapsam dengelemenin yapılmadığı durumda RMSE değeri daha düşük bulunmuştur. RMSE değerleri arasındaki fark oldukça düşük

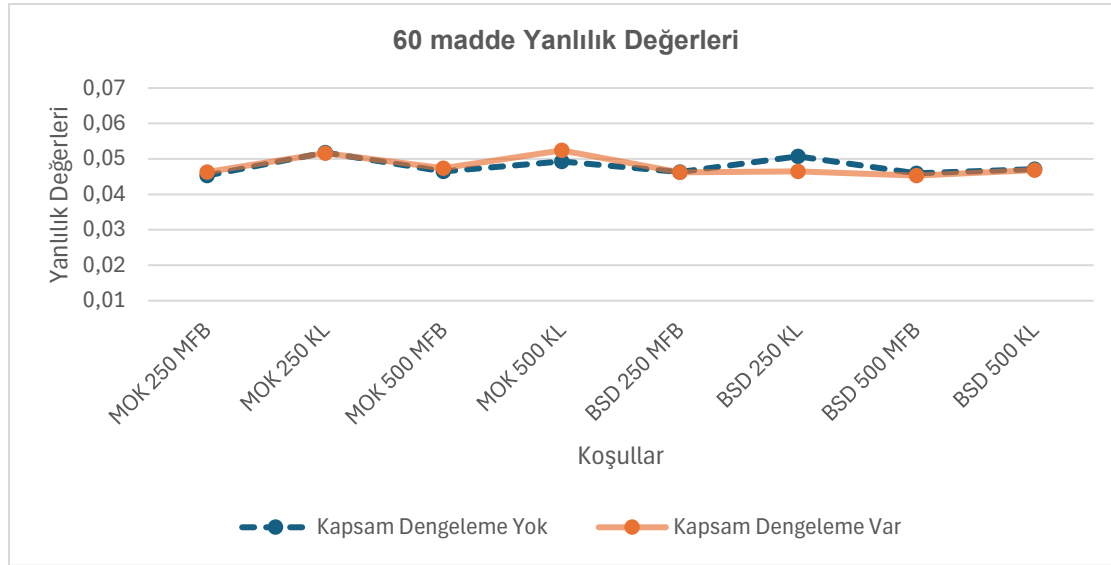
olmakla birlikte 0.0003 – 0.008 arasındadır. $SH \leq 50$ sonlandırma kuralından elde edilen en yüksek RMSE değeri (0.378), 500 kişilik örnekleme, MOK yöntemi ve MFB madde seçim yöntemi kullanılan kapsam dengeleme yapılan durumda elde edilmiştir. En düşük RMSE değeri ise (0.345), 500 kişilik örnekleme BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan durumda elde edilmiştir. Testin $SH \leq 50$ kuralı ile sonlandırıldığı koşulların tamamında kapsam dengelemenin yapılmadığı durumda RMSE değeri daha düşük bulunmuştur. RMSE değerleri arasındaki fark oldukça düşük olmakla birlikte 0.002 – 0.01 arasındadır.

Farklı test uzunluklarına ait yanlılık değerleri Şekil 29 ve 30'da verilmiştir.

Şekil 29

Kısa Testten Elde Edilen Yanlılık Değerleri



Şekil 30**Uzun Testten Elde Edilen Yanlılık Değerleri**

Tüm koşullar göz önünde bulundurulduğunda yanlılık değerleri testin 60 maddede sonlandırıldığı koşullarda 20 maddede sonlandırıldığı koşullara göre daha yüksek bulunmuştur.

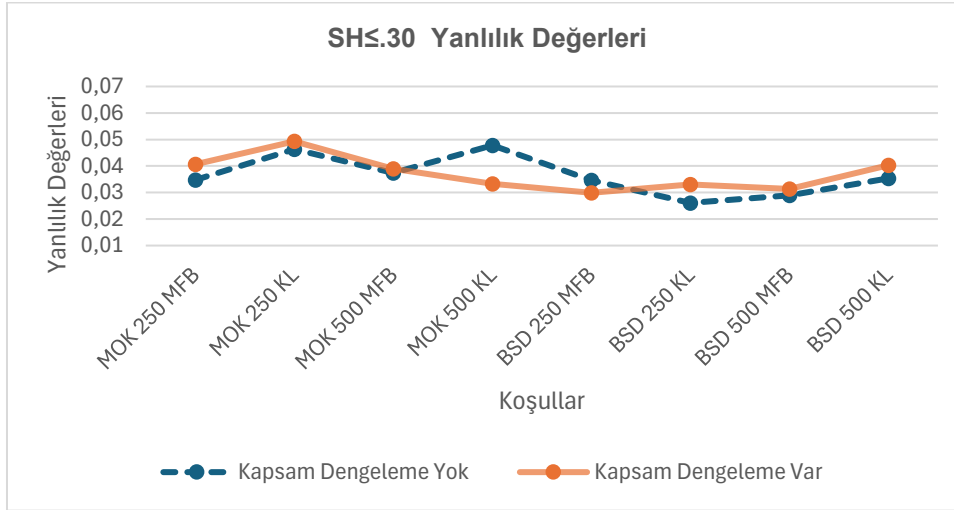
Her iki test uzunluğunda da BSD kestirim yöntemi kullanılan koşulların neredeyse tamamında kapsam dengeleme yapılması yanlılık değerini az miktarda düşürmüştür. Bunun yanında MOK yöntemi ve KL madde seçim yönteminin kullanıldığı; 500 kişilik örnekleme testin 20 maddede sonlandırıldığı ve 250 kişilik örnekleme testin 60 maddede sonlandırıldığı koşullarda kapsam dengeleme yapılması yanlılık değerini bir miktar düşürmüştür. Testin 20 maddede sonlandırıldığı koşullarda en yüksek yanlılık değeri (0.049) 250 kişilik örnekleme, ML kestirim yöntemi ve KL madde seçim yöntemi kullanılan kapsam dengeleme yapılan durumda elde edilmiştir. En düşük yanlılık değeri ise (0.031) 500 kişilik örnekleme BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanılan, kapsam dengeleme yapılmayan durumda elde edilmiştir. Testin 60 maddede sonlandırıldığı koşullarda en yüksek yanlılık değeri (0.052) 500 kişilik örnekleme, MOK yöntemi ve KL madde seçim yöntemi kullanılan kapsam dengeleme yapılan durumda elde edilmiştir. En düşük yanlılık değeri ise (0.126) 500

kişilik örnekleme BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanılan, kapsam dengeleme yapılan durumda elde edilmiştir.

Farklı standart hatalara ait yanlılık değerleri Şekil 31, 32 ve 33'te verilmiştir.

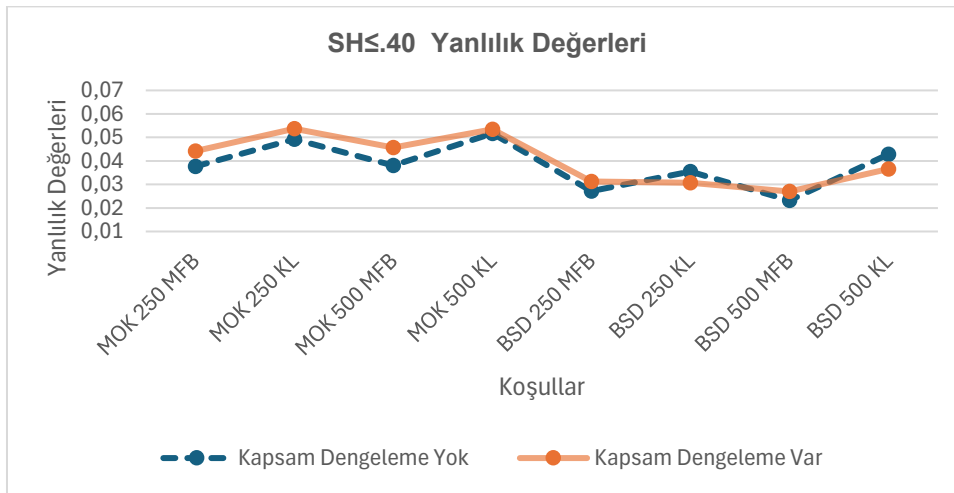
Şekil 31

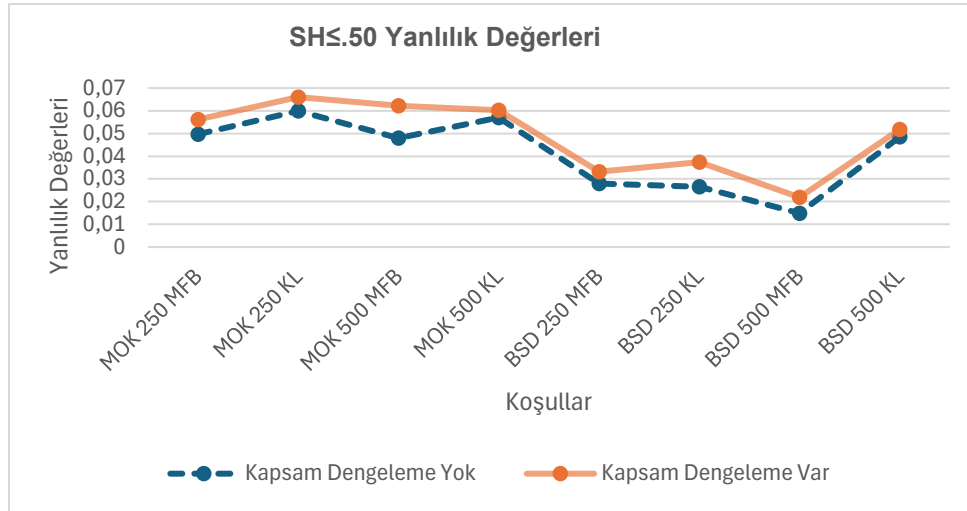
SHs.30 Kuralından Elde Edilen Yanlılık Değerleri



Şekil 32

SHs.40 Kuralından Elde Edilen Yanlılık Değerleri



Şekil 33**SH \leq .50 Kuralından Elde Edilen Yanlılık Değerleri**

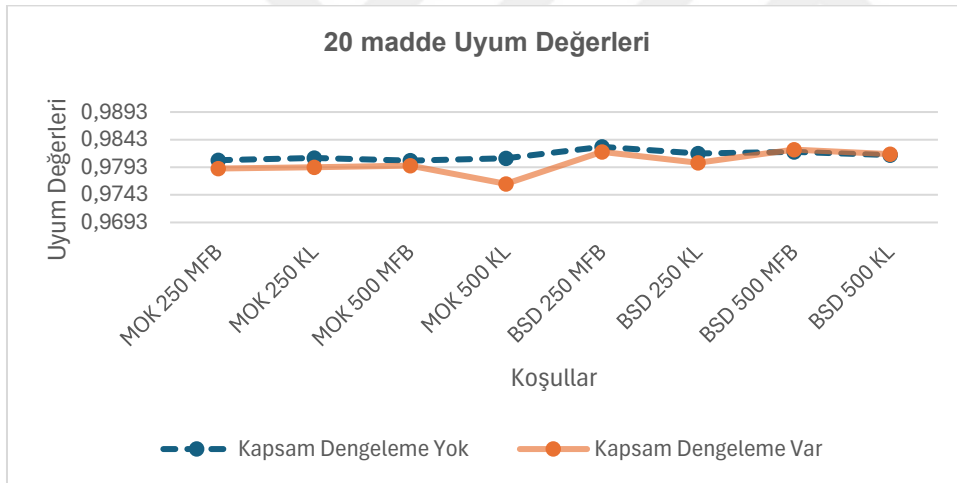
Üç ayrı standart hataya göre sonlandırılma koşulu göz önüne alındığında MOK yöntemi kullanılan koşullarda, kapsam dengeleme durumu fark etmeksizin, en yüksek yanlılık değerleri SH \leq .50, en düşük yanlılık değerleri ise SH \leq .30 sonlandırma kuralı tercih edildiğinde elde edilmiştir. Kapsam dengeleme yapılan ve yapılmayan koşullardan elde edilen yanlılık değerleri birbirine oldukça yakındır. SH \leq .50 sonlandırma kuralı tercih edildiğinde kapsam dengeleme yapılması tüm koşullarda yanlılığı az miktarda (0.003-0.01) artırmıştır. SH \leq .40 sonlandırma kuralı tercih edildiğinde BSD kesrim yöntemi ve KL madde seçim yöntemi kullanılan koşullar dışında kapsam dengeleme yapılması yanlılığı bir miktar (0.002-0.008) artırmıştır. SH \leq .30 sonlandırma kuralı tercih edildiğinde ise 500 kişilik örnekleme MOK kestirim yöntemi ve KL madde seçim yöntemi ile 250 kişilik örnekleme BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanılan koşullar dışında yine kapsam dengeleme yapılması yanlılık değerini bir miktar (0.002-0.007) artırmıştır. SH \leq .30 sonlandırma kuralından elde edilen en yüksek yanlılık değeri (0.049), 250 kişilik örnekleme, MOK yöntemi ve KL madde seçim yöntemi kullanılan kapsam dengeleme yapılan durumda elde edilmiştir. En düşük yanlılık değeri ise (0.026) 250 kişilik örnekleme BSD kestirim yöntemi ve KL madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan durumda elde edilmiştir. SH \leq .40 sonlandırma kuralından elde edilen en yüksek yanlılık değeri (0.054), 250 kişilik örnekleme, MOK yöntemi ve KL madde seçim

yöntemi kullanılan kapsam dengeleme yapılan durumda elde edilmiştir. En düşük yanlılık değeri ise (0.023), 500 kişilik örnekleme BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan durumda elde edilmiştir. Benzer şekilde $SH \leq 50$ sonlandırma kuralından elde edilen en yüksek yanlılık değeri yine (0.066), 250 kişilik örnekleme, MOK yöntemi ve KL madde seçim yöntemi kullanılan kapsam dengeleme yapılan durumda elde edilmiştir. En düşük yanlılık değeri ise (0.015), 500 kişilik örnekleme BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan durumda elde edilmiştir.

Farklı standart hatalara ait uyum değerleri Şekil 34 ve 35'te verilmiştir.

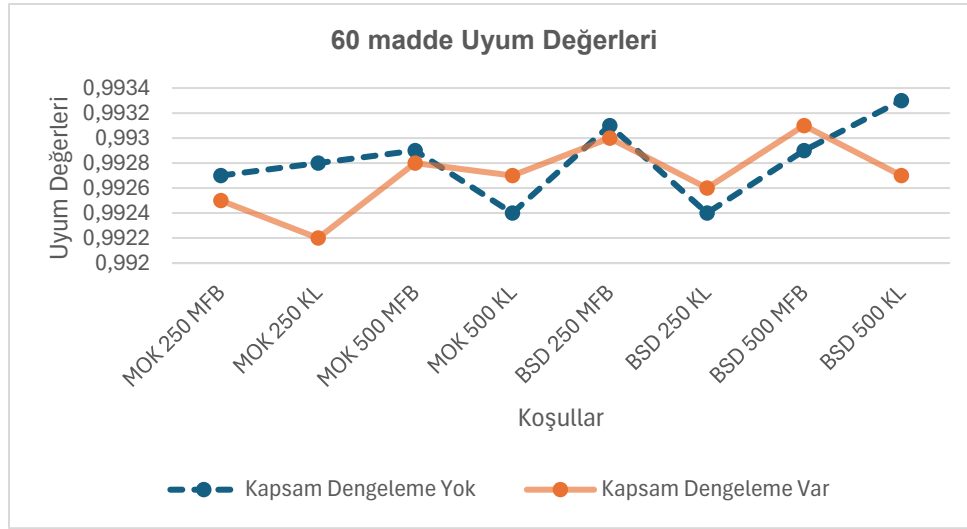
Şekil 34

Kısa Testten Elde Edilen Uyum Değerleri



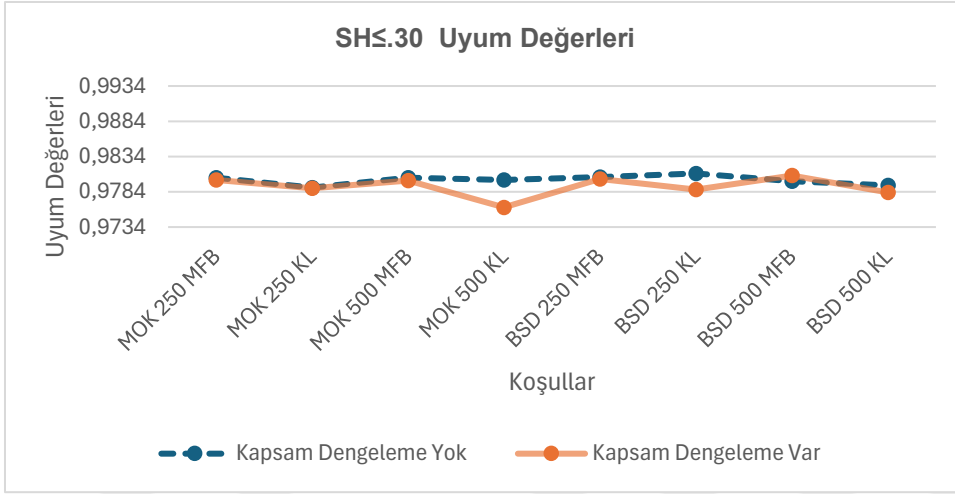
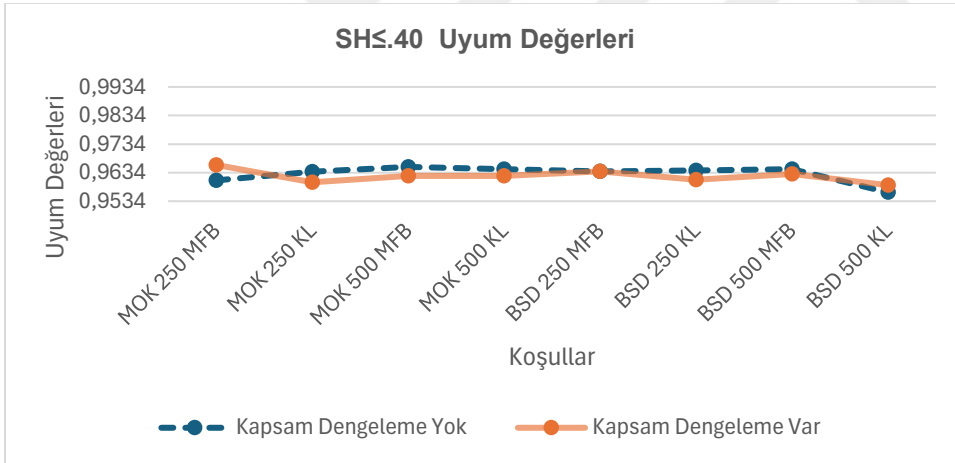
Şekil 35

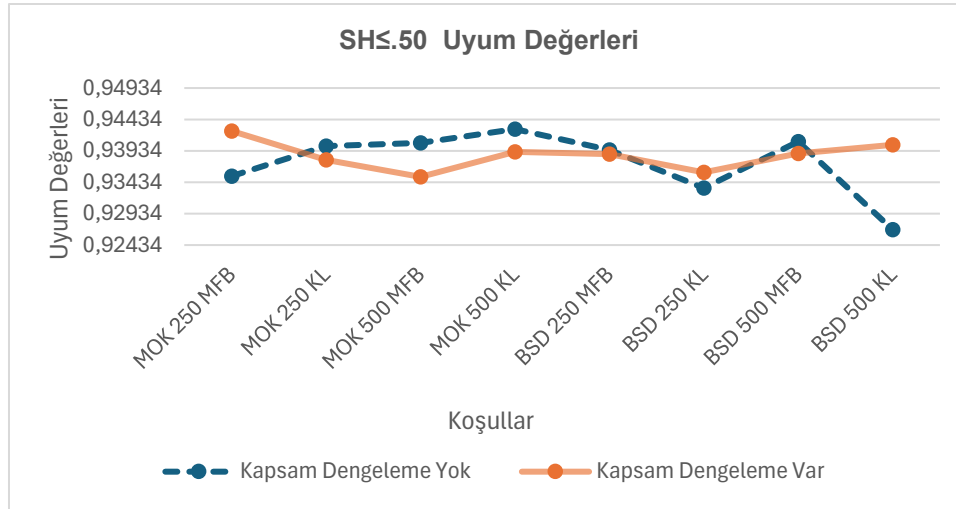
Uzun Testten Elde Edilen Uyum Değerleri



Şekilde de görüldüğü gibi genel olarak her iki test uzunluğunda da kapsam dengeleme yapılmasının uyum değerini etkilemediği söylenebilir. Testin 20 maddede sonlandırıldığı koşulda 500 kişilik örnekleme BSD kestirim yöntemi kullanılan koşullar dışında kapsam dengeleme yapılmaması uyum değerini az da olsa (0-0.005) artırmıştır. Testin 60 maddede sonlandırıldığı koşullarda ise kapsam dengeleme yapılan ve yapılmayan her iki duruma ait koşullardan elde edilen uyum değerleri neredeyse aynıdır ve oldukça yüksek bulunmuştur. Testin 20 maddede sonlandırıldığı koşullarda en yüksek uyum değeri (0.983) 250 kişilik örnekleme, BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan durumda elde edilmiştir. En düşük uyum değeri ise (0.976) 500 kişilik örnekleme MOK yöntemi ve KL madde seçim yöntemi kullanılan, kapsam dengeleme yapılan durumda elde edilmiştir. Testin 60 maddede sonlandırıldığı koşullarda en yüksek uyum değeri (0.993) 500 kişilik örnekleme, BSD kestirim yöntemi ve KL madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan durumda elde edilmiştir. En düşük uyum değeri ise (0.992) 250 kişilik örnekleme MOK yöntemi ve KL madde seçim yöntemi kullanılan, kapsam dengeleme yapılan durumda elde edilmiştir.

Farklı standart hatalara ait uyum değerleri Şekil 36,37 ve 38'de verilmiştir.

Şekil 36**SHİ.30 Kuralından Elde Edilen Uyum Değerleri****Şekil 37****SHİ.40 Kuralından Elde Edilen Uyum Değerleri**

Şekil 38**SH \leq .50 Kuralından Elde Edilen Uyum Değerleri**

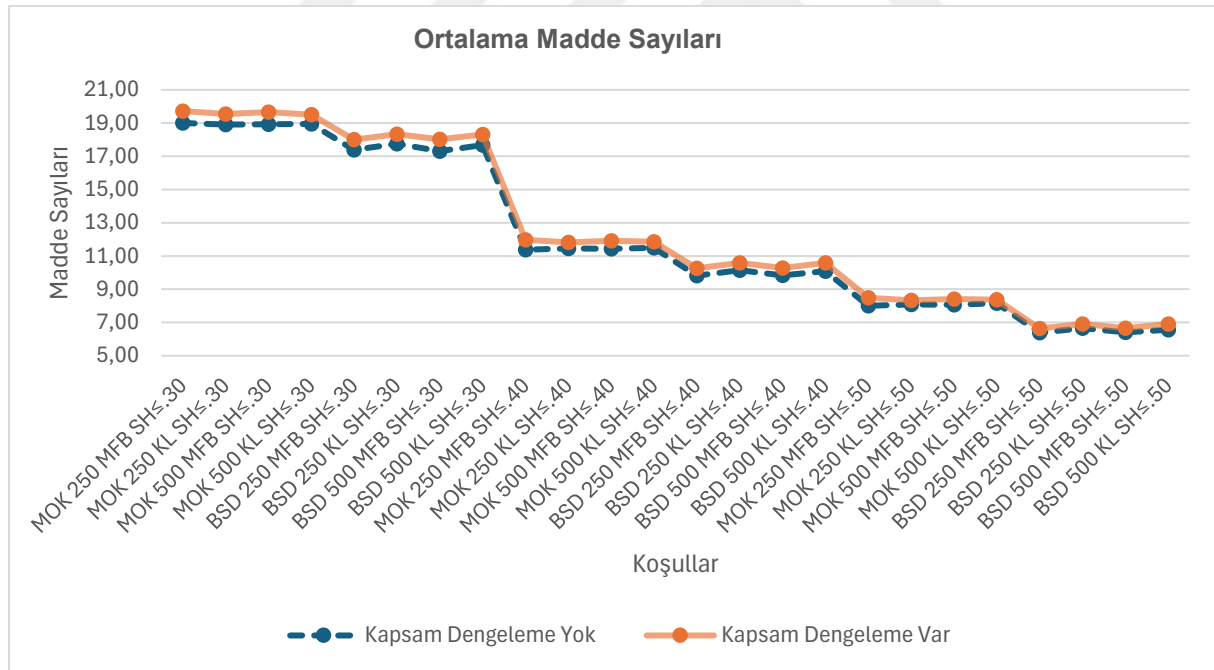
Şekilde görüldüğü gibi genel olarak her iki test uzunluğunda da kapsam dengeleme yapılmasının uyum değerini etkilemediği söylenebilir. Sonlandırma kurallarına ait koşulların her birinde kapsam dengeleme yapılan ve yapılmayan durumlardan elde edilen uyum değerleri birbirine oldukça yakındır. Üç ayrı standart hataya göre sonlandırılma koşulu göz önüne alındığında, kapsam dengeleme durumu fark etmeksizin, en yüksek uyum değerleri SH \leq .30, en düşük uyum değerleri ise SH \leq .50 sonlandırma kuralı tercih edildiğinde elde edilmiştir. SH \leq .30 sonlandırma kuralında kapsam dengeleme yapılan ve yapılmayan durumlardan elde edilen uyum değeri farkı oldukça düşüktür (0-0.004). En yüksek uyum değeri (0.981) 250 kişilik örnekleme, BSD kestirim yöntemi ve KL madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan durumda elde edilmiştir. En düşük uyum değeri ise (0.976) 500 kişilik örnekleme MOK yöntemi ve KL madde seçim yöntemi kullanılan, kapsam dengeleme yapılan durumda elde edilmiştir. Benzer şekilde SH \leq .40 sonlandırma kuralında kapsam dengeleme yapılan ve yapılmayan durumlardan elde edilen uyum değeri farkı oldukça düşüktür (0-0.005). En yüksek uyum değeri (0.966) 250 kişilik örnekleme, MOK yöntemi ve MFB madde seçim yöntemi kullanılan kapsam dengeleme yapılan durumda elde edilmiştir. En düşük uyum değeri ise (0.957) 500 kişilik örnekleme BSD kestirim yöntemi ve KL madde seçim yöntemi kullanılan, kapsam dengeleme yapılmayan durumda elde edilmiştir. Testin SH \leq .50 kuralıyla

sonlandırıldığı koşullarda da kapsam dengeleme yapılan ve yapılmayan durumlardan elde edilen uyum katsayıları arasındaki farkın oldukça küçük olmasının yanında en büyük fark (0.01), kapsam dengeleme yapılan durum lehine, 500 kişilik örnekleme BSD kestirim yöntemi ve KL madde seçim yöntemi kullanılan koşulda ortaya çıkmıştır. $SH \leq 50$ kuralıyla testin sonlandırıldığı koşullarda en yüksek uyum değeri (0.943) 500 kişilik örnekleme MOK yöntemi ve KL madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan koşulda elde edilmiştir. En düşük uyum değeri (0.927) ise 500 kişilik örnekleme BSD kestirim yöntemi ve KL madde seçim yöntemi kullanılan kapsam dengeleme yapılmayan koşulda elde edilmiştir.

Aşağıda Şekil 39'da standart hataya göre durdurma kuralı tercih edilen koşullara ilişkin ortalama madde sayıları verilmiştir.

Şekil 39

Farklı Standart Hata Değerlerine Ait Ortalama Madde Sayıları



Şekilde de görüldüğü gibi kapsam dengeleme yapılması madde sayısını tüm koşullarda bir maddeden daha az (0.22-0.73) olacak şekilde artırmıştır. Kapsam dengelemenin standart hataya göre durdurma kuralı uygulandığında madde sayısını dikkate değer şekilde artırmadığı görülmüştür. En yüksek ortalama madde sayısı (19.73), $SH \leq 30$ kuralıyla testin sonlandırıldığı

250 kişilik örnekleme, MOK kestirim yöntemi ve MFB madde seçim yöntemi kullanılıp kapsam dengeleme yapılan koşulda elde edilmiştir. En düşük ortalama madde sayısı ise (6.38), $SH \leq 50$ kuralıyla testin sonlandırıldığı, 250 kişilik örnekleme, BSD kestirim yöntemi ve MFB madde seçim yöntemi kullanılıp kapsam dengeleme yapılmayan koşulda elde edilmiştir. Üç farklı sonlandırma kuralında da kapsam dengeleme durumu fark etmeksizin BSD kestirim yöntemi kullanıldığında ortalama madde sayılarında düşüş görülmüştür. Standart hataya göre sonlandırma kurallarının tamamında kapsam dengeleme yapılan koşullarda MOK yöntemi tercih edildiğinde; KL madde seçim yönteminde MFB'ye göre ortalama madde sayıları daha düşük bulunmuştur. Aynı koşullarda BSD kestirim yöntemi tercih edildiğinde ise tam aksine KL madde seçim yöntemi kullanılan koşullarda ortalama madde sayısı artış göstermiştir. Kapsam dengeleme yapılmayan koşulların neredeyse tamamında, yetenek kestirim yöntemi fark etmeksizin, KL madde seçim yönteminin tercih edilmesi ortalama madde sayısını artırmıştır. Örneklem büyüklüğünün ise çalışma kapsamında ele alınan farklı koşullarda ortalama madde sayısını etkilemediği bulunmuştur.

Bu çalışmada simülasyon tabanlı metodolojiler kullanılarak kapsam dengelemenin farklı örneklem büyüklükleri, madde seçim yöntemleri, yetenek kestirim yöntemleri ve test sonlandırma kuralları altında tahmin edilen ve gerçek yetenek düzeyleri arasındaki korelasyon, RMSE ve yanlılık gibi temel ölçütleri nasıl etkilediğinin incelenmesi amaçlanmıştır. Çalışmanın bulguları, BBT alanında yapılan ve özellikle kapsam dengelemenin ölçme doğruluğu ve test verimliliği üzerindeki etkilerini inceleyen Yasuda ve Hull (2021), Leung ve arkadaşları (2003), Özer Taymur (2023), Sari ve Manley (2017) tarafından yapılan çalışmalarda rapor edilen bulgular ile büyük ölçüde örtüşmektedir.

Yasuda ve Hull (2021), Force Concept Inventory (FCI) için BBT uygulamasında kapsam dengelemenin dâhil edilmesinin, özellikle test uzunluğunun kısa olduğu durumlarda, RMSE değerinde hafif bir artışa neden olduğunu bildirmiştir. Bu artış 13 maddeden uzun testler için %1'den az olarak rapor edilmiştir. Benzer şekilde Özer Taymur (2023) madde havuzunda pratik sınırlandırmalardan olan kapsam dengeleme yapılmasının ölçme kesinliği kestirim

değerlerinden RMSE değerlerinde az da olsa bir artışa neden olduğunu raporlamıştır. Bu bulgu, kapsam dengelemenin RMSE ve yanlılıkta küçük bir artışa neden olduğunun gözlemlendiği çalışma sonuçları ile oldukça tutarlıdır. Her iki çalışma da kapsam dengelemenin belirli bir hata payı getirdiğini kabul etmekte, ancak test uzunluğu arttıkça bu etkinin azaldığını vurgulamaktadır. Yasuda ve arkadaşları, test uzunluğu 13 maddeden fazla olduğunda, kapsam dengelemenin RMSE üzerindeki artışının ihmal edilebilir düzeyde olduğunu belirtmiştir. Bu sonuç, daha uzun testlerin, kapsam dengelemenin ölçme doğruluğu üzerindeki olumsuz etkilerini azalttığına dair elde edilen bulgularla örtüşmektedir. Benzer şekilde, Leung ve arkadaşları (2003), üç kapsam dengeleme yöntemini KBBT, MKBBT ve MÇM karşılaştıran bir çalışma gerçekleştirmiş ve kapsam dengeleme yönteminin ölçme verimliliği üzerinde sistematik bir etkisi olmamakla birlikte, MÇM gibi bazı yöntemlerin madde kullanım sıklığı oranını azaltmada ve madde havuzunun dengeli kullanımını sağlamada daha etkili olduğunu bulmuştur. Yapılan çalışma da özellikle örneklem büyüklüğü ve madde seçme yöntemleri bağlamında bu bulguları doğrulamaktadır. Leung ve arkadaşları gibi kapsam dengelemenin ölçme verimliliğini etkileyebileceği, ancak uygun yöntemler kullanıldığında bu etkinin nispeten küçük olduğu tespit edilmiştir. Sari ve Manley'nin (2017) çalışmaları, kapsam dengelemenin, yetenek tahminlerinde hafif sapmalara neden olma potansiyeline rağmen, testin farklı içerik alanları arasında geçerliğini korumak için kritik bir öneme sahip olduğunu göstererek bu çalışmanın bulgularını desteklemektedir. Söz konusu çalışmalar, tüm içerik alanlarının bireyselleştirilmiş testlerde yeterince temsil edilmesini sağlamanın önemini vurgulamaktadır; bu ilke mevcut çalışmanın da temelini oluşturmaktadır. Çalışmada kapsam dengelemenin tahmin edilen ve gerçek yetenek düzeyleri arasındaki korelasyon üzerinde minimal bir etkiye sahip olduğu gösterilmiş olup, Sari ve Manley'nin kapsam dengelemenin özellikle yüksek riskli testlerde vazgeçilmez olduğunu savunan sonuçlarıyla örtüşmektedir. Özer Taymur (2023) ve Bulut ve Kan (2012) tarafından yapılan çalışmaların, yetenek kestirim yöntemi olarak BSD yönteminin kullanılmasının ölçme kesinliği kestirim değerlerinden olan RMSE ve yanlılık değerlerini düşürebileceği bulgusu da mevcut çalışmadan elde edilen bulgularla benzerlik taşımaktadır.

Yapılan araştırma, BBT alanındaki bahsi geçen temel çalışmalarla birçok önemli bulguyu paylaşmaktadır. Bu çalışmaların genel olarak ortaya koyduğu gözlem, kapsam dengelemenin RMSE, yanlılık ve test verimliliği açısından hafif düzeyde kayıplara neden olsa da bu etkilerin genellikle oldukça hafif olduğu ve dikkatli bir madde seçim yöntemi ile uygun test tasarımıyla hafifletilebileceğidir. Yasuda ve arkadaşları (2021), Leung ve arkadaşları (2003), Sari ve Manley (2017), Özer Taymur'un (2023) bulgularıyla bu sonuçların örtüşmesi, kapsam dengelemenin BBT'de kapsam geçerliğini sağlamada etkili ve güvenilir bir strateji olduğunu, ölçme doğruluğunu ise önemli ölçüde tehlikeye atmadığını güçlendirmektedir.

Bu çalışmada elde edilen bulgular, bireyselleştirilmiş testlerde kapsam dengelemenin test adilliği ve geçerlik açısından kritik bir role sahip olduğunu göstermektedir. Literatürdeki diğer çalışmalarla tutarlılık içinde, kapsam dengelemenin özellikle testin tüm içerik alanlarını eşit bir şekilde temsil etmesini sağladığı ve bu şekilde test adilliği ilkesine hizmet ettiği görülmektedir. Ancak, kapsam geçerliğinin düşük olduğu bir testin güvenilirliğinin yüksek olmasının, doğru ve adil ölçme sonuçları üretmesi açısından yeterli olmadığı açıktır. Geçerliği düşük bir test, güvenilirliği yüksek olsa bile ölçmek istediği yeteneği tam olarak temsil edemediğinde karar süreçlerini olumsuz etkileyebilir. Geçerlik ve güvenilirlik arasındaki bu hassas denge, yüksek riskli uygulamalarda testin adil ve doğru sonuçlar sunabilmesi için özel bir önem taşımaktadır. Bu çalışmanın sonuçları, literatürde kapsam dengelemenin ölçme kesinliği üzerindeki etkilerinin genellikle minimal olduğu ve uygun yöntemler kullanıldığında bu etkinin daha da hafifletilebileceği yönündeki bulgularla uyumludur. Mevcut çalışmada yanlılık, RMSE ve uyum değerlerinin birbirine oldukça yakın olduğu bulunmuştur. Bunun temel nedeni, simülasyon ortamında kontrol edilen koşullar ve ideal madde parametrelerinin kullanılmasıdır. Ancak gerçek veri ile çalışıldığında test uygulama koşullarındaki farklılıklar, testin uygulandığı örneklemin heterojen yapısı ve madde havuzundaki olası sınırlamalar nedeniyle bu değerler arasında daha belirgin farklar görülebilir.

Bölüm 5

Sonuç ve Öneriler

Bu bölümde araştırmanın bulgularına ilişkin sonuçlara ve önerilere yer verilmiştir.

Sonuçlar

Bu araştırmada BBT uygulamalarında kapsam dengeleme yapılan ve yapılmayan durumlarda farklı örneklem büyüklükleri, madde seçim yöntemleri, yetenek kestirim yöntemleri ve sonlandırma kurallarının birlikte kullanılmasının bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanlılık ve RMSE değerleri üzerindeki etkisinin incelenmesi amaçlanmıştır. Çalışma kapsamında oluşturulan 80 koşul için inceleme yapılmış ve araştırmanın dört alt problemine cevap aranmıştır. Araştırmanın sonuçları aşağıda yer almaktadır.

- Kapsam dengeleme, testte tüm alanların veya içeriklerin yeterli şekilde temsil edilmesini sağlamak amacıyla tasarlanmış olmasına rağmen, genel olarak ortalama hata ve yanlılık değerlerinde hafif artışlara neden olmuştur. Bu artışlar oldukça küçük bulunmuştur ve yetenek kestirimlerinin doğruluğunu önemli ölçüde etkilemeyeceği söylenebilir. Bu durum, farklı test koşullarında tahmin edilen ve gerçek yetenek düzeyleri arasındaki yüksek korelasyon katsayılarından anlaşılmaktadır.
- Her iki örneklem büyüklüğünde de BSD yetenek kestirim yönteminin MOK yöntemine göre ortalama hata değeri daha düşüktür.
- Küçük örnekleme kapsam dengeleme yapılan ve yapılmayan koşullarda yanlılık değerinin farklılaşmadığı görülmüştür. Geniş örnekleme ise MOK yöntemi ve KL madde seçim yöntemi, $SH \leq 30$ ve 20 madde de sonlandırma kuralı uygulandığında kapsam dengeleme yapılması yanlılık değerini düşürmüştür.
- Her iki örneklem büyüklüğünde de kapsam dengeleme yapılması bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon katsayılarında fark yaratmamıştır.

- Tüm koşullar göz önünde bulundurulduğunda bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki en yüksek korelasyon ve en düşük ortalama hata değerleri testin 60 maddede sonlandırıldığı koşullardan elde edilirken en düşük korelasyon ve en yüksek ortalama hata değerleri testin $SH \leq 50$ kuralıyla sonlandırıldığı koşullardan elde edilmiştir.
- Geniş örnekleme BSD yetenek kestirim yönteminin kullanılması; MOK yöntemi ile birlikte MFB madde seçim yönteminin kullanılması ve $SH \leq 50$ kuralıyla testin sonlandırılması kapsam dengelemenin ortalama hata değeri üzerindeki etkisini artırmıştır.
- Geniş örnekleme BSD kestirim yöntemi ile birlikte KL madde seçim yönteminin kullanılması ortalama hata değerlerinde artışa neden olmuştur.
- Kapsam dengeleme yapılmayan durumlarda küçük örnekleme MFB madde seçim yöntemi ile birlikte testin 20 ve 60 maddede sonlandırıldığı koşullar haricinde; kapsam dengeleme yapılan durumlarda ise geniş örnekleme KL madde seçim yöntemi ile birlikte testin 20 maddede ve $SH \leq 30$ kuralıyla sonlandırıldığı koşullar dışında BSD kestirim yöntemi ML yöntemine göre daha düşük yanlılık değerleri vermiştir.
- Küçük örnekleme kapsam dengeleme yapılan durumlarda, KL madde seçim yöntemi MOK yetenek kestirim yöntemi ile birlikte kullanıldığında ortalama hata değeri artış gösterirken BSD kestirim yöntemi ile birlikte kullanıldığında düşüş gösterdiği görülmüştür.
- MFB madde seçim yöntemi kullanılan koşullarda ortalama hata değeri, MOK yöntemi tercih edildiğinde BSD'a göre daha yüksek bulunmuştur. Fakat KL madde seçim yöntemi kullanıldığında elde edilen değerler birbirine oldukça yakındır.
- Her iki örneklem büyüklüğünden elde edilen yanlılık değerleri kapsam dengeleme yapılan ve yapılmayan durumlarda birbirine oldukça yakın iken BSD kestirim yöntemi ile birlikte KL madde seçim yöntemi ve standart hataya göre sonlandırma kuralları uygulanan koşullarda büyük örnekleme daha büyük yanlılık değerleri elde edilmiştir.
- Çalışma kapsamında ele alınan farklı test sonlandırma kurallarına ilişkin elde edilen değerler incelendiğinde; tüm koşullarda en düşük ortalama hata değerinin 60 madde ile

test sonlandırıldığında elde edildiği görülmüştür. Testin $SH \leq 30$ kuralıyla ve 20 maddede sonlandırıldığı durumlarda bu değerler oldukça yakın bulunmuşken $SH \leq 50$ kuralında RMSE değeri en yüksek değerleri almıştır.

- Yanlılık değerlerine bakıldığında ise; $SH \leq 30$, $SH \leq 40$ ve 20 maddede testi sonlandırma kurallarında değerler birbirine oldukça yakınken $SH \leq 50$ kuralı ile ve 60 maddede testin sonlandırıldığı koşullarda bu değer diğer sonlandırma kurallarının uygulandığı koşullara göre nispeten daha yüksek bulunmuştur.
- Bireylerin gerçek ve kestirilen yetenek düzeyleri arasındaki en yüksek korelasyon testin 60 maddede sonlandırıldığı koşullardan elde edilirken, en düşük korelasyon testin $SH \leq 50$ kuralı ile sonlandırıldığı koşullardan elde edilmiştir.
- Kapsam dengeleme yapılmasının, testin 60 maddede sonlandırıldığı koşullarda, ölçmenin kesinliğine ilişkin ortalama hata, yanlılık ve uyum değerlerinde meydana getirdiği değişim minimumdur. Diğer sonlandırma kurallarında da değerler arasındaki farkın birbirine yakın ve oldukça küçük olduğu görülmektedir.
- Araştırma kapsamında ele alınan tüm koşullar incelendiğinde; en düşük ortalama hata değeri geniş örnekleme kapsam dengelemenin olmadığı, BSD kestirim yöntemi ile MFB madde seçim yöntemi birlikte kullanılıp testin 60 maddede sonlandırıldığı koşulda elde edilmiştir. En yüksek ortalama hata değeri ise geniş örnekleme kapsam dengelemenin yapıldığı, MOK yöntemi ile MFB madde seçim yöntemi birlikte kullanılıp testin $SH \leq 50$ kuralıyla sonlandırıldığı koşulda elde edilmiştir.
- Tüm koşullarda en düşük yanlılık değeri geniş örnekleme kapsam dengelemenin olmadığı, BSD kestirim yöntemi ile MFB madde seçim yöntemi birlikte kullanılıp testin $SH \leq 50$ kuralıyla sonlandırıldığı koşulda elde edilmiştir. En yüksek yanlılık değeri ise küçük örnekleme kapsam dengelemenin yapıldığı, MOK yöntemi ile KL madde seçim yöntemi birlikte kullanılıp testin $SH \leq 50$ kuralıyla sonlandırıldığı koşulda elde edilmiştir.

- Standart hataya göre test sonlandırıldığı durumlarda en fazla madde küçük örnekleme MOK yöntemi ile birlikte MFB madde seçim yönteminin kullanılıp kapsam dengelemenin yapıldığı koşullarda uygulanmıştır. En az madde ise geniş örnekleme BSD kestirim yöntemi ile birlikte MFB madde seçim yönteminin kullanılıp kapsam dengelemenin yapılmadığı koşullarda uygulanmıştır.
- Araştırma, kapsam dengelemenin, belirli bir doğruluk eşiğine ulaşmak için, özellikle standart hatanın belirli bir eşik değerinin altına düştüğü sıkı sonlandırma kuralları altında, gereken madde sayısını genel olarak artırdığı bulunmuştur. Bununla birlikte, madde sayısındaki bu artış farklı koşullar arasında ortalama olarak 1 maddeyi geçmemektedir. Bu durum, BBT uygulamalarında kapsam dengeleme yapılmasının test verimliliğini önemli ölçüde azaltmadığını göstermektedir.
- Standart hataya göre sonlandırma kurallarının tümünde kapsam dengeleme yapılan koşullarda MOK yöntemi ile birlikte KL madde seçim yöntemi kullanıldığında ortalama madde sayıları daha düşük bulunmuştur. Anı koşullarda BSD kestirim yöntemi ile birlikte KL madde seçim yöntemi kullanılan koşullarda aksine ortalama madde sayısı artış göstermiştir.

Öneriler

Araştırmadan elde edilen bulgular ve araştırma sonuçlarına göre uygulayıcılara ve araştırmacılara yönelik önerilerde bulunulmuştur.

Uygulayıcılara Yönelik Öneriler

- Çalışma kapsamında ele alınan BSD ve MOK yetenek kestirim yöntemlerinden BSD'nin daha az yanlılık ile kestirimler yaptığı görülmüştür. Bu nedenle kullanılacak veri seti göz önünde bulundurularak BSD yetenek kestirim yönteminin kullanılması önerilebilir.
- Çalışmada geniş örnekleme BSD kestirim yöntemi ile birlikte KL madde seçim yönteminin kullanılmasının ortalama hata değerlerinde artışa neden olduğu görülmüştür.

Bu nedenle geniş örnekleme BSD yetenek kestirim yöntemi kullanılıyorsa MFB madde seçim yönteminin tercih edilmesi önerilir.

- Çalışmada küçük örnekleme kapsam dengeleme yapılan durumlarda, KL madde seçim yöntemi BSD kestirim yöntemi ile birlikte kullanıldığında ortalama hata değerinde düşüş görülmüştür. Dolayısıyla Küçük örneklemlerde kapsam dengeleme yapılırken bu iki yöntemin birlikte kullanılması önerilir.
- Çalışmada kapsam dengeleme yapıp standart hataya göre test sonlandırıldığında MOK kestirim yöntemi ile birlikte KL madde seçim yöntemi kullanıldığında test verimliliğinde artış gözlenmiştir. Bu nedenle standart hataya göre sonlandırma yapılan kapsam dengeleme çalışmalarında bu yöntemlerin birlikte kullanımı önerilebilir.
- Farklı sonlandırma koşullarından elde edilen ortalama hata, yanlılık ve kestirilen yetenek düzeyleri arasındaki korelasyon göz önünde bulundurulduğunda $SH \leq 20$ ve $SH \leq 30$ sonlandırma kurallarının tercih edilmesi önerilebilir.
- Çalışmada kapsam dengeleme yapılmasının, 60 maddelik uzun testte ölçmenin kesinliğine ilişkin ortalama hata, yanlılık ve uyum değerlerinde meydana getirdiği değişimin minimum olduğu gözlemlendiğinden özellikle farklı içerik alanlarının temsilinin kritik olduğu uzun testlerde kapsam dengeleme yapılması önerilebilir.
- Kapsam dengelemenin, ortalama hata ve yanlılık değerlerinde hafif artışlara neden olurken tahmin edilen ve gerçek yetenek düzeyleri arasındaki korelasyonu önemli ölçüde etkilemediği görülmüştür. Bu nedenle BBT uygulamalarında içerik temsilinin ve adilliğin kritik öneme sahip olduğu durumlarda kapsam dengeleme yapılması önerilebilir.
- Uygulayıcılar, test tasarımında kapsam dengelemenin erken aşamalarda düşünmeli ve özellikle geniş ölçekli sınavlar ile çok alanlı testlerde tüm içerik alanlarının yeterince temsil edilmesini sağlamalıdır. Kapsam dengeleme, testin doğruluğu üzerinde doğrudan büyük bir etkiye sahip olmasa da testin kapsam geçerliğini artırarak güvenilir ve geçerli sonuçlar

elde edilmesine yardımcı olur. Bu süreç, testin verimliliğini fazla etkilemeden kapsam geçerliğini sağlama adına kritik bir rol oynar.

Araştırmacılara Yönelik Öneriler

- Mevcut çalışma, simülasyon tabanlı metodolojilere dayalı olup gerçek test ortamlarındaki karmaşıklık ve değişkenliği tam olarak yansıtmamaktadır. Bulgularımız, simülasyon koşullarında elde edilen kapsam dengeleme etkilerinin nispeten sınırlı olduğunu göstermektedir. Bu nedenle, özellikle yüksek riskli test ortamlarında kapsam dengelemenin gerçek verilerle ve farklı test bağlamlarında incelenmesi önerilir. Böylelikle bulgular, gerçek hayat uygulamaları için daha genel geçer hale gelebilir.
- Çalışmada kullanılan iki kategorili maddeler ve 3PLM, ölçme kesinliğini belirlemede başarılı olsa da çok kategorili maddelerle yapılan uygulamalar üzerine yapılan çalışmaların bu tür test sistemlerinin uygulanabilirliği ve etkinliği hakkında daha geniş bir perspektif sunabileceği öngörülmektedir. Çok kategorili maddelerle kapsam dengelemenin etkilerinin araştırılması önerilmektedir.
- Bu çalışmada madde parametreleri araştırmacının belirlediği değerlere göre tek biçimli dağılım kullanılarak üretilmiştir. Farklı dağılımlara ve parametrelere sahip maddeler ile madde havuzu oluşturularak simülasyon çalışmaları yapılabilir.
- Çalışmada bireylere ait yetenek parametreleri normal dağılım gösterecek şekilde araştırmacı tarafından üretilmiştir. Farklı dağılıma sahip yetenek grupları için benzer koşullarda ölçme kesinliğinin nasıl değiştiği çalışılabilir.
- Çalışmada iki farklı örneklem büyüklüğünde farklı koşullarda kapsam dengelemenin etkisi incelenmiştir. Daha küçük ve daha geniş örneklem büyüklükleri kullanılarak farklı örneklem büyüklüklerinde çalışmanın nasıl sonuçlanacağı araştırılabilir.
- Çalışmada testi başlama kuralı $-0.5 \leq b \leq 0.5$ olarak belirlenmiştir. Farklı test başlatma kuralları tercih edilebilir.

- Mevcut çalışmada kapsam dengeleme stratejilerinden KBBT yöntemi kullanılmıştır. Gelecekteki araştırmalarda farklı kapsam dengeleme yaklaşımlarını içeren algoritma ve stratejiler kullanarak yöntem karşılaştırması yapılabilir.
- Çalışmada kapsam dengelemenin, test güvenliği, madde kullanım sıklığı oranları ve değerlendirmenin adilliği gibi BBT'nin diğer önemli yönleri üzerindeki etkisi ele alınmamıştır. Özellikle yüksek riskli test ortamlarında kapsam dengelemenin uygulanabilirliği için kritik olan bu hususların gelecekteki araştırmalarda derinlemesine incelenerek bu ek faktörleri hesaba katan daha kapsamlı modeller çalışılabilir.
- Çalışmada kapsam dengeleme yapılırken içerik alanları eşit oranlarla ağırlıklandırılmıştır. İçerik alanlarının farklı oranlarla ağırlıklandırılmasıyla benzer bir çalışma gerçekleştirilebilir.
- Çalışmada MFB ve KL madde seçim yöntemleri kullanılmıştır. Gelecek çalışmalarda diğer madde seçme yöntemleri de çalışmaya dahil edilerek daha kapsamlı bir araştırma yapılabilir.

Kaynaklar

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.
- Babcock, B. & Weiss, D. J. (2009). Termination criteria in computerized adaptive tests: Variable-length CATs are not biased. Proceedings of the 2009 GMAC conference on computerized adaptive testing.
- Babcock, B. & Weiss, D. J. (2012). Termination criteria in computerized adaptive tests: Do variable-length CATs provide efficient and effective measurement? *Journal of Computerized Adaptive Testing*, 1, 1–5. <https://doi.org/10.7333/1212-0101001>
- Baker, F.B. (1992). The basics of item response theory. *ERIC Clearinghouse on Assessment and Evaluation*.
- Bock, D. R. & Mislevy, R. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6(4), 431-444. <https://doi.org/10.1177/014662168200600405>
- Bulut, O. & Kan, A. (2012) Application of computerized adaptive testing to entrance examination for graduate studies in Turkey. *Eurasian Journal of Educational Research*, 49, 61-80.
- Chang, H.H. & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20(3), 213-229.
- Chang, H. H. (2004). Understanding computerized adaptive testing. In D. Kaplan (Ed.). *The sage handbook of quantitative methodology for the social sciences* (pp.117-133).
- Chang, S. & Ansley, T. (2003). A comparative study of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 40(1), 71-103.

- Chen, S., Ankenmann, R. D. & Spray, J. A. (1999). *Exploring the relationship between item exposure rate and test overlap rate in computerized adaptive testing*. The National Council of Educational Measurement toplantısında sunulan bildiri, Montreal, Canada.
- Chen, S. K., Hou, L., & Dodd, B. G. (1998). A comparison of maximum likelihood estimation and expected a posteriori estimation in CAT using the partial credit model. *Educational and Psychological Measurement*, 58(4), 569.
- Chen, Y., & Ankenman, R. D. (2004). Effects of practical constraints on item selection rules at the early stages of computerized adaptive testing. *Journal of Educational Measurement*, 41(2), 149-174.
- Cheng Y. & Chang, H-H. (2007). Two-phase item selection procedure for flexible content balancing in CAT. *Applied Psychological Measurement*, 31(6), 467-482.
- Crocker, L. & Algina, J. (1986). Introduction to item response theory. L. Crocker ve J. Algina (Ed.), *Introduction to classical and modern test theory*. Cengage Learning.
- Davey, T. (2011). *A guide to computer adaptive testing systems*. Council of Chief State School Officers.
- Davis, L.L. (2002). *Strategies for controlling item exposure in computerized adaptive testing with polytomously scored items*. (Doctoral Dissertation). University of Texas, Austin.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. The Guilford Press.
- Demars, C. (2010). *Item response theory*. New York: Oxford University Press, Inc.
- Demir, S. (2019). *Bireyselleştirilmiş bilgisayarlı sınıflama testlerinde sınıflama doğruluğunun incelenmesi*. (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Economides, A.A. & Roupas, C. (2007). Evaluation of computer adaptive testing systems. *International Journal of Web Web-Based Learning and Teaching Technologies*, 2(1).

- Eggen, T. J. H. M. (2004). *Contributions to the theory and practice of computerized adaptive testing* (Doctoral Dissertation). Ipskamp B.V., Enschede.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum
- Feinberg, R. A., & Rubright, J. D. (2016). Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice*, 35(2), 36-49. <https://doi.org/10.1111/emip.12111>
- Flaugher, R. (2000). Item pools. In H. Wainer (Ed.), *Computerized Adaptive Testing: A primer* (pp. 37-60). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Georgiadou, E., Triantafillou, E. & Economides, A.A. (2007). A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *The Journal of Technology, Learning, and Assessment*, 5(8).
- Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L. & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21(4), 347–360. <https://doi.org/10.1111/j.1745-3984.1984.tb01039.x>
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R Package. *Journal of Statistical Software*, 33(2), 1–22. <https://doi.org/10.18637/jss.v033.i02>
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory principles and applications*. Boston: Kluwer.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory* (1. Baskı). SAGE Publications, Inc.
- Han, K. (2012). SimulCAT: Windows application that simulates computerized adaptive test administration. *Applied Psychological Measurement*, 36, 64-66. <https://doi.org/10.1177/0146621611414407>

Han, K.T. (2018). Components of the item selection algorithm in computerized adaptive testing. *Journal of Educational Evaluation for Health Professions* 15(7).

<https://doi.org/10.3352/jeehp.2018.15.7>

Harwell, M., Stone, C. A., Hsu, T.C., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement*, 20(2), 101-125.

<https://doi.org/10.1177/014662169602000201>

Kingsbury, G. G. & Zara, A. R. (1989). Procedures for selecting item for computerized adaptive tests. *Applied Measurement in Education* 2(4), 359-375.

Kingsbury, G. G., Bontempo, B., & Zara A. R. (2009). A comparison of CAT with LOFT methods for certification examinations. NOCA Yıllık Eğitim Konferansında sunum.

Leung, C.K., Chang, H.H., & Hau, K.T. (2000). *Content balancing in stratified computerized adaptive testing designs*. Amerikan Eğitim Araştırmaları Derneği Yıllık Toplantısında Sunulan Bildiri, New Orleans.

Leung, C.K., Chang, H.H., & Hau, K.T. (2003a). Computerized adaptive testing: A comparison of three content balancing methods. *Journal of Technology, Learning, and Assessment*, 2(5).

Leung, C.K., Chang, H.H., & Hau, K.T. (2003b). Incorporation of content balancing requirements in stratification designs for computerized adaptive testing. *Educational and Psychological Measurement*, 63(2), 257-270.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates.

Luecht, R. M. (2003). Exposure control using adaptive multi-stage item bundles. *Psychometrika*, 68(3), 245-260. <https://doi.org/10.1007/BF02294769>

Magis, D., Yan, D. & von-Davies, A. (Eds.). (2017). *Computerized adaptive and multistage testing with R. Using packages catr and mstr*. Switzerland: Springer.

- Magis, D., Raiche, G. & Barrada, J. R. (2018). Package 'catR': Generation of IRT response patterns under computerized adaptive testing. <https://cran.r-project.org/web/packages/catR/catR.pdf>
- Mills, C. N., & Stocking, M. L. (1996). Practical issues in large-scale computerized adaptive testing. *Applied Measurement in Education*, 9(4), 287–304. https://psycnet.apa.org/doi/10.1207/s15324818ame0904_1
- Özdemir, B. (2015). İçerik ağırlıklandırmasının maddeler-arası boyutluluk modeline dayalı çok boyutlu bilgisayar ortamında bireyselleştirilmiş test yöntemleri üzerindeki etkisinin incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(2), 365-384.
- Özer Taymur, M. (2023). *BOBUT uygulamalarında farklı koşullardan elde edilen ölçme kesinliği kestirim değerlerinin karşılaştırılması*. (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- R Core Team (2022). R: A language and environment for statistical computing, (Version 4.2.2) [Computer software], Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>
- Ree, M. J., & Jensen, H. E. (1983). Effects of sample size on linear equating of item characteristic curve parameters, In Weiss, D. (Ed). *New horizons in testing latent trait test theory and computerized adaptive testing* (pp.135-146). London: Academic Press. <https://doi.org/10.1016/B978-0-12-742780-5.50017-2>
- Rudner, L. M. (1998). *Interactive computer adaptive testing*. <http://edres.org/scripts/cat/catdemo.htm>
- Sari, H. İ., & Manley, A. C. (2017). Examining content control in adaptive tests: Computerized adaptive testing vs. computerized adaptive multistage testing. *Educational Sciences: Theory & Practice*, 17(5). <https://doi.org/10.12738/estp.2017.5.0484>

- Sarı, H. İ. (2019). Investigating consequences of using item pre-knowledge in computerized multistage testing. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 39(2), 1113-1134. <https://doi.org/10.17152/gefad.535376>
- Segall, D. O. (2004). Computerized adaptive testing. In Kempf-Leanard (Eds.). *The Encyclopedia of Social Measurement*.
- Shin, C. D., Chien, Y., & Way, W. D. (2012). *A comparison of three content balancing methods for fixed and variable length computerized adaptive tests*. NCME'de sunulan bildiri, Canada.
- Song, T. (2010). *The effect of fitting a tridimensional IRT model to multidimensional data in content-balanced computerized adaptive testing* (Doctoral Dissertation) Michigan State University.
- Stafford, R. E, Runyon, C. R., Casabianca, J. M. & Dodd, B. G. (2019). Comparing computer adaptive testing stopping rules under the generalized partial-credit model. *Behavior Research Methods*, 51,1305–1320.
- Suhardi, I. (2020). Alternative item selection strategies for improving test security in computerized adaptive testing of the algorithm. *Research and Evaluation in Education*, 6(1).
- Şahin, A., Özbaşı D. (2017). Effects of content balancing and Item selection method on ability estimation in computerized adaptive tests. *Eurasian Journal of Educational Research*, 17(69), 21-36.
- Şenel, S. (2021). *Bilgisayar Ortamında Bireye Uyarlanmış Testler*. Pegem Akademi, Ankara.
- Thissen, D. & Mislevy, R. J. (2000). Testing algorithms. H. Wainer, (Ed.), *Computerized adaptive testing: A primer* (pp. 101-133), Lawrence Erlbaum Associates.
- Thompson, N. A. (2009). Item selection in computerized classification testing. *Educational and Psychological Measurement*, 69(5), 778-793. <https://doi.org/10.1177/0013164408324460>

- Thompson, N. A., & Weiss, D. A. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research, and Evaluation*, 16(1). <https://doi.org/10.7275/wqzt-9427>
- Tian, J., Miao, D., Zhu, X. & Gong, J. (2007). An introduction to the computerized adaptive testing. *Us-China Education Review* 4(1), 72-81.
- Urry V.W. (1970). *A monte carlo investigation of logistic test models* (Doctoral Dissertation). University of Purdeu.
- Urry, V. W. (1977). Tailored testing: A successful application of latent trait theory. *Journal of Educational Measurement*, 14(2), 181–196. <https://doi.org/10.1111/j.1745-3984.1977.tb00035.x>
- vander Linden W.J. & Glas, C.A.W. (2000). *Computerized Adaptive Testing: Theory and Practice*. Kluwer Academic Publishers.
- van der Linden, W. J. (2005). Linear models for optimal test design. *Springer-Verlag*. <https://doi.org/10.1007/0-387-29054-0>
- vander Linden W.J. (2010). Constrained adaptive testing with shadow tests. In van der Linden W.J ve Glas C. A.W. (Eds.). *Elements of adaptive testing*. Springer.
- Veerkamp, W. J. J., & Berger, M. P. F. (1997). Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics*, 22(2), 203–226. <https://doi.org/10.2307/1165378>
- Wainer, H. (2000). Introduction and history. H. Wainer (Ed.), *Computerized adaptive testing: A primer* (2. Baskı). Lawrence Erlbaum Associates.
- Wainer, H. & Mislevy, R. J. (2000). Item response theory, item calibration, and proficiency estimation. H. Wainer (Ed.), *Computerized adaptive testing: A primer* (2. Baskı.). Lawrence Erlbaum Associates.

- Wainer, H., & Thissen, D. (1987). Estimating ability with the wrong model. *Journal of Educational Statistics*, 12(4), 339–368. <https://doi.org/10.2307/1165054>
- Wang, T. & Vispoel, W. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(2), 109-135. <https://doi.org/10.1111/j.1745-3984.1998.tb00530.x>
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, 6, 473-492. <https://doi.org/10.1177/014662168200600408>
- Weiss, D. J. & Kingsbury, G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361-375. <https://doi.org/10.1111/j.1745-3984.1984.tb01040.x>
- Wise, S. L., & DeMars, C. (2005). The validity of adaptive tests in measuring student performance. *Educational Measurement: Issues and Practice*, 24(3), 15-22. <https://doi.org/10.1111/j.1745-3992.2005.00015.x>
- Wise, S. L., Kingsbury, G. G. & Webb, N. L. (2015). Evaluating content alignment in computerized adaptive testing. *Educational Measurement: Issues and Practice*, 34(4), 41-48. <https://doi.org/10.1111/emip.12094>
- Yasuda, J.& Hull, M. (2021). *Balancing content of computerized adaptive testing for the force concept inventory*. Fizik Eğitimi Araştırma Konferansı Bildirisi (PERC).
- Yi, Q., Chang, H. (2003). a-Stratified CAT design with content blocking. *British Journal of Mathematical and Statistical Psychology*, 56(2), 359-78. <https://doi.org/10.1348/000711003770480084>
- Zheng Y, Chang C-H & Chang H-H. (2013). Content-balancing strategy in bifactor computerized adaptive patient-reported outcome measurement. *Qual Life Res*, 22(3).

Zheng, Y., & Chang, H-H. (2014). On-the-fly assembled multistage adaptive testing. *Applied Psychological Measurement*, 39(2), 105-118.

<https://doi.org/10.1177/0146621614544519>



EK-A: Madde Havuzunun Farklı İçerik Alanlarına Ait Maddelerin Betimsel İstatistikleri**1.İçerik Alanı Maddeleri Betimsel İstatistikleri**

Madde parametreleri	Minimum Değeri	Maksimum Değeri	Ortalama	Standart Sapma
a	0.509	1.999	1.258	0.448
b	-2.988	2.975	0.062	1,744
c	0.051	0.199	0.125	0.045

2.İçerik Alanı Maddeleri Betimsel İstatistikleri

Madde parametreleri	Minimum Değeri	Maksimum Değeri	Ortalama	Standart Sapma
a	0.514	1.986	1.239	0.445
b	-2.951	2.962	0.078	1,604
c	0.051	0.199	0.125	0.045

3.İçerik Alanı Maddeleri Betimsel İstatistikleri

Madde parametreleri	Minimum Değeri	Maksimum Değeri	Ortalama	Standart Sapma
a	0.502	1.999	1.202	0.450
b	-2.971	2.964	0.107	1,681
c	0.050	0.200	0.122	0.044

4.İçerik Alanı Maddeleri Betimsel İstatistikleri

Madde parametreleri	Minimum Değeri	Maksimum Değeri	Ortalama	Standart Sapma
a	0.503	1.992	1.200	0.439
b	-2.933	2.976	-0.014	1,755
c	0.050	0.198	0.124	0.044

5.İçerik Alanı Maddeleri Betimsel İstatistikleri

Madde parametreleri	Minimum Değeri	Maksimum Değeri	Ortalama	Standart Sapma
a	0.503	1.994	1.224	0.450
b	-2.989	2.951	-0.130	1,682
c	0.051	0.200	0.122	0.045

EK-B: BSD Yöntemi Kullanılan Koşullardan Elde Edilen Ölçme Kesinliğine İlişkin Değerler

Koşul	Kapsam Dengeleme Yok				Kapsam Dengeleme Var			
	RMSE	Yanlılık	Uyum	Ortalama Madde Sayısı	RMSE	Yanlılık	Uyum	Ortalama Madde Sayısı
N=250, 20 madde, MFB	0,1900	0,0368	0,983	-	0,1935	0,0322	0,9821	-
N=250 60 madde, MFB	0,1282	0,0463	0,9931	-	0,1283	0,0462	0,993	-
N=250, SH \leq .30, MFB	0,2025	0,0346	0,9805	17,4	0,2028	0,0299	0,9802	18
N=250, SH \leq .40, MFB	0,2694	0,0272	0,9638	9,82	0,2697	0,0313	0,9638	10,27
N=250, SH \leq .50, MFB	0,3460	0,0279	0,9395	6,38	0,3483	0,0331	0,9388	6,64
N=500, 20 madde, MFB	0,1868	0,0313	0,9821	-	0,1928	0,0325	0,9825	-
N=500, 60 madde, MFB	0,1255	0,0459	0,9929	-	0,1281	0,0453	0,9931	-
N=500, SH \leq .30, MFB	0,1968	0,0289	0,9799	17,32	0,2021	0,0314	0,9807	18,02
N=500, SH \leq .40, MFB	0,2693	0,0232	0,9646	9,84	0,2759	0,0270	0,9629	10,28
N=500, SH \leq .50, MFB	0,3453	0,0149	0,9408	6,40	0,3507	0,0219	0,9389	6,66
N=250,20 madde, KL	0,1981	0,0431	0,9818	-	0,1993	0,0354	0,9801	-
N=250, 60 madde, KL	0,1266	0,0507	0,9924	-	0,1277	0,0465	0,9926	-
N=250, SH \leq .30, KL	0,2038	0,0261	0,9810	17,76	0,2063	0,0331	0,9787	18,34
N=250, SH \leq .40, KL	0,2816	0,0355	0,9642	10,15	0,2791	0,0307	0,9609	10,58
N=250, SH \leq .50, KL	0,3608	0,0265	0,9334	6,66	0,3687	0,0374	0,9359	6,91
N=500, 20 madde, KL	0,1970	0,0436	0,9815	-	0,2022	0,0415	0,9817	-
N=500, 60 madde, KL	0,1274	0,0471	0,9933	-	0,1294	0,0468	0,9927	-
N=500, SH \leq .30, KL	0,2058	0,0353	0,9793	17,68	0,2121	0,0403	0,9783	18,31
N=500, SH \leq .40, KL	0,2801	0,0429	0,9565	10,08	0,2827	0,0366	0,959	10,58
N=500, SH \leq .50, KL	0,3602	0,0485	0,9268	6,57	0,3728	0,0518	0,9403	6,90

EK-C: MOK Yöntemi Kullanılan Koşullardan Elde Edilen Ölçme Kesinliğine İlişkin Değerler

Koşul	Kapsam Dengeleme Yok				Kapsam Dengeleme Var			
	RMSE	Yanlılık	Uyum	Ortalama Madde Sayısı	RMSE	Yanlılık	Uyum	Ortalama Madde Sayısı
N=250, 20 madde, MFB	0,2044	0,0357	0,9806	-	0,2135	0,0394	0,9791	-
N=250, 60 madde, MFB	0,1327	0,0453	0,9927	-	0,1349	0,0463	0,9925	-
N=250, SH≤.30, MFB	0,2045	0,0348	0,9804	19,01	0,2070	0,0407	0,9801	19,73
N=250, SH≤.40, MFB	0,2861	0,0376	0,9607	11,37	0,2830	0,0442	0,9661	11,98
N=250, SH≤.50, MFB	0,3732	0,0497	0,9353	8,00	0,3758	0,0561	0,9425	8,48
N=500,20 madde, MFB	0,2068	0,0385	0,9805	-	0,2132	0,0416	0,9796	-
N=500,60 madde, MFB	0,1315	0,0465	0,9929	-	0,1329	0,0474	0,9928	-
N=500, SH≤.30, MFB	0,2067	0,0374	0,9804	18,94	0,2097	0,0390	0,9800	19,67
N=500, SH≤.40, MFB	0,2794	0,0381	0,9654	11,43	0,2825	0,0457	0,9623	11,91
N=500, SH≤.50, MFB	0,3730	0,0480	0,9406	8,07	0,3777	0,0622	0,9352	8,41
N=250,20 madde, KL	0,2006	0,0462	0,9810	-	0,2094	0,0494	0,9793	-
N=250, 60 madde, KL	0,1309	0,0518	0,9928	-	0,1343	0,0516	0,9922	-
N=250, SH≤.30, KL	0,2088	0,0463	0,9790	18,92	0,2102	0,0493	0,9789	19,55
N=250, SH≤.40, KL	0,2788	0,0493	0,9637	11,46	0,2871	0,0537	0,9600	11,82
N=250, SH≤.50, KL	0,3619	0,0599	0,9401	8,09	0,3647	0,0660	0,9379	8,34
N=500, 20 madde, KL	0,2013	0,0484	0,9809	-	0,2052	0,0356	0,9763	-
N=500, 60 madde, KL	0,1318	0,0493	0,9924	-	0,1327	0,0524	0,9927	-
N=500, SH≤.30, KL	0,2073	0,0478	0,9801	18,96	0,2051	0,0333	0,9762	19,52
N=500, SH≤.40, KL	0,2790	0,0517	0,9646	11,49	0,2824	0,0535	0,9622	11,86
N=500, SH≤.50, KL	0,3602	0,0570	0,9428	8,16	0,3617	0,0602	0,9392	8,38

EK-Ç: BSD Yöntemi Kullanılan Koşullardan Elde Edilen Ölçme Keskinliğine İlişkin

Değerlere Ait Standart Hatalar


Koşul	Kapsam Dengeleme Yok			Kapsam Dengeleme Var		
	RMSE	Yanlılık	Uyum	RMSE	Yanlılık	Uyum
N=250, 20 madde, MFB	0,0013	0,0016	0,0002	0,0013	0,0015	0,0002
N=250 60 madde, MFB	0,0014	0,001	0,0002	0,0007	0,001	0,0001
N=250, SH≤.30, MFB	0,0017	0,0018	0,0003	0,0019	0,0017	0,0004
N=250, SH≤.40, MFB	0.0026	0.0023	0.0008	0.0025	0.0025	0.0007
N=250, SH≤.50, MFB	0.0029	0.0027	0.0010	0.0027	0.0031	0.0009
N=500, 20 madde, MFB	0.0009	0.0011	0.0002	0.0012	0.0013	0.0002
N=500, 60 madde, MFB	0.0006	0.0008	0.0001	0.0009	0.0010	0.0001
N=500, SH≤.30, MFB	0.0010	0.0012	0.0002	0.0013	0.0014	0.0002
N=500, SH≤.40, MFB	0.0011	0.0018	0.0003	0.0012	0.0019	0.0003
N=500, SH≤.50, MFB	0.0014	0.0023	0.0005	0.0017	0.0023	0.0006
N=250,20 madde, KL	0.0013	0.0014	0.0002	0.0013	0.0014	0.0003
N=250, 60 madde, KL	0.0007	0.0009	0.0001	0.0009	0.0011	0.0001
N=250, SH≤.30, KL	0.0012	0.0016	0.0003	0.0016	0.0015	0.0003
N=250, SH≤.40, KL	0.0018	0.0019	0.0005	0.0017	0.0024	0.0005
N=250, SH≤.50, KL	0.0027	0.0034	0.0010	0.0020	0.0028	0.0008
N=500, 20 madde, KL	0.0009	0.0010	0.0001	0.0009	0.0010	0.0002
N=500, 60 madde, KL	0.0006	0.0006	0.0001	0.0010	0.0008	0.0001
N=500, SH≤.30, KL	0.0008	0.0012	0.0002	0.0022	0.0011	0.0006
N=500, SH≤.40, KL	0.0015	0.0017	0.0005	0.0011	0.0015	0.0003
N=500, SH≤.50, KL	0.0015	0.0018	0.0007	0.0016	0.0023	0.0006

EK-D: MOK Yöntemi Kullanılan Koşullardan Elde Edilen Ölçme Kesinliğine İlişkin

Değerlere Ait Standart Hatalar

Koşul	Kapsam Dengeleme Yok			Kapsam Dengeleme Var		
	RMSE	Yanlılık	Uyum	RMSE	Yanlılık	Uyum
N=250,20 madde, MFB	0,003	0,0018	0,0007	0,003	0,0019	0,0007
N=250 60 madde, MFB	0,0023	0,001	0,0004	0,0022	0,001	0,0004
N=250, SH≤.30, MFB	0,0029	0,0019	0,0007	0,0028	0,0016	0,0006
N=250, SH≤.40, MFB	0,0024	0,0024	0,0007	0,0025	0,0027	0,0006
N=250, SH≤.50, MFB	0,0042	0,0027	0,0014	0,0041	0,0035	0,0011
N=500, 20 madde, MFB	0,0021	0,0013	0,0004	0,0015	0,0013	0,0003
N=500, 60 madde, MFB	0,0015	0,0007	0,0002	0,0006	0,0005	0,0001
N=500, SH≤.30, MFB	0,0020	0,0013	0,0004	0,0015	0,0011	0,0003
N=500, SH≤.40, MFB	0,0019	0,0015	0,0005	0,0019	0,0018	0,0005
N=500, SH≤.50, MFB	0,0028	0,0019	0,0008	0,0033	0,0023	0,0010
N=250,20 madde, KL	0,0015	0,0017	0,0003	0,0022	0,0020	0,0004
N=250, 60 madde, KL	0,0009	0,0011	0,0001	0,0022	0,0011	0,0003
N=250, SH≤.30, KL	0,0021	0,0019	0,0004	0,0021	0,0017	0,0005
N=250, SH≤.40, KL	0,0019	0,0022	0,0004	0,0024	0,0027	0,0007
N=250, SH≤.50, KL	0,0035	0,0028	0,0012	0,0035	0,0031	0,0012
N=500, 20 madde, KL	0,0010	0,0013	0,0002	0,0013	0,0011	0,0003
N=500, 60 madde, KL	0,0008	0,0008	0,0001	0,0006	0,0008	0,0001
N=500, SH≤.30, KL	0,0011	0,0013	0,0002	0,0010	0,0013	0,0002
N=500, SH≤.40, KL	0,0012	0,0015	0,0003	0,0016	0,0016	0,0005
N=500, SH≤.50, KL	0,0023	0,0023	0,0007	0,0026	0,0021	0,0009

EK E: Araştırma Etik Komisyon İzin Muafiyeti Formu

	Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması/Araştırma Etik Komisyon İzin Muafiyeti Formu	F46
		12/ 11 / 2024
Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Ana Bilim Dalı Başkanlığına		
Tez/Araştırma Başlığı	Bilgisayarda Bireyselleştirilmiş Test Uygulamalarında Kapsam Dengelemenin Ölçme Keskinliğine Etkisi	
Yukarıda başlığı/konusu verilen tez/araştırma çalışmam,		
<ol style="list-style-type: none"> 1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır. 2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir. 3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir. 4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırmalar niteliğinde değildir. 5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir. 		
Çalışmada kullanacağım veriler:		
<input type="checkbox"/> Kamusal erişime açık (buraya yazınız):		
<input type="checkbox"/> Özel izin ve onaya tabi (buraya yazınız):		
<input checked="" type="checkbox"/> Üretilmiş veri (buraya yazınız): R Programlama Dili kullanılarak oluşturulan simülasyon verisi		
<input type="checkbox"/> Diğer (buraya yazınız):		
Yükseköğretim Kurumları Etik Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.		
Gereğini saygılarımla arz ederim.		
İlkay Üçgül Öcal (Araştırmacı Adı Soyadı, İmzası)		
Araştırmacı Bilgileri		
Adı Soyadı	İlkay Üçgül Öcal	
Öğrenci İse No	N20143700	
Ana Bilim Dalı	Eğitim Bilimleri Ana Bilim Dalı	
Programı	Eğitimde Ölçme ve Değerlendirme	
Çalışma Türü	<input checked="" type="checkbox"/> Tez <input checked="" type="checkbox"/> Tezden Üretilen Yayın <input type="checkbox"/> Araştırma Makalesi	
Statüsü	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr. <input type="checkbox"/> Diğer	
Danışman Görüşü ve Onayı*		
Prof. Dr. Nuri Doğan		
*Tez ve tezden üretilen yayın ve araştırma makalelerinde gerekli		
Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü. Beytepe Yerleşkesi, 06800, Çankaya / ANKARA Telefon: 0(312) 297 85 72 Belgegeçer: 0(312) 297 85 66 e-Aç: http://ebe.hacettepe.edu.tr/ e-Posta: ebe@hacettepe.edu.tr		

EK-F: Etik Beyanı

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

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

beyan ederim.

...../...../2024

(İmza)

İlkay ÜÇGÜL ÖCAL

EK-G: Doktora Tez Çalışması Orijinallik Raporu

13/12/2024

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

Tez Başlığı: Bilgisayarda Bireyselleştirilmiş Test Uygulamalarında Kapsam Dengelemenin Ölçme Kesinliğine Etkisi

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
13/12/2024	92	130330	28/11/2024	%7	2551256981

Uygulanan filtreler:

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

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

Ad Soyadı: İlkay ÜÇGÜL ÖCAL

Öğrenci No.: N20143700

Ana Bilim Dalı: Eğitim Bilimleri

İmza

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

Prof. Dr. Nuri DOĞAN

EK-H: Dissertation Originality Report

13/12/2024

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

Thesis Title: Effects of Content Balancing on Measurement Precision in Computerized Adaptive 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
13/12/2024	92	130330	28/11/2024	%7	2551256981

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: İlkey ÜÇGÜL ÖCAL

Student No.: N20143700

Department: Educational Sciences

Program: Educational Measurement and Evaluation

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

Signature

ADVISOR APPROVAL

APPROVED
Prof. Dr. Nuri DOĞAN

EK-I: Yayınlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm 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.

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

..... / /

(imza)

İlkay ÜÇGÜL ÖCAL

"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 tezini 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 tezini 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.

