



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

MADDE TAKIMLARINDAN OLUŞAN TESTLERDE BİLGİSAYARDA
BİREYSELLEŞTİRİLMİŞ TEST UYGULAMA DESENLERİNİN KARŞILAŞTIRILMASI

Sebahat GÖREN

Doktora Tezi

Ankara, 2024

Liderlik, arařtırma, inovasyon, kaliteli eęitim ve deęiřim ile

Daha ileriye ... En İyiyeye ...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

MADDE TAKIMLARINDAN OLUŞAN TESTLERDE BİLGİSAYARDA
BİREYSELLEŞTİRİLMİŞ TEST UYGULAMA DESENLERİNİN KARŞILAŞTIRILMASI

A COMPARISON OF COMPUTERIZED ADAPTIVE TESTING DESIGNS FOR TESTLET
BASED TESTS

Sebahat GÖREN

Doktora Tezi

Ankara, 2024

Kabul ve Onay

Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Sebahat GÖREN'in hazırladığı "Madde Takımlarından Oluşan Testlerde Bilgisayarda Bireyselleştirilmiş Test Uygulama Desenlerinin Karşılaştırılması" başlıklı bu çalışma jürimiz tarafından **Eğitim Bilimleri Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiştir.

Jüri Başkanı	Prof. Dr. Hülya KELECİOĞLU	İmza
Jüri Üyesi (Danışman)	Prof. Dr. Selahattin GELBAL	İmza
Jüri Üyesi	Prof. Dr. Hakan Yavuz ATAR	İmza
Jüri Üyesi	Prof. Dr. Neşe GÜLER	İmza
Jüri Üyesi	Doç. Dr. Kübra ATALAY KABASAKAL	İmza

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim, Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından 06 / 06 / 2024 tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca / / tarihi itibarıyla kabul edilmiştir.

Prof. Dr. İsmail Hakkı MİRİCİ
Eğitim Bilimleri Enstitüsü Müdürü

Öz

Bu çalışmada madde takımlarından oluşan testlerin Bilgisayarda Bireyselleştirilmiş Test (BBT) uygulamaları farklı desenler oluşturularak incelenmiştir. Bu bağlamda öncelikle bilgisayar tabanlı olarak Bilkent Üniversitesi hazırlık sınıfı öğrencilerine uygulanan ve dokuz formdan oluşan İngilizce seviye belirleme sınavından elde edilen gerçek veri seti kullanılmıştır. Her bir form, ölçekleme yapmak amacıyla dokuz ortak madde içerdiğinden eş zamanlı kalibrasyon yöntemiyle madde havuzu oluşturulmuştur. Madde takımlarının bağımsız birer madde olarak ele alındığı BBT, madde takımlarının birlikte ve madde takımı tepki modelinin temel alındığı MT-BBT deseni ve madde takımlarının yani okuma metinlerinin yine birlikte ele alındığı fakat tek boyutlu madde tepki kuramı modeli temelli OM-BBT deseni olmak üzere üç farklı BBT deseni oluşturulmuştur. Bu test desenleri madde havuzundaki parametreler kullanılarak madde takımı sayısı (6-9-12), örneklem büyüklüğü (200-500-1000) ve yetenek kestirim yöntemlerinin (EAP-MAP) kullanıldığı farklı koşullarda ölçme kesinliği ve doğruluğu, madde havuzu kullanımı açısından incelenmiştir. Bu koşullardan elde edilen gerçek ve kestirilen θ değerleri arasındaki korelasyon, BIAS, MAB, RMSE ve SH değerleri karşılaştırılmıştır. Madde havuzu kullanımı için madde kullanım sıklığı, kullanılmayan madde sayısı ve test çakışma oranları hesaplanmıştır. BBT'nin hem ölçme kesinliği hem de madde havuzu kullanımı açısından diğer desenlere göre daha etkili olduğu, OM-BBT deseninin ise genel olarak BBT deseni ile benzer sonuçlar ürettiği sonucuna ulaşılmıştır. Madde takımlarının yerel bağımlılık derecelerinin düşük olmasından dolayı MT-BBT deseninden beklenen sonuçlar elde edilememiştir. M-BBT analizlerinin karmaşık ve çok zaman almasından dolayı düşük ya da orta düzeyde madde takımı etkisine sahip veri setlerinde OM-BBT deseni kullanılabilir. Elde edilen tüm bulgular detaylı olarak tartışılmış, uygulayıcı ve araştırmacılara önerilerde bulunulmuştur.

Anahtar sözcükler: madde takımı, bilgisayarda bireyselleştirilmiş test, madde tepki kuramı, madde takımı tepki kuramı

Abstract

In this study, CAT applications of testlet based tests were examined under different designs. In this context, the real data set obtained from the computer-based English proficiency exam, consisting of nine forms, administered to students at Bilkent University. Since each form contains nine anchor items for scaling purposes, an item pool was created through concurrent calibration method. Three distinct CAT designs were developed, namely, the CAT where testlets were treated as independent items, the Testlet Based-CAT(T-CAT) where testlets were treated together and based on the Testlet Response Theory (TRT), and the Passage Based-CAT(P-CAT) where testlets were treated together but using the unidimensional Item Response Theory (IRT). These test designs were examined for measurement precision and accuracy, item pool utilization under different conditions. The simulation conditions included the number of testlets (6-9-12), sample size (200-500-1000) and ability estimation methods (EAP-MAP). The correlation, BIAS, MAB, RMSE and SH values between true and estimated θ values obtained from these conditions were compared. Item usage frequency, number of unused items and test overlap rates were calculated. The findings indicated that CAT was more effective while the P-CAT produced similar results to the CAT design in general. Conversely with the T-CAT, the expected results could not be achieved due to the low local dependency degrees of the testlets. Since T-CAT analyses are complex and time-consuming, the P-CAT may be preferred in terms of usefulness for assessments with low or medium testlet effects. All findings obtained were discussed and recommendations were made for practitioners and researchers.

Keywords: testlet, computer adaptive test, item response theory, testlet response theory

Teşekkür

Doktora sürecimin en başından sonuna kadar her konudaki yol göstericiliği, sonsuz sabrı, anlayışlı ve babacan tavrıyla her zaman yanımda olan ve her defasında bana “iyi ki” dedirten kıymetli danışmanım Prof. Dr. Selahattin Gelbal’a öncelikli olarak çok teşekkür ederim.

Tez izleme sürecimde tezimin şekillenmesini sağlayan, yapıcı yaklaşımları ile ufkumu genişleten, karakter ve duruşlarıyla kendilerini her zaman örnek aldığım Prof. Dr. Hülya Kelecioğlu ve Prof. Dr. Hakan Yavuz Atar hocalarıma;

Tez savunma sürecimde tezimi titizlikle inceleyen, önerileriyle tezime katkı sunan, enerjileri ve pozitif bakış açıları ile beni motive eden Prof. Dr. Neşe Güler ve Doç. Dr. Kübra Atalay Kabasakal hocalarıma;

Doktora eğitimim boyunca tüm sorularımı sabırla cevaplandıran ve danışmanı olduğu Tübitak projesine beni yönlendiren Prof. Dr. Nuri Doğan’a; bu projede veri toplamama imkân sağlayan proje yürütücüsü Doç. Dr. Levent Yakar’a;

Eğitim hayatımda akademik katkılarının yanı sıra yardımsever ve sevecen tavırlarıyla zor anlarımda bana destek olan Prof. Dr. Burcu Atar; Doç. Dr. Ayfer Sayın ve Doç. Dr. Sevda Çetin hocalarıma çok teşekkür ederim.

Tez yazma sürecimin en kritik aşamalarında çözümler üreten, çok uzaklarda da olsa varlığıyla beni her zaman güçlü ve güvende hissettiren canım dostum Dr. Ümmügül Bezirhan’a; desteğini yine uzaklardan esirgemeyerek tezime katkı sağlayan arkadaşım Dr. Onur Demirkaya’ya;

Doktora eğitiminde yollarımızın kesiştiği, tüm iyi ve kötü anlarımda yanımda olan yol arkadaşım Dr. Öğr. Üyesi Merve Yıldırım Seheryeli’ne; veri toplama sürecinde bana destek olan Dr. Demet Alkan’a gönülden teşekkür ederim.

Bu süreçte sorduğum soruları özveri ve samimiyetle yanıtlayarak süreci kolaylaştıran Doç. Dr. Eren Can Aybek, Dr. Öğr. Üyesi Başak Erdem Kara ve Dr. Özge Erşan Çınar'a;

Doktora eğitimim süresince burs olanağı sağlayan TÜBİTAK'a çok teşekkür ederim.

Tüm engelleri birlikte aştığım, her daim bana inanan ve güvenen, bu hayattaki en büyük şansım anneme; bugünlere gelmemde sonsuz emeği olan, eksikliğini her zaman hissettiğim iyi yürekli babama ve tüm zor anlarımı bir bakışıyla güzelleştiren, yaşama sevincim olan güzel kızıma tüm kalbimle teşekkür ederim.

Kızım Irmak'a ve anneme...

İçindekiler

Kabul ve Onay.....	ii
Öz.....	iii
Abstract.....	iv
Teşekkür.....	v
Tablolar Dizini.....	x
Şekiller Dizini.....	xi
Simgeler ve Kısaltmalar Dizini.....	xii
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi	7
Araştırma Problemi	9
Sınırlılıklar	10
Tanımlar.....	10
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	11
Madde Tepki Kuramı.....	11
Madde Tepki Kuramı Modelleri	13
Madde ve Test Bilgi Fonksiyonu	17
Madde Takımı	18
Madde Takımı Tepki Kuramı ve Modelleri.....	20
Bilgisayarda Bireyselleştirilmiş Test.....	24
Madde Takımı Tabanlı BBT	35
İlgili Araştırmalar	38
İlgili Araştırmalar Özeti	44
Bölüm 3 Yöntem.....	45
Araştırmanın Türü	45
Verilerin Elde edilmesi.....	45

Veri Analizi	47
Madde Havuzunun Oluřturulması	47
Yetenek Parametrelerinin Elde Edilmesi	56
Arařtırma Deseni	57
Deęerlendirme Kriteri	62
Bölüm 4 Bulgular, Yorumlar ve Tartıřma.....	65
Arařtırma Probleminin Ölçme Kesinlięi ve Doğruluęuna İliřkin Bulgular ve Yorumlar	65
Birinci Alt Arařtırma Problemine İliřkin Bulgular ve Yorumlar	69
İkinci Alt Arařtırma Problemine İliřkin Bulgular ve Yorumlar	73
Üçüncü Alt Arařtırma Problemine İliřkin Bulgular	76
Arařtırma Probleminin Madde Havuzu Kullanımına İliřkin Bulgular ve Yorumlar	81
Dördüncü Alt Arařtırma Problemine İliřkin Bulgular ve Yorumlar	83
Beřinci Alt Arařtırma Problemine İliřkin Bulgular ve Yorumlar	84
Altıncı Alt Arařtırma Problemine İliřkin Bulgular ve Yorumlar	85
Bölüm 5 Sonuç ve Öneriler.....	87
Sonuç.....	87
Öneriler	90
Kaynaklar	94
Ek-A: Her Bir Forma Göre Model Uyum İndeksleri.....	109
Ek-B: Birey-Madde Etkileřimi Haritası	110
EK-C: Etik Beyanı.....	111
EK-D: Etik Komisyon İzni	112
EK-E: Yüksek Lisans/Doktora Tez Çalıřması Orijinallik Raporu	115
EK-F: Thesis/Dissertation Originality Report	116
EK-G: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	117

Tablolar Dizini

Tablo 1 <i>Formlara Göre Madde ve Cevaplayan Sayısı</i>	46
Tablo 2 <i>Yerel Bağımlılık İçeren Madde Takımlarının Yüzdesi</i>	49
Tablo 3 <i>Boyutluluk Analizi Sonuçları</i>	51
Tablo 4 <i>Madde Havuzuna İlişkin Veri Toplama Deseni</i>	53
Tablo 5 <i>Araştırma Kapsamında Manipüle Edilen Değişkenler</i>	58
Tablo 6 <i>Farklı BBT Desenlerine Göre Tüm Koşullardan Elde Edilen Bulgular</i>	66
Tablo 7 <i>BBT Desenlerine Göre RMSE, BIAS, MAB ve r Değerleri</i>	67
Tablo 8 <i>Madde takımı sayısına Göre RMSE, BIAS, MAB ve SH Değerleri</i>	69
Tablo 9 <i>Örneklem Büyüklüğüne Göre RMSE, BIAS, MAB ve SH Değerleri</i>	73
Tablo 10 <i>Yetenek Kestirim Yöntemine Göre RMSE, BIAS, MAB ve SH Değerleri</i>	77
Tablo 11 <i>Madde Kullanım Sıklığı, Kullanılmayan Madde ve Test Çakışma Oranları</i>	82
Tablo 12 <i>Madde Takımı Sayısına Göre Madde Kullanım Sıklığı, Kullanılmayan Madde ve Test Çakışma Oranları</i>	83
Tablo 13 <i>Örneklem Büyüklüğüne Göre Madde Kullanım Sıklığı, Kullanılmayan Madde ve Test Çakışma Oranları</i>	84
Tablo 14 <i>Yetenek Kestirim Yöntemine Göre Madde Kullanım Sıklığı, Kullanılmayan Madde ve Test Çakışma Oranları</i>	85

Şekiller Dizini

Şekil 1 Doğrusal Madde Takımı	19
Şekil 2 Madde Takımı Modelinin Yapısı.....	22
Şekil 3 BBT Uygulamalarının Temel İşleyiş Diyagramı	25
Şekil 4 Test Bilgi Fonksiyonu Grafiği	55
Şekil 5 Gerçek Veri Setine ait Yetenek Dağılımı	56
Şekil 6 Yetenek Parametrelerinin Dağılımı	57
Şekil 7 Madde Takımı Sayısına Göre RMSE ve MAB Değerleri	70
Şekil 8 Madde Takımı Sayısına Göre SH ve r Değerleri Grafiği.....	72
Şekil 9 Örneklem Büyüklüğüne Göre RMSE ve MAB Değerleri.....	74
Şekil 10 Örneklem Büyüklüğüne Göre SH ve r Değerleri.....	75
Şekil 11 Yetenek Kestirim Yöntemine Göre RMSE ve MAB Değerleri	78
Şekil 12 Yetenek Kestirim Yöntemine Göre SH ve r Değerleri.....	79
Şekil 13 Madde Takımı Sayısı ve Örneklem Büyüklüklerine Göre Yetenek Ölçeği Boyunca Farklı BBT Desenlerinden Elde edilen BIAS Değerleri-MAP	80

Simgeler ve Ksaltmalar Dizini

BBT: Bilgisayarda Bireyselleştirilmiş Test

OM-BBT: Okuma metni tabanlı Bilgisayarda Bireyselleştirilmiş Test

MT-BBT: Madde takımı tabanlı Bilgisayarda Bireyselleştirilmiş Test

MTK: Madde Tepki Kuramı

MTTK: Madde Takımı Tepki Kuramı

MAB: Mean Absolute Bias

EAP: Beklenen Sonsal Kestirim Yöntemi

MAP: Maksimum Sonsal Kestirim Yöntemi

RMSE: Root Mean Square Error

YDSBS: Yabancı Dil Seviye Belirleme Sınavı

Bölüm 1

Giriş

Çalışmanın bu bölümünde; problem durumu, amaç ve önem, alt problemler, sınırlılıklar ve tanımlar sırasıyla verilmiştir.

Problem Durumu

Eğitim ve psikoloji alanında uygulanan testler aracılığıyla gerçekleştirilen ölçme ve değerlendirme işlemleriyle bireylerin yetenekleri tahmin edilerek amaca uygun doğru kararlar vermek nihai amacımızdır. Bireylerin gerçek yeteneklerinin bilinmesi mümkün olmadığından yapılan tahmin işlemi süreci sonunda gerçek yeteneklerin, en az hata ile kestirilmesi gerekmektedir. Ülkemizde uygulanan ilköğretimden ortaöğretime geçişte LGS Liselere Giriş Sınavı (LGS), ortaöğretimden yükseköğretime geçişte Yükseköğretim Kurumları Sınavı (YKS), Yabancı Dil Testi (YDT), yükseköğretimi tamamlayıp kamuda görev yapmak için Kamu Personeli Seçme ve Yerleştirme Sınavı (KPSS), lisansüstü eğitime devam etmek için Akademik Lisansüstü Eğitim Sınavı (ALES), yabancı dil için YDS Yabancı Dil Sınavı (YDS), Yükseköğretim Kurumları Yabancı Dil Sınavı (YÖKDİL) gibi ulusal sınavlardan hatasız ve güvenilir sonuçlar elde edilmesi önem arz etmektedir. Kâğıt-kalem testi olarak adlandırılan kâğıt üzerinde uygulanan bu sınavlarda adaylardan, belli bir amaca hizmet eden sabit sayıda ve aynı maddelerin yer aldığı formlardaki tüm maddeleri belli bir sürede cevaplandırması istenir. Adayların yetenek düzeyleri göz önünde bulundurulmadan bu maddelere verdikleri doğru cevaplara göre toplam puanları hesaplanır ve bu puana göre değerlendirme yapılır. Adayların yetenek düzeyleri farklı olduğundan bazı maddeler adaylara çok kolay ya da çok zor gelebilir. Böylece bu maddeler çok az bilgi sağlayabilir. Ayrıca Klasik Tepki Kuramı (KTK)'nda yüksek ayırt edicilik ve orta güçlükteki maddelerin seçildiği göz önünde bulundurulursa yetenek düzeyinin uç noktalarında bulunan adayların yetenekleri oldukça hatalı kestirilebilir. Bu durum da adayların gerçek

yeteneklerinin doğru bir şekilde kestirilmesini ve buna bağlı güvenilir değerlendirmeler yapılabilmesini zorlaştırabilir.

Uzun yıllar boyunca uluslararası alanda da öğrenci yeteneklerinin kestirilmesinde, ülkemizdeki gibi kâğıt kalem testleri yani geleneksel testler yaygın olarak kullanılmıştır (Weiss & Kingsbury, 1984; Yan ve diğerleri, 2014). Fakat son yıllarda değişen ve gelişen teknoloji ile bilgisayar kullanımı artmış ve bu durumun sağladığı avantajlar sebebiyle geleneksel kâğıt-kalem testlerine alternatif olarak bilgisayarda uygulanan testlerin çeşitliliği ve sayısında da artış meydana gelmiştir (Aybek & Çıkrıkçı, 2018; Erdem-Kara & Doğan, 2022; Frey ve diğerleri, 2016; İnce-Aracı, 2022; Luecht & Sireci, 2011; Luo & Kim, 2018; Şahin & Gelbal, 2020; Şenel, 2018). Bilgisayarda uygulanan sınavlar; adaylara yer ve zaman konusunda esneklik sağlaması, yenilikçi madde türlerinin uygulanmasına olanak tanınması, kolay puanlanabilir olması, çok sayıda adayın test sonuçlarının daha doğru ve hızlı bir şekilde raporlanmasını kolaylaştırması gibi birçok avantaja sahiptir (Bergstorm & Lunz, 1992; Chalhoub-Deville & Deville, 1999). Özellikle eğitimde uygulanan tüm sınavların bilgisayar tabanlı test ve muadilleri olarak eğitime entegre edilmesi bu noktada oldukça önemlidir. Bu testlerin en önemlilerinden biri hem ulusal hem de uluslararası sınavlarda kullanımı gittikçe yaygınlaşan Bilgisayarda Bireyselleştirilmiş Testler (BBT)'dir. Bu testler, sınava giren bireylerin kendi yetenek düzeyine uygun maddeleri almasını sağlayarak hem testin uzunluğunu hem de süresini kısaltmakta ve kâğıt kalem testlerine göre daha yüksek ölçme verimliliği ve etkililiği sağlayabilmektedir (Wainer ve diğerleri, 2000; Weiss, 1982). Çünkü bireyler, yetenek düzeylerinin çok altında ya da üstünde olan maddelerle karşılaşmadıklarından, onların yetenekleri hakkında daha fazla bilgi sağlanabilmektedir. Ayrıca anında dönüt vermesi, test eşitlemeye ihtiyaç duyulmadan puan karşılaştırmalarının yapılması, adaya istediği zaman sınava girme esnekliği sağlaması gibi avantajlara da sahiptir (Patsula, 1999; Wainer ve diğerleri, 2000). Madde Tepki Kuramı (MTK) temelli BBT uygulamalarında, tüm bu avantajlara rağmen bazı özel durumlarda sorunlar ile karşılaşlabilmektedir. Bu sorunlardan biri de bu çalışmanın odak noktası olan, ortak bir

uyarana baėlı madde gruplarından oluřan ve madde takımı (testlet) adı verilen bu madde gruplarının BBT uygulamalarında nasıl kullanılacaėıdır.

İlk kez Wainer (1990) tarafından ortaya atılan madde takımı, aynı ieriėe baėlı birden fazla maddeden oluřan yapılardır. Bu yapılar grafik, tablo, diyagram, harita, okuma metni ve senaryo gibi herhangi bir ortak uyarana atıfta bulunan yeniliki maddelerden oluřabilir (Kang ve diėerleri, 2022; Wang & Wilson, 2005). Eėitim ve psikoloji alanında zellikle karmařık biliřsel sreler ieren okuduėunu anlama becerilerinin llmesinde, her biri madde demetleri/kmeleri hlinde gruplanmıř maddelerden oluřan testler kullanılır (Rosenbaum, 1988). Bu baėlamda, ortak bir madde kkn paylařan maddeler; madde demeti, madde takımı ya da madde kmesi olarak adlandırılır. Bu madde takımları ile iliřkili bir uyarana sahip ierik hakkında birden fazla soru sorularak hem zaman ve aba aısından ekonomiklik saėlanır hem de st dzey becerilerin daha iyi llebilmesi hedeflenir (Demars, 2010; Wainer & Wang, 2000). Bu nedenle okuduėunu anlama becerisinin llmesi ve deėerlendirilmesinde, bir veya birden fazla okuma metnine atıfta bulunan bir dizi madde takımlarının kullanımı sıklıkla tercih edilir.

Tm avantajlarına raėmen oktan semeli maddelere dayalı madde takımlarından elde edilen puan yorumları geerlik aısından eleřtirilmektedir (Katz & Lautenschlager, 2001). zellikle madde takımlarının kullanımı, yerel baėımsızlık varsayımının ihlli nedeniyle MTK iin bir zorluk teřkil eder. Yerel baėımsızlık, sınava giren kiřinin yeteneėine baėlı olarak, bir maddeye doėru yanıt verme olasılıėının diėer herhangi bir maddeye doėru yanıt verme olasılıėından istatistiksel olarak baėımsız olduėu anlamına gelir (Hambleton & Swaminathan, 1985). Bu nedenle yerel baėımsızlıėın ihlli maddeler arası bir baėımlılık olduėuna iřaret eder. Birok geleneksel psikometrik yaklařıma alternatif olarak bu ortak metin ieren maddeler arasındaki baėımlılıėın modellenmesi gerekmektedir (Baldonado ve diėerleri, 2015). nk yerel madde baėımlılıėı, bir madde takımı iinde yer alan maddelerin ana zelliėidir (Wilson, 1988). Yerel baėımlılıėın varlıėı ise lm kesinliėinin fazla ya da dřk kestirilmesine (Marais & Andrich, 2008; Sireci, Thissen, & Wainer, 1991; Wainer &

Thissen, 1996; Yen, 1993), madde parametre tahmininde yanlışlığa (Ackerman, 1987; Tuerlinckx & De Boeck, 2001; Wainer & Wang, 2000), madde uyumsuzluğuna (Marais & Andrich, 2008), bağlama ve eşitleme hatalarına (Lee ve diğerleri, 2001; Li ve diğerleri, 2006) sebep olmaktadır. Bu nedenle yerel bağımlılığın kontrol altına alınması, geçerli ve güvenilir kestirimler yapılabilmesi açısından önemlidir.

Madde takımlarından oluşan testlerde yerel bağımlılık ile başa çıkmak için birkaç yaklaşım önerilmektedir. Bunlardan biri madde takımlarını çok kategorili puanlanan bir madde olarak ele almaktır. Bu yaklaşım ile madde takımı sayısı kadar çok kategorili puanlanmış madde elde edilir. Birçok BBT uygulama çalışmalarında bu yöntem tercih edilmesine rağmen (Çoban, 2020; Dodd ve diğerleri, 1995; He & Min, 2024; Jiao ve diğerleri, 2012; Jodoin ve diğerleri, 2006; Penfield, 2006; van Rijn ve diğerleri, 2002) madde takımının çok kategorili tek bir madde (süper madde) olarak bu şekilde tanımlanmasının yerel bağımlılık sorunu ile yeterince baş edememesine neden olabilmektedir. Çünkü bu yaklaşım ile yerel bağımlılık miktarı dikkate alınmamaktadır. Ancak Wainer (1995)'a göre bir madde takımı içindeki maddeler arasındaki yerel bağımlılık orta düzeydeyse ve test büyük oranda bağımsız maddeler içeriyorsa bu yaklaşım uygun olabilir. Diğer önemli bir sorun ise bir madde takımındaki toplam puanların aynı olması durumunda aynı kategoride yer almasıdır. Bu durumda, bir madde takımı içindeki maddelerin bilgileri eksik hesaplanabilir ve bu durum da bilgi kaybına yol açabilir (Wainer & Lewis, 1990). Bu bilgi kaybını önlemek için yerel bağımlılık miktarlarını dikkate alan ve her madde takımı boyunca her bir bireye özgü yerel bağımlılık miktarını belirten ek bir parametrenin yer aldığı Madde Takımı Tepki Kuramı (MTTK) gibi karmaşık modellere ihtiyaç vardır (Wainer ve diğerleri, 2000).

Alan yazında madde takımlarından oluşan testlerin BBT uygulamalarında, maddelerin bağımsız birer madde olarak ayrı ayrı ele alındığı çalışmalara da rastlanmaktadır (Alkan, 2023; Kaya, 2022, Murphy ve diğerleri, 2010). Bu durumun asıl sebebi MTTK modellerinin daha zor ve karışık olmasından dolayı araştırmacılar tarafından fazla tercih edilmemesidir. Bir diğer önemli neden ise bu yöntemin madde havuzundaki

madde sayısını artırarak BBT performansının daha iyi sonuçlanmasını sağlamaktır. Fakat madde takımları içindeki her bir madde ortak bir uyarıcı (metin, şekil, grafik vb.) içerdiklerinden maddeler bağımsız ele alındığından her bir madde bu ortak uyarıcıyı da içerecektir. Bu durumda gerçek zamanlı BBT uygulaması yapıldığında, bireyler ortak uyarıcıya birden fazla kez maruz kalacak bu durum da BBT uygulamalarının ekonomiklik ilkesi ile çelişecektir. Yani bireyler sınav anında ortak uyarıcı ile birden fazla kez karşılaşabileceklerinden bu durum hem zaman kaybına hem de fazladan okuma yüküne sebep olacaktır. Bir de aynı metin ile birden fazla kez karşılaşmak da bu uyarıcıya aşinalığı artıracak bu da maddelerin güçlüğü gibi psikometrik özelliklerini etkileyecektir.

Özellikle dil becerilerinin ölçülmesinde ortak uyarıcıya bağlı madde takımları sıklıkla kullanılmaktadır. Örneğin, Educational Testing Service (ETS)'den Yabancı Dil Olarak İngilizce Testi (TOEFL), Cambridge Üniversitesi Yerel Sınavlar Sendikası (UCLES)'nin Cambridge sınavları, Uluslararası İngilizce Dil Test Sisteminin (IELTS) tümü ve ülkemizde yapılan YDS, YÖKDİL gibi sınavlarda tek bir okuma metnine bağlı birden fazla madde kullanılmaktadır. Artık tüm dünyanın BBT desenlerinin avantajlarından yararlanabilmek için uyarılabilir test uygulamalarına geçtiği bu dönemde, bu alanda yapılan çalışmaların sayısı artış göstermektedir. Hatta ülkemizde de bu geçiş sürecinin hazırlıklarını ilk olarak YÖKDİL ve YDS gibi yabancı dil sınavlarının elektronik ortamda bilgisayar tabanlı olarak uygulanmaya başlanmasıyla görmekteyiz. Fakat MTK temelli BBT uygulamalarında standart MTK modelleri, madde takımı etkilerinin göz ardı edilmesinden dolayı diğer madde takımlarında olduğu gibi bu İngilizce okuma metni içeren madde takımlarıyla da düzgün çalışmayabilir ve parametre tahminlerinde bozulmalara ve güvenilirliğin fazla tahmin edilmesine yol açabilir (Li ve diğerleri, 2006). Çünkü bir madde takımı içindeki madde yanıtlarının tamamen bağımsız olmadığı, ortak uyarıcı aracılığıyla ilişkili bu maddeler arasında yerel bağımlılığın olabileceği unutulmamalıdır. Böylece madde takımlarından oluşan BBT uygulamalarında tek boyutlu MTK modellerinin kullanılması da yerel bağımlılık varsayımının ihlâlinden dolayı madde/madde takımı bilgi fonksiyonlarının yüksek

hesaplanmasına sebep olmaktadır (Thissen ve diğerleri, 1989). Alan yazında İngilizce seviye belirleme sınavlarının BBT uygulamalarını yapmak amacıyla farklı modellerin ve desenlerin uygulanarak elde edilen sonuçların karşılaştırılmasını amaçlayan çalışmalara ihtiyaç vardır.

Madde takımlarından oluşan testlerin BBT uygulamalarına nasıl entegre edileceği önemli bir sorundur. Alan yazında ise madde takımı tabanlı BBT desenlerinin uygulandığı birden fazla yöntem bulunmaktadır. Bunlardan en sık kullanılanı ise madde takımıdaki her bir maddeyi ortak uyararla birlikte ayrı ayrı ele alan BBT uygulamalarıdır. Çünkü BBT uygulamaları madde tabanlı olduğu için madde seçiminde tek bir maddeye dayalı yetenek kestirimi yapabilme algoritmasını içermektedir. Bu nedenle genellikle BBT uygulamalarında madde takımları sağladığı avantajlardan ötürü (madde havuzu büyüklüğü, kolay hesaplanabilir olması vb.) bağımsız birer madde olarak düşünülerek ele alınmaktadır. Fakat bu durum gerçek zamanlı uygulanan BBT çalışmalarının mantığı ile çelişmekte ve yerel bağımsızlık ilkesi ihlâlinden dolayı yanlış yetenek kestirimlerine yol açabilmektedir. Madde takımı tabanlı BBT uygulama yöntemlerinden diğeri ise madde takımlarını bir bütün halinde düşünerek tek bir madde değil madde takımına verilen cevaba göre yetenek kestirimlerinin yapıldığı fakat yine tek boyutlu MTK modellerinin kullanıldığı ortak metin tabanlı BBT uygulamalarıdır (OM-BBT). Ek algoritmalar yazılarak geliştirilen bu BBT deseni ile ortaya çıkan zaman ve çaba açısından ekonomiklik sorununa çözüm getirilmiştir. Çünkü bu desen ile bireyin tek bir madde takımına verdiği cevaptan sonra yetenek kestirimi yapılmakta ve buna göre yeni bir madde takımı verilmektedir. Bu durumda aynı okuma metni ya da grafik, şekil vb. uyarılar bireyin karşısına sadece bir kez çıkarak zaman ve emek kaybı oluşmamaktadır. Fakat bu desen madde takımları arasındaki bağımlılığı göz ardı ederek MTK'nin yerel bağımsızlık ihlâli sorununa çözüm getirememektedir. Madde takımları arası yerel bağımlılığı da hesaba katarak MTTK modeline göre yetenek kestirimlerinin yapıldığı madde takımı tabanlı BBT (MT-BBT) uygulama deseni ise diğeri bir yöntemdir.

Özet olarak madde takımlarından oluşan bir BBT uygulamasının nasıl dizayn edileceği ve modelleneceği ile ilgili çalışmalar alan yazında yeterli değildir. Bu nedenle bu çalışmada öncelikle yabancı dil sınavından elde edilen gerçek bir veri seti üzerinden madde takımlarının bağımsız birer madde olarak ele alındığı BBT deseni, madde takımlarının birlikte ve madde takımı tepki modelinin temel alındığı madde takımı tabanlı BBT deseni (MT-BBT) ve son olarak madde takımlarının yani okuma metinlerinin yine birlikte ele alındığı fakat tek boyutlu madde kuramı modeli temelli okuma metni tabanlı BBT deseni (OM-BBT) olmak üzere üç farklı BBT deseni oluşturulmuştur. Farklı koşullar eklenerek bu üç farklı BBT deseninden elde edilen ölçme kesinliği ve madde havuzu kullanımına ilişkin sonuçlar karşılaştırılmış ve detaylı bir şekilde incelenmiştir.

Araştırmanın Amacı ve Önemi

Bu araştırmanın amacı madde takımlarından oluşan bir İngilizce seviye belirleme testini farklı BBT desenleri ve test koşulları oluşturarak elde edilen ölçme kesinliği ve madde havuzu kullanımına ilişkin sonuçları karşılaştırmaktır. Gelişen teknoloji ile dünyada kâğıt kalem testlerinin yerini yetenek kestirimi, test süresi, değerlendirme süreci, madde takımı sayısı vb. konularda sağladığı avantajlar sebebiyle BBT uygulamaları almaya başlamıştır. Ülkemizde ise bilgisayar tabanlı sınav uygulamalarına başlanmış olup (e-YDS, e-YÖKDİL) BBT uygulamalarına da yakın gelecekte başlanılacağı düşünülmektedir. Bu nedenle bu geçiş sürecinde madde takımı içeren BBT uygulamalarını temel alan çalışmaların yapılması oldukça önem teşkil etmektedir.

Madde Tepki Kuramı (MTK) temelli BBT uygulamalarının tüm avantajlarından faydalanabilmek için BBT uygulamalarının tüm aşamalarının titizlikle yapılması gerekmektedir. Madde havuzuna uygun MTK modeli ve madde seçim yöntemi kullanılması, geçerli ve güvenilir sonuçlar elde edilmesini sağlar. Özellikle madde takımlarından oluşan bir madde havuzunun BBT uygulamalarında nasıl kullanılacağı önemli bir sorundur. Alan yazında yer alan çalışmalar madde takımlarından oluşan madde havuzundaki maddeleri

bağımsız birer madde olarak değerlendirmişlerdir (Alkan, 2023; Kaya, 2022; Murphy ve diğerleri, 2010). Bunun en büyük sebebi geniş ve yeterli madde havuzuna sahip BBT uygulamalarının etkili kestirimler yapabilmesidir. Simülasyon çalışmalarında madde etkili kestirimler yapabilmek için bağımsız maddeler kullanılabilir fakat gerçek zamanlı BBT uygulamalarında bu durum özellikle BBT'nin ekonomiklik özelliğine aykırı olup bireylerin bir ortak metni birden fazla maddede görmesine sebep olur. Ayrıca yerel bağımlılığın özellikle yüksek olduğu madde takımları nedeniyle yanlış belirlenen parametrelerin kullanılması veya uygulama aşamasında yanıtlardaki yerel bağımlılığın kontrol edilmemesi durumunda, BBT puanlarının doğruluğu olumsuz etkilenebilmektedir (Yao, 2019). Bu nedenle madde takımlarından oluşan BBT çalışmalarında madde takımlarının nasıl ele alınacağı ve bu durumda madde havuzuna uygun MTK modelinin nasıl seçileceği oldukça önemlidir.

Bilgisayarda bireyselleştirilmiş testlerde madde takımı tabanlı maddelerin uygulanması, geleneksel iki veya çok kategorili MTK modellerinden farklı olarak daha karmaşık tasarlanmış psikometrik modeller gerektirir. Bu nedenle, madde takımı temelli maddeler BBT uygulamalarına bir parça veya bütün olarak dâhil edildiğinde, MTTK modellerinin gizil özellikleri tahmin etmede geleneksel MTK modellerine göre daha iyi performans gösterdiği öne sürülmektedir (DeMars, 2010; Wang ve diğerleri, 2002; Yao, 2019; Zhang, 2010). Bunun yanı sıra madde takımı tabanlı BBT uygulamalarında MTTK modellerinin kullanılmasının MTK modelleri ile benzer sonuçlar elde edildiğine dair alan yazında çalışmalar da mevcuttur (Murphy ve diğerleri, 2010; Sireci ve diğerleri, 1991; Wainer ve diğerleri, 2007; Wainer & Thissen, 1996; Yen, 1993). Bu nedenle bu çalışma ile gerçek veri setinden üretilmiş madde havuzu yardımıyla madde takımlarından oluşan BBT uygulamaları farklı model ve test koşulları altında ölçme kesinliği, madde havuzu kullanımı açısından karşılaştırılmıştır. Farklı BBT desenlerinin performansını incelerken gerçek veri seti ve madde parametrelerinin kullanılması, sınava giren bireyler hakkında daha doğru psikometrik özellikler elde edilmesini sağlamaktadır (Wang ve diğerleri, 1999). Madde takımı tabanlı BBT uygulamaları, yeni bir yaklaşım olduğundan bu alanda yapılan

alıřmaların uluslararası literatürde sınırlı olduėu görölmektedir (Choi & van der Linden, 2018; Keng, 2008; Zheng & Chang, 2015). Hatta ölkemizde madde takımı BBT uygulamaları ile ilgili herhangi bir alıřmaya rastlanmamıř olup ulusal alan yazında da bu konuyla ilgili eksiklik olduėu dikkat çekmektedir. Arařtırmada ele alınan ü farklı desen karřılařtırmasına ilk kez tek bir alıřmada yer verildiėinden bu alıřmanın özgün etkileri olacaėı hedeflenmektedir. Ayrıca gerek veri setinden elde edilen parametreler kullanıldıėından, alıřmanın uygulanabilirliėi ve bu alanda yapılacak benzer alıřmalara öncü olması aısından alan yazına önemli katkı sunacaėı düşünölmektedir.

Arařtırma Problemi

Madde takımlarından oluřan BBT uygulamalarında, farklı madde takımı sayısı, yetenek kestirim yöntemleri ve örnekleme büyüklüėü kořulları altında BBT, MT-BBT ve OM-BBT desenlerinin ölçme kesinliėi ve madde havuzu kullanımı nasıl deėiřmektedir?

Alt Problemler

- 1) Madde takımı sayısına göre BBT, MT-BBT ve OM-BBT desenlerinin gerek ve kestirilen yetenek düzeyler arasındaki korelasyon, RMSE, BIAS, SH deėerleri nasıl deėiřmektedir?
- 2) Örnekleme büyüklüėüne göre BBT, MT-BBT ve OM-BBT desenlerinin gerek ve kestirilen yetenek düzeyler arasındaki korelasyon, RMSE, BIAS, SH deėerleri nasıl deėiřmektedir?
- 3) Yetenek kestirimi yöntemine göre BBT, MT-BBT ve OM-BBT desenlerinin gerek ve kestirilen yetenek düzeyler arasındaki korelasyon, RMSE, BIAS, SH deėerleri nasıl deėiřmektedir?
- 4) Farklı test uzunluklarında BBT, MT-BBT ve OM-BBT desenlerinin madde kullanım sıklıėı, kullanılmayan madde sayısı ve test akıřma oranları nasıl deėiřmektedir?

- 5) Farklı örneklem büyüklüğünde BBT, MT-BBT ve OM-BBT desenlerinin madde kullanım sıklığı, kullanılmayan madde sayısı ve test çakışma oranları nasıl değişmektedir?
- 6) Farklı yetenek kestirimlerinde BBT, MT-BBT ve OM-BBT desenlerinin madde kullanım sıklığı, kullanılmayan madde sayısı ve test çakışma oranları nasıl değişmektedir?

Sınırlılıklar

Bu çalışma, her bir veri setindeki veri sayısının az olmasından dolayı Rasch model temelli MTK ve MTTK modelleri ile sınırlıdır.

Tanımlar

Madde Takımı: Tek bir uyarana bağlı birden fazla maddeden oluşan setler.

Bilgisayarda Bireyselleştirilmiş Test (BBT): Bireyin testte yer alan her bir maddeye verdiği cevaptan sonra yetenek kestiriminin gerçekleştiği ve bu geçici yetenek düzeyine uygun maddelerin geldiği belli bir algoritmaya dayalı test desenidir.

Madde Takımı Tepki Kuramı: Madde takımı etkisini dikkate alan ve bifaktör modelin sınırlandırılmış hali olan madde tepki kuramı modelidir.

Madde Takımı tabanlı Bilgisayarda Bireyselleştirilmiş Test (MT-BBT): Madde takımlarından oluşan ve madde takımı seçiminde MTTK modelinin BBT uygulamasıdır. Madde takımlarından oluşan ve madde seçim kriteri olarak madde takımına ait bilgilerin kullanıldığı MTTK modeli temelli BBT uygulamasıdır.

Ortak Metin tabanlı Bilgisayarda Bireyselleştirilmiş Test (OM-BBT): Tek bir içeriğe bağlı okuma parçalarını içeren madde takımlarından oluşan, madde takımı seçimine dayalı MTK modelinin kullanıldığı BBT uygulamasıdır.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde öncelikle araştırmada yer alan BBT uygulamalarının psikometrik anlamda temelini oluşturan MTK, iki kategorili MTK modelleri, madde takımı kavramı ve madde takımlarına uygulanan MTTK modelleri açıklanmıştır. Ardından BBT uygulamaları, aşamaları ve yaklaşımları tanıtılmıştır. Son olarak da madde takımı tabanlı BBT uygulamaları ile ilgili araştırmalara yer verilmiştir.

Madde Tepki Kuramı

Eğitim alanında ölçme ve değerlendirmeler, Klasik Test Kuramı (KTK), Genellenebilirlik Kuramı ve Madde Tepki Kuramı (MTK) temel alınarak yapılır. Her bir bireyin örtük özelliğini doğru ve güvenilir biçimde tanımlama amacı güden bu kuramlardan ilk geliştirileni ve özellikle sınıf içi ölçmelerde en çok kullanılanı KTK'dir (Crocker & Algina, 1986). KTK uygulamalarının diğer ölçme kuramlarına göre daha pratik olması ve daha az varsayım gerektirmesinden dolayı tercih edilmektedir. Diğer taraftan birçok noktada da birçok eleştiriye maruz kaldığı aşikârdır. En sık yapılan eleştirilerden bazıları, KTK'nin bir birey üzerinde yapılan tekrarlı ölçmelere dayandırılması, bu kuramdan elde edilen madde ve test istatistiklerinin testin uygulandığı gruba bağlı olması ve bütün yetenek aralığı için tek bir hata kestirimi yapılması gibi sınırlılıklardır (Crocker & Algina, 1986; Lord & Novick, 1968).

KTK'nin yukarıda ifade edilen sınırlılıklarına çözüm sunabilmek amacıyla geliştirilen ve kullanımı hızla yaygınlaşmış bir diğer ölçme kuramı Örtük Özellikler Kuramı olarak da bilinen MTK'dır. MTK, ölçülen gizil özellik ya da yetenek ile test maddesine verilen tepki arasındaki ilişkiyi tanımlayan matematiksel bir model kurarak bireyin tepkileri sayesinde gizil özelliğini ya da yeteneğini kestiren bir kuramdır (Embretson & Reise, 2000). Bireyin belli bir maddede göstermiş olduğu performans ile bunun altındaki özellikler kümesiyle arasındaki ilişki madde karakteristik fonksiyonu (MKF), bu fonksiyonun grafiği de madde karakteristik eğrisi (MKE) ile tanımlanır (Demars, 2010).

MTK; yetenek düzeyi arttıkça, bir maddenin doğru cevaplandırılma olasılığının da arttığı varsayımı üzerine kurulmuştur (de Ayala, 2009). MTK'de yetenek ya da özellik olarak adlandırılan puanlar θ ile gösterilir ve θ eşit aralıklı bir ölçekte $-\infty$ ile ∞ arasında değer alabilirken genellikle uygulama kolaylığı açısından $(-3,+3)$, $(-3.5, 3.5)$ gibi aralıklar tercih edilir. Madde karakteristik eğrisinde θ , bireyin madde özelliklerinden ve bireyin maddeye verdiği cevaptan kestirilen yetenek düzeyidir ve her zaman testteki maddelerden bağımsızdır (van der Linden & Hambleton, 1997). Yani MTK'nin testin bütününden ziyade maddelere odaklandığı söylenebilir. Bu da birey-madde etkileşiminin sağlanmasında oldukça avantaj sağlamaktadır. Diğer bir önemli avantajı, madde parametre kestirimlerinin, testin uygulandığı örneklemden bağımsız olmasıdır.

MTK Varsayımları

MTK'nin en önemli iki varsayımı, tek boyutluluk ve yerel bağımsızlıktır (Embretson & Reise, 2000). Tek boyutluluk varsayımında, bireylerin test maddeleriyle olan etkileşimlerinin birey özelliğini betimleyen tek bir faktör ile yeterli biçimde temsil edilebileceği kabul edilmektedir (Crocker & Algina, 1986). Bir testin tek boyutluluk varsayımının tamamen karşılanması, yani sadece tek bir özelliği ölçmesi ve başka hiçbir özelliği ölçmemesi oldukça güçtür. Ancak yine de genellikle testlerle ölçülen baskın bir faktör bulunmaktadır. Bu baskın faktör, testle asıl ölçülmek istenen özellik olarak tanımlanır. Tek boyutlu MTK modelleri testle baskın bir faktörün ölçüldüğünü, bu faktörün bireyin test performansını etkileyen ana özellik olduğunu kabul eder (Hambleton & Swaminathan, 1985). Hatta testin çok boyutlu olmasına rağmen tek bir baskın faktöre sahip olabileceği istatistiksel olarak hiyerarşik omega (ω_H) katsayısı ile hesaplanabilir (Rodriguez ve diğerleri, 2016). Hatta bu varsayımın adı çok boyutlu modellerin yaygın olarak kullanılmaya başlanmasıyla birlikte boyutluluk adını almıştır. Boyutluluk varsayımı; faktör yüklerinin özdeğerleri, yamaç birikinti grafiği, paralel analiz ve artıkların analizi gibi yöntemlerle incelenebilir.

Yerel bağımsızlık varsayımı, belirli yetenek düzeyinde herhangi bir maddeye verilen tepkinin başka bir maddeye verilen tepkiden istatistiksel anlamda bağımsız olması yani o

maddeye verilen cevabın diğer maddeleri herhangi bir şekilde etkilememesidir (Hambleton & Swaminathan, 1985). Bu varsayım sağlanmadığında, ham puanlar ile madde ve yetenek parametresi kestirimleri de hatalı olur ve maddeler yerel bağımlı duruma gelir. Örneğin bir madde, testteki diğer maddelerin doğru bir şekilde cevaplandırılmasına etki edecek bilgiler içerdiğinde ya da testte ortak bir uyarana bağlı birden fazla madde yer aldığına ise bu varsayım ihlâl edilebilmektedir (Hambleton ve diğerleri, 1991; Sireci ve diğerleri, 1991). Yerel bağımsızlık varsayımı ihlâli problemi ile baş etme yollarından biri bağımlılığı göz ardı etme ve uygun bir MTK modeli ile analizlere devam etmektir. Diğerleri ise bağımlılığı dikkate alarak çok kategorili MTK modelleri, bifaktör modeli ve madde takımı MTK modelini kullanmaktır.

Yerel madde bağımlılığının belirlenmesinde madde çiftleri arasından hesaplanan $Q3$, χ^2 ve G istatistiği gibi yöntemler geliştirilmiştir (Yen, 1993; Chen & Thissen, 1997). $Q3$ istatistiğine göre madde çiftleri arasında hesaplanan artıkların korelasyonu $[0,2]$ 'den büyük olduğunda yerel bağımlılığa dikkat edilmelidir (Yen & Fitzpatrick, 2006). $LD \chi^2$ istatistiğine göre ise 4'ü aşan değerler maddeler arasında yerel bağımlılığı, 10'u aşan değerler ise maddeler arasında yüksek derecede yerel bağımlılığı gösterir (Min & He, 2014).

Ek olarak hız testi olma durumu, yetenek ve madde parametrelerinin değişmezliği diğer MTK varsayımları olarak sıralanabilir (Hambleton & Swaminathan, 1985). MTK modellerinin varsayımlarının oldukça güç ve karşılanması zor olduğu görülmektedir. Bu varsayımların sağlanmadığı durumlarda model veri uyumu zayıf olabilir ve uygulanan modelin geçerliği pek sağlanamayabilir (Hambleton & Swaminathan, 1985). Etkili sonuçlar elde etmek için model veri uyumunun sağlandığı modeller seçilerek uygulamalara devam edilmelidir.

Madde Tepki Kuramı Modelleri

MTK'de kendine özgü MKF ya da MKE ile tanımlanan çok sayıda model vardır. Eğitimde kullanışlılığı sebebiyle yaygın olarak çoktan seçmeli maddeler kullanıldığından

MTK modellerinden iki kategorili modeller sıklıkla tercih edilir. İki kategorili MTK modelleri tipik olarak çoktan seçmeli maddelere sahip testlerle ilişkilendirilir ancak aynı zamanda madde yanıtları “doğru-yanlış” veya “katılıyorum-katılmıyorum” gibi ikili kategoriler hâlinde sınıflandırılabilen herhangi bir teste de uygulanabilir (Embretson & Reise, 2000). Modeller çeşitli biçimlerde sınıflanabilir ancak maddeyi tanımlarken kullanılacak parametre sayısı en önemli ayrımdır. En yaygın iki kategorili MTK modelleri; tek parametrelili lojistik (1PL; Rasch, 1960), iki parametrelili lojistik (2PL; Birnbaum, 1968) ve üç parametrelili lojistik (3PL; Birnbaum, 1968) modelleridir.

En basit iki kategorili MTK modeli olan Standart Rasch modeli, ilk olarak Georg Rasch (1960) tarafından geliştirilmiştir ve bu modelde madde ayırt edicilikleri 1’e eşit, şans parametresi de 0 olarak kabul edilir. Bir bireyin bir maddeyi doğru cevaplama olasılığı, bir maddenin zorluğuna ve sınava giren kişinin yeteneğine göre belirlenir. Bu modele göre, yeteneği madde güçlüğüne eşit olan herhangi bir kişi için maddeyi doğru yanıtlama olasılığı 0,5’tir. Teknik olarak Rasch model için tüm maddelerde ayırt edicilik değeri 1’dir ve ölçek sabiti olan D parametresi modeli anlaşılır kılmak için modelden tümüyle çıkarılmıştır (DeMars, 2010).

Rasch MTK modelin denklemi:

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

(1) denkleminde yer alan ifadeler;

$P_i(\theta)$: θ yetenek düzeyindeki bireylerden random olarak seçilen bir bireyin i maddesini doğru cevaplandırma olasılığı,

b_i : i maddesinin madde güçlük parametresi

e : Değeri yaklaşık 2,718 olan matematiksel sabit

Rasch model en basit model olduğundan parametre tahmini için diğer MTK modellerine göre daha küçük bir örneklem büyüklüğü gerektirir. Ayrıca Rasch modelinde

maddeler sabit etkiler olarak ele alındığından madde parametrelerine ilişkin dağılımsal varsayımlara gerek duyulmamaktadır (Wang & Wilson, 2005). Analizdeki göreceli basitlik ve verimlilik nedeniyle Rasch modelleri pratikte yaygın olarak kullanılmaktadır.

Standart Rasch modeli (Rasch, 1960), $P(\theta)$ elde etmenin bir fonksiyonu olarak gizil yetenek θ ile güçlük parametresi olan b arasındaki farkla ilgilidir ve onu diğer 1PL modellerinden ayırt edilebilir kılan iki kısıtlamaya sahiptir: Bunlardan biri tüm güçlük parametrelerinin toplamı sıfıra eşittir. Diğeri aynı sayıda doğru cevaba sahip bireylerin, aynı gizil yeteneğe (θ) sahip olmasıdır. Sonuç olarak toplam puan, Rasch modeli altında sınava giren kişinin gizli yeteneği yani θ 'yı tahmin etmek için yeterli bir istatistik olabilir (Rasch, 1960). Bu iki kısıtlama göz önüne alındığında, standart Rasch modeli eğitim uygulamalarında oldukça pratik ve gerçek test ortamında daha fazla uygulanabilirliğe sahiptir.

Bir parametrelili modelde (1PLM) tüm maddeler için madde ayırt ediciliklerinin eşit olduğunu varsayar ve (1) denkleminde $P(\theta)$ kestiriminde madde ayırtıcılık parametrelerinin ortalaması ve D sabiti eklenir ve 1PLM denklemi:

$$P_i(\theta) = \frac{e^{D\bar{a}(\theta-b_i)}}{1+e^{D\bar{a}(\theta-b_i)}} \quad (2)$$

$\bar{a}$: Testteki tüm maddelerin ayırt ediciliklerinin ortalaması

D: Ölçek sabiti (1,702)

İki parametrelili modelde (2PLM; Birnbaum, 1968) (2) denkleminde her bir maddenin ayırt edicilik parametresi eklenir ve 2PLM denklemi:

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \quad (3)$$

a_i : i maddesinin madde ayırt edicilik parametresi

Denklem (3)'te görüldüğü gibi her bir maddenin ayırt ediciliği hesaplanarak bu parametre değeri yetenekten güçlük parametresinin çıkartılmasıyla elde edilen değere bir

çarpan olarak eklenmiştir. Böylece ayırt ediciliği yüksek olan maddeler, yetenek ve güçlük parametresi arasındaki farkı ve buna bağlı olarak maddenin doğru cevaplama olasılığını etkilemektedir.

Üç parametrelili modelde (3PLM; Birnbaum, 1968) (3) denkleminde her bir maddenin şans/tahmin parametresi eklenir ve 3PLM denklemi:

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad (4)$$

c_i : i maddesinin madde şans parametresi

3PLM'de yer alan madde şans parametresi ile düşük yetenek düzeyine sahip bireylerin bir maddeyi doğru yanıtlama olasılığı hesaplanır.

Özet olarak iki kategorili lojistik modellerin en genel hali 3PLM'dir. $c_i=0$ olduğunda, 2PLM elde edilir. 2PLM'ye a_i değerlerinin eşit olduğu varsayımı eklenirse, 1PLM elde edilir. Tüm maddelerin ayırt edicilikleri 1'e eşitlenirse de Rasch model elde edilir. Veri setinin hangi model ile analiz edileceğine; gerekli sayıtların incelenmesi, model veri uyumu ölçütlerinin değerlendirilmesi, parametre kestirimlerinin amacının belirlenmesinden sonra karar verilir.

Eğitim ve psikolojide doğru-yanlış gibi iki kategorili puanlanamayan ikiden çok kategorili kısmi puanlanan maddeler de mevcuttur. Çok kategorili olarak puanlanan bu tür maddeler için Aşamalı Tepki Modeli (ATM[Graded Response Model-GRM]), Kısmi Puan Modeli (KPM[Partial Credit Model-PCM]), Genelleştirilmiş Kısmi Puan Modeli (GKPM[Generalized Partial Credit Model-GPCM]), Derecelendirme Ölçeği Modeli (DÖM[Rating Scale Model-RSM]), Adlandırmalı Tepki Modeli (ATM[Nominal Response Model-NRM]) (Embretson & Reise, 2000, Samejima, 1969) ve ortak bir uyarana bağlı birden çok maddenin yer aldığı ortak köklü maddelerde de yerel bağımsızlık ihlâlüne karşın bu çok kategorili MTK modelleri kullanılabilir.

Madde ve Test Bilgi Fonksiyonu

MTK'de bilgi fonksiyonları, belli bir maddenin herhangi bir yetenek düzeyinde ne kadar etkili ölçüm yapılabildiğini tanımlar. Yetenek ölçeğinin her bir noktasında bir maddenin verdiği bilginin hesaplanması ölçmenin kesinliğinin yorumlanması açısından önemlidir. Bir maddenin yetenek düzeyinin herhangi bir noktasında MTK'de hem ölçmenin standart hatası hem de güvenilirlik bilgi fonksiyonları kullanılarak hesaplanır (Demars, 2010). Maddelerden elde edilen yüksek bilgi değeri parametre kestirimi hakkında daha fazla bilgiye sahip olunduğunu ifade eder (Hambleton ve diğerleri, 1991).

$$I_i(\theta) = \sum_{i=1}^n \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad (5)$$

$$Q_i(\theta) = 1 - P_i(\theta)$$

$I_i(\theta)$: i. maddenin madde bilgi fonksiyonu

$P_i(\theta)$: Bireyin i maddesini doğru yanıtlama olasılığı

$P'_i(\theta)$: Bireyin i maddesini doğru yanıtlama olasılığının türevi

$Q_i(\theta)$: Bireyin i maddesini yanlış yanıtlama olasılığı

Bilgi madde düzeyinde elde edilir fakat bilgi fonksiyonu test düzeyinde olduğundan madde bilgi fonksiyonlarının toplanmasıyla test bilgi fonksiyonları hesaplanır. Bilgi fonksiyonları her bir yetenek düzeyine göre hesaplandığından bir madde için tek bir değer ifade etmez.

$$I_T(\theta) = \sum I_i(\theta) \quad (6)$$

$I_T(\theta)$: Test bilgi fonksiyonu

Testteki madde sayısı da doğal olarak test bilgi fonksiyonunu etkilemektedir; daha fazla madde, daha yüksek test bilgi fonksiyonu elde edilmesini sağlar.

Çok kategorili olarak puanlanan maddeler için ise her bir kategoriye ait maddenin sağladığı bilgi bir fonksiyon aracılığıyla hesaplanmaktadır. Bu çok kategorili maddeler için madde bilgi fonksiyonu denklemi;

$$I_i(\theta) = \sum I_{ix}(\theta)P_{ix}(\theta) \quad (7)$$

$I_{ix}(\theta)$: i. maddenin kategori bilgi fonksiyonu

$P_{ix}(\theta)$: i. maddenin her bir kategorisi için madde tepki fonksiyonu

Test bilgi fonksiyonu da iki kategorili maddelerde olduğu gibi madde bilgi fonksiyonlarının toplamı olarak hesaplanmaktadır.

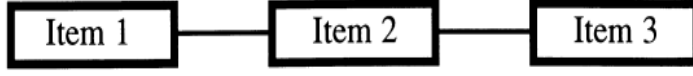
Madde Takımı

Eğitim ve psikoloji alanında özellikle karmaşık bilişsel süreçler gerektiren okuduğunu anlama becerilerinin ölçülmesinde, bir okuma parçasına bağlı birden fazla madde içeren testler yaygın olarak kullanılmaktadır (Rosenbaum, 1988). Ortak bir uyarıcı paylaşan ve birbirleriyle ilişkili olan bu maddeler, madde takımı (testlet) olarak adlandırılmıştır (Wainer & Kiely, 1987). Madde takımı, bir birim olarak geliştirilen ve sınava giren kişinin izleyebileceği önceden belirlenmiş sabit sayıda yol içeren, tek bir içerik alanıyla ilgili bir grup maddedir. Bu madde türünde, ortak bir uyarıcı olarak kullanılan resim, metin veya senaryo gibi içeriklerle birden fazla madde oluşturulur. Madde grubu/kümesi/demeti ifadesinin kullanılmasının nedeni tek bir içeriğe bağlı olarak bireylerin okuduğunu anlama becerilerinin ölçülmesinde tek maddenin kullanılmasının test süresi ve adayın verimliliği açısından uygun olmamasıdır. Bu durumlarda, sınava girenlerin herhangi bir test maddesini yanıtlamadan önce uyarıcıyı işlemek için önemli miktarda zamana ihtiyacı vardır.

Birden fazla madde içeren madde takımları, testin amacına göre adayların yanıtlarına bağlı olarak ardışık ve bir dizi puanlamayla sonuçlanan hiyerarşik bir dallanma şemasını izleyebileceği gibi Şekil 1'deki gibi tüm adaylara uygulanan birkaç maddeden oluşan yalnızca tek bir doğrusal yol da içerebilir (Wainer & Kiely, 1987)

Şekil 1

Doğrusal Madde Takımı



Özetle, bir madde takımı (Wainer ve Kiely, 1987) veya madde kümesi (Rosenbaum, 1988), bir birim olarak geliştirilen ve sınava giren kişinin izleyebileceği önceden belirlenmiş sabit sayıda yolu içeren, tek bir içerik alanıyla ilgili bir grup maddeyi ifade eder. Test geliştirme açısından bakıldığında, madde takımlarının kullanımı test verimliliğinin artırılmasına yardımcı olmaktadır. (Wainer ve diğerleri, 2007). Madde takımları sayesinde ortak bir içerik ile birden fazla soru sorularak zaman ve çaba açısından ekonomiklik sağlanır ve özellikle BBT uygulamalarında bağlam etkisi, sıra etkisi gibi sorunlara çözüm üretilebilir (Ho & Dodd, 2012; Wainer & Wang, 2000). Çünkü bireyler ortak uyarana sahip içeriği yalnızca bir kez değerlendirmek zorundadır ve daha sonra bu içerikteki bilgileri madde takımında yer alan tüm maddelerde kullanabilmektedir. Bu nedenle eğitimde matematik, İngilizce, Türkçe vb. birçok alanda madde takımlarından oluşan okuma metni bazlı çoktan seçmeli test yapısı kullanımı oldukça yaygındır.

Madde Takımlarında Yerel Bağımlılık

Eğitim alanında madde takımlarının asıl kullanım sebebi, test puanlarından elde edilen sonuçların geçerliğini arttırmaktır. Fakat ortak uyaran temelli bu maddeler, aynı madde takımında yer alması durumunda ölçülmesi hedeflenen örtük özelliğin etkisi dışında birbirleri ile ilişkili olabilir ve böylece geçerlik sorunu ortaya çıkabilir. Okuduğunu anlama becerisini ölçmek için okuma metinlerinin yer aldığı bir testte, öğrencinin performansını öğrencinin metnin konusuna olan ilgisi etkileyebilir (Yen, 1993). Bu durumda aynı madde takımındaki maddeler, temel MTK varsayımlarından biri olan yerel bağımsızlığı ihlâl etmektedir. Bunun nedeni, bir madde takımı içindeki maddelerin tamamen bağımsız olmaması yani ortak bir metin, grafik, şekil vb. bir uyaran aracılığıyla ilişkili olmalarıdır (Rosenbaum, 1988). Bu durumda madde takımı içerisindeki herhangi bir maddeye verilen

tepki, aynı madde takımı içerisindeki diğer maddelere verilen tepkiyi de etkileyebilmektedir (Wainer & Kiely, 1987).

Yerel madde bağımlılığı (local dependency), maddeler arasında özel bir bağımlılık olduğunda ortaya çıkan madde takımlarının ana özelliğidir (Wilson, 1988). Madde takımı kaynaklı yerel bağımlılığa madde takımı etkisi adı verilir (Wainer & Kiely, 1987). Varyans ne kadar fazlaysa, madde takımının içerdiği etki de o kadar fazladır (Wainer & Wang, 2000).

Kim ve diğerlerine (2007) göre, yerel bağımlılık hesaplanmasında dört farklı madde çifti türü ele alınır. Bunlar; (i) bağımsız maddeler arası (ii) bir madde takımı ve bir bağımsız madde arası (iii) bir madde takımı içindeki iki madde arası (iv) farklı madde takımlarındaki iki madde arası yerel bağımlılık dereceleridir. Aynı madde takımındaki iki madde çifti arasında tespit edilen yerel bağımlılığının kaynağı, genellikle yapıyla alakasız varyansların kaynağını oluşturan bir nüans faktörü olarak kabul edilen madde takımı faktörüdür. Diğer türlerde ise genellikle dil testlerinin ana yapısıyla ilgili yardımcı bir faktör olarak kabul edilen alt beceri faktörü olma eğilimindedir.

Madde takımı verilerinde yerel bağımlılığı ele almak için yaygın bir yaklaşım, madde takımını ölçüm birimi olarak tanımlamak ve daha sonra çok kategorili MTK modellerinden birini uygulamaktır (Wainer & Lewis, 1990). Bu nedenle madde takımlarını içeren bu sorunların üstesinden gelmek için yerel bağımlılık derecelerini dikkate alan MTTK gibi daha karmaşık modeller gereklidir.

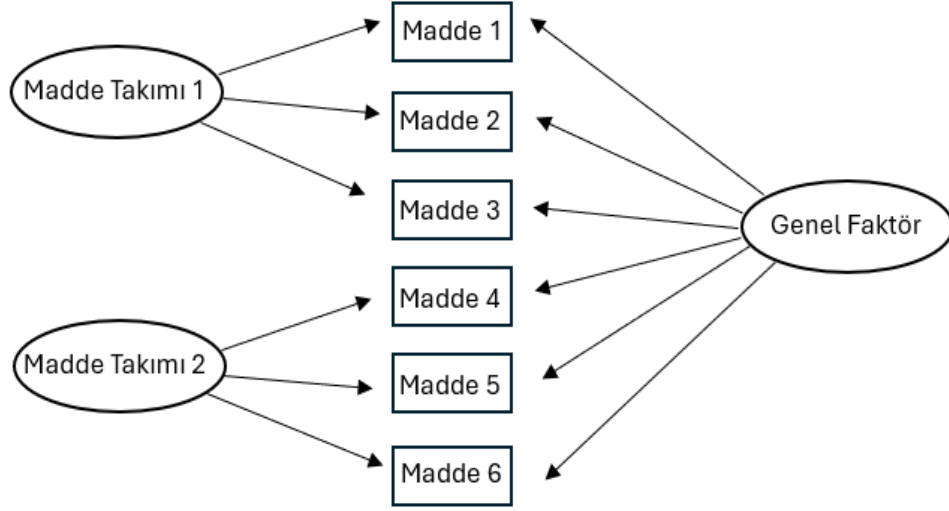
Madde Takımı Tepki Kuramı ve Modelleri

Ortak bir uyarana dayalı madde takımlarından oluşan testler, parametre tahminlerinde hatalara ve güvenilirliğin fazla tahmin edilmesine yol açabilir (Li ve diğerleri, 2006). Bunun nedeni, yerel bağımlılık derecesini belirten madde takımı etkilerinin göz ardı edilmesidir (DeMars, 2010). Yerel bağımlılığın varlığı, uygun şekilde ele alınmazsa, teste bağlı psikometrik sonuçlar olumsuz etkilenebilir. Test bilgisinin ve güvenilirliğin yüksek kestirilmesine (Braeken, 2011; Chen & Wang, 2007; Sireci ve diğerleri, 1991; Thissen ve

diğerleri, 1989; Yen, 1993), madde parametre ve yetenek kestirimlerinde yanlışlığa (Tuerlinckx & De Boeck, 2001; Wainer & Wang, 2000); geçti kaldı gibi kararlarda sınıflama hatalarına (Keller ve diğerleri, 2003) sebep olabilir. Bu sorunun üstesinden gelmek amacıyla literatürde birkaç çözüm önerisi yer almaktadır. Bunlardan biri ortak uyarana bağlı maddelerin her birini bağımsız birer madde olarak ele almaktır. Fakat bu yaklaşım ekonomiklik ilkesine aykırı olup istatistiksel açıdan da uygun değildir. Çünkü aynı okuma parçasına birden çok maddede yer verilmesi ek bir okuma yükünü beraberinde getirip zaman ve çaba açısından ekonomiklik ilkesi ile çelişir. Ayrıca bir okuma parçasının birden çok maddede verilmesi o uyarana ait aşinalığı da beraberinde getirecek bu da geçerlik sorununa neden olacaktır. Diğer bir yaklaşım ise madde takımlarında yer alan maddeleri birleştirmek ve her bir madde takımını çok kategorili tek bir madde olarak ele almaktır (Zhang, 2010). Madde takımlarının tek bir ölçü birimi olarak kullanılmasıyla “süper maddeler” oluşturulur (Thissen & Wainer, 2001) ve bu durumda, her süper madde, çok kategorili puanlanabilen tek bir madde olarak ele alınabilir. Bu durumda Aşamalı Tepki Modeli ([GRM], Samejima, 1969) veya Kısmi Kredi Modeli gibi çok kategorili MTK modelleri uygulanabilir ([PCM], Masters, 1982). Fakat madde takımlarına çok kategorili MTK modelleri uygulayarak çok kategorili puanlama kullanmak nihâi çözüm değildir. Toplam puan yaklaşımının kullanılması kalibrasyon sürecini kolaylaştırırsa da maddelerin madde takımlarında birleştirilmesi bilgi kaybına neden olabilmektedir. Çok kategorili MTK modellerinin madde takımlarındaki maddeler arasındaki bağımlılığı ele almak için çok etkili bir çözüm olmadığı anlaşıldığından; fazladan bir madde takımı etki parametresini içeren bir MTK modeli önerilmiştir. (Bradlow ve diğerleri 1999; Wainer ve diğerleri, 2000; Wang ve diğerleri, 2002; Wainer ve diğerleri, 2007; Wainer & Wang, 2000). Örnek teşkil etmesi amacıyla her biri beş madde içeren iki madde takımından oluşan MTTK modelinin yapısı Şekil 2’de gösterilmiştir.

Şekil 2

Madde Takımı Modelinin Yapısı



Şekil 2’de görüldüğü üzere örnek madde takımı 1 ve 2’de üçer madde yer almakta ve tüm maddeler hem özel faktöre hem de tek bir genel faktöre yük vermektedir. Ayrıca özel faktörlerdeki maddelerin faktör yüklerinin ölçeği genel boyut yükü ile sınırlandırılır. Madde-birey etkileşimine modelde yer veren MTTK modelleri maddeler arası varyansın da hesaba katılmasını sağlar. Modelde yer alan madde takımı etki parametresi olarak adlandırılan birey-madde etkileşimi parametresi madde takımı içinde yer alan tüm maddelerde ortak iken, madde takımları arasında farklıdır (Dresher, 2002). Bu etki parametresinin varyansındaki büyüklük, madde takımında yer alan maddelerin yerel bağımlılık derecesinin belirlenmesinde rol oynar (Wainer & Wang, 2000). Yerel madde bağımlılığını hesaba katan bu MTTK modellerinin en basiti Rasch MTTK modelidir (Wang ve diğerleri, 2002). Bu modelde de MTK modellerindeki gibi a ve b parametreleri olup $c_i=0$ ve $a_i=1$ olarak kabul edilir. Fakat burada MTK modellerinden farklı olarak bireye özgü madde takımı parametresi ($\gamma_{jd(i)}$) yer almaktadır. Rasch MTTK modeli ve madde bilgi fonksiyonu denklem (8) ve (9) ile sırasıyla verilmiştir.

$$P(x_{ji} = 1 | \theta_i, b_i) = \frac{\exp(\theta_j - b_i - \gamma_{jd(i)})}{1 + \exp(\theta_j - b_i - \gamma_{jd(i)})} \quad (8)$$

$$I(x_{ji} = 1 | \theta_i, b_i) = \left(\frac{\exp(t_{ji})}{1 + \exp(t_{ji})} \right)^2 \frac{1}{\exp(t_{ji})} \quad (9)$$

$$t_{ji} = (\theta_j - b_i - \gamma_{jd(i)})$$

b_i : i maddesinin madde güçlük parametresi

$\gamma_{jd(i)}$: j bireyinin d madde takımına yuvalanmış i maddesinin madde takımı etki parametresi

Yerel bağımsızlık varsayımı sağlandığında, madde takımı etki varyansı sıfır yani tüm bireyler için $\gamma_{jd(i)} = 0$ olur ve bu durumda MTTK modeli tek boyutlu MTK modeline dönüşür. Bireye özgü madde takımı etki parametresi, gizil özellik ve madde parametrelerinden bağımsızdır. Madde takımı etki parametresinin ($\gamma_{jd(i)}$), bireye ve madde takımına özgü bir parametre olduğuna dikkat edilmelidir. Yani bu, belli bir madde takımı içindeki maddelerin yerel bağımlılıklarının sınava giren her aday için değiştiği anlamına gelir. Bu nedenle, $\gamma_{jd(i)}$ 'nin varyansı tipik olarak her madde takımı için tahmin edilir ve her madde takımı içindeki maddelerin yerel bağımlılık derecesinin bir göstergesi olarak kullanılır. Ayrıca madde takımı etkisinin varyansları madde takımları boyunca değişmektedir.

En basit MTTK modellerinden olan Rasch MTTK modeline ayırt edicilik parametresi eklenmesiyle 2PL MTTK modeli elde edilir. Aslında bu model bilinen iki parametrelili lojistik modele madde takımı etki parametresinin eklenmesiyle oluşan bir modeldir (Bradlow ve diğerleri, 1999).

a_i , i maddesinin madde ayırt edicilik parametresi olmak üzere 2PL MTTK denklem 10'da verilmiştir.

$$P(x_{ji} = 1 | \theta_i, a_i, b_i) = \frac{\exp(a_i(\theta_j - b_i - \gamma_{jd(i)}))}{1 + \exp(a_i(\theta_j - b_i - \gamma_{jd(i)}))} \quad (10)$$

$$I(x_{ji} = 1|\theta_i, \alpha_i, b_i) = \alpha_i^2 \left(\frac{\exp(t_{ji})}{1 + \exp(t_{ji})} \right)^2 \frac{1}{\exp(t_{ji})}$$

$$t_{ji} = \alpha_i(\theta_j - b_i - \gamma_{jd(i)})$$

Wainer ve diğerleri (2000) c_i şans parametresini 2PL-MTTK modeline ekleyerek 3PL-MTTK modelini oluşturmuştur.

c_i :i maddesinin madde şans parametresi olmak üzere;

$$P(x_{ji} = 1|\theta_i, \alpha_i, b_i, c_i) = c_i + (1 - c_i) \frac{\exp(\alpha_i(\theta_j - b_i - \gamma_{jd(i)}))}{1 + \exp(\alpha_i(\theta_j - b_i - \gamma_{jd(i)}))} \quad (11)$$

$$I(x_{ji} = 1|\theta_i, \alpha_i, b_i, c_i) = \alpha_i^2 \left(\frac{\exp(t_{ji})}{1 + \exp(t_{ji})} \right)^2 \frac{1 - c_i}{c_i + \exp(t_{ji})}$$

$$t_{ji} = \alpha_i(\theta_j - b_i - \gamma_{jd(i)})$$

Şans parametresinin (c_i) sıfır olması durumunda 3PL-MTTK, 2PL-MTTK'nin özel bir hâli olur. 3PL MTTK modeli diğer MTTK modellerine göre daha fazla parametre içerdiğinden hesaplama algoritmaları daha karmaşıktır. 2PL-MTTK modelindeki gibi madde takımı etki varyansı madde takımlarında sabit olmayıp madde takımları arasında değişmektedir (Wainer ve diğerleri, 2000). Bu nedenle, 3PL-MTK'de artık test takımı etkisinin varyansını tanımlayan tek bir parametre varsayılmamaktadır. Ancak $c_i = 0$ ve her madde takımı için $\sigma_{\gamma d(i)}^2 = \sigma_{\gamma}^2$ olduğunda 2PL-MTTK modeli 3PL-MTTK modelinin basitleştirilmiş bir versiyonudur.

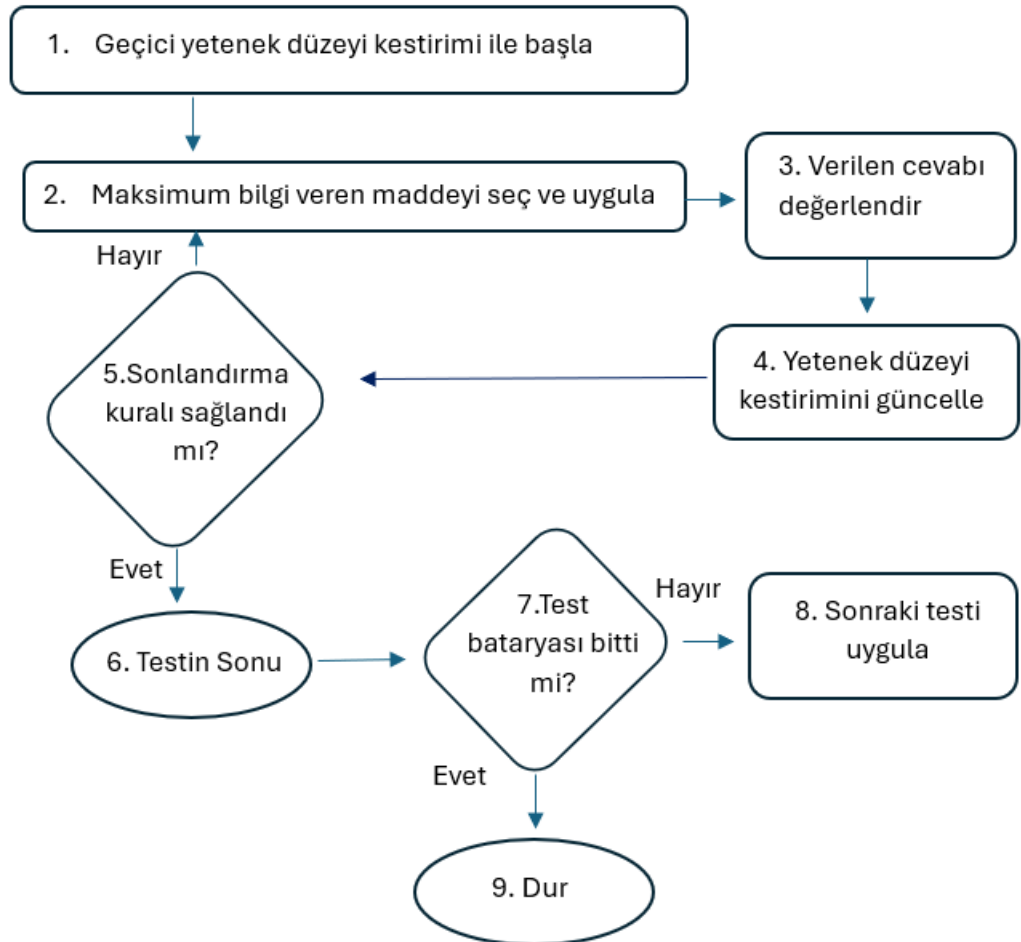
Bilgisayarda Bireyselleştirilmiş Test

Geleneksel kâğıt kalem testlerine alternatif olarak bilgisayar ortamında gerçekleştirilen ve son yıllarda oldukça popüler olan BBT, bireylerin kendi yetenek düzeyine uygun maddeleri cevaplandığı ölçme sistemidir. Belli bir algoritmaya dayanan BBT ile bireyin yetenek düzeyi ile bireye yöneltilen maddelerin güçlük düzeyleri eşleştirilerek daha güvenilir ölçme sonuçları elde edilmesi sağlanır. Bu sayede bireyler kendi yetenek düzeylerine uygun olan maddeler ile karşılaşılır. Bireyin her bir maddeye verdiği cevaptan

sonra geçici olarak kestirilen yetenek düzeyine göre yeni maddeler bireye yönlendirildiği için geleneksel kâğıt kalem testlerinden farklı olarak test formları anında oluşturulur. Sonuç olarak, BBT uygulamalarının en büyük avantajı, sınava giren bireyin yeteneğinin ölçüm hassasiyetini artırırken madde sayısını kısaltabilmesidir (Wainer, 2000; Weiss, 1982). Buna ek olarak BBT sistemi, tüm bilgisayar tabanlı sınavlar gibi bilgisayar teknolojisinin faydalarından yararlanır ve sınava girenler ve test yöneticileri için planlamada esneklik, yenilikçi madde formatlarının kullanım imkânı, test eşitlemeye ihtiyaç duyulmadan puanların karşılaştırılabilirliği, yüksek test güvenliği, otomatik veri toplama ve anında puan raporlama gibi çeşitli avantajlar sağlar. (Bergstorm & Lunz, 1999). BBT uygulamasının temel işleyiş diyagramı detaylı bir şekilde Şekil 3'te verilmiştir (Thissen ve diğerleri, 2000).

Şekil 3

BBT Uygulamalarının Temel İşleyiş Diyagramı



BBT uygulaması yapılırken öncelikle madde havuzuna uygun MTK modeli ve uygun bir yetenek kestirim yöntemi belirlenir, sonra uygulamaya başlanacak ilk madde seçilir, daha sonra bireyin cevabına göre diğer maddeler seçilir ve uygulanır. Son olarak da belli bir koşula göre test sonlandırılır. Tüm bu işlemlerin uygulanması belli bir algoritmaya dayalı gerçekleşir. Hangi güçlükte bir maddenin uygulanacağı ise bireylerin uygulama anında göstermiş oldukları performansa bağlıdır. Böylece bireyler yeteneğinin çok üstünde ya da çok altında sorularla karşılaşmaz ve yetenek düzeylerine uygun maddeleri alır. Sonuç olarak BBT uygulamaları ile bireyin cevaplarından yola çıkarak bireyin yeteneğine en yakın olduğu noktaya ulaşılması hedeflenir. Böylece test daha az madde ile daha kısa sürede sonlanır. Geleneksel kâğıt-kalem uygulamalarında birey, kendi yetenek düzeyine göre daha zor sorularla karşılaştığından bu durum onun testi alma motivasyonu olumsuz yönde etkileyebilir. Fakat BBT uygulamaları ile bireyler kendi yetenek düzeyi hakkında daha fazla bilgi sağlayan daha az madde ile daha güvenilir sonuçlar elde edilebilir (Kalender, 2009; Thompson & Weiss, 2011).

Ulusal ve uluslararası bağlamda, BBT uygulamaları çok sayıda test kuruluşu ve kurumu tarafından tercih edilmektedir. Bu uygulamalar başarı, yetenek, tutum, kişilik ve ilgi gibi değişkenlerin ölçülmesine uygun olarak tasarlanmaktadır. Örneğin, Uluslararası Öğrenci Değerlendirme Programı (Programme for International Student Assessment [PISA]), lisansüstü eğitim almak isteyen bireylerin girdiği uluslararası kabulü olan sınavlardan olan Graduate Record Examinations (GRE) ve Graduate Management Admission Test (GMAT), Uluslararası Yetişkin Yeterliliklerini Değerlendirme Programında (PIAAC), İngilizce yeterliğin ölçüldüğü Test of English as a Foreign Language (TOEFL), yurt dışına lisans eğitimi almak isteyenlerin girdiği sınavlardan Scholastic Aptitude Test (SAT) BBT veya BBT desenlerinden ÇAT olarak uygulanmaktadır (Educational Test Service, 2023; Luecht & Sireci, 2011). Sadece eğitim alanında değil ABD’de depresyon ve diğer zihinsel sağlık sorunlarını ölçmek için sağlık hizmetlerinde de BBT uygulamalarının doğrudan kullanıldığı bilinmektedir (Fliege ve diğerleri, 2005). Bu unsurlar göz önünde

bulundurularak, bilgisayar tabanlı test ve BBT uygulamalarının gelecekte özellikle eğitim sistemlerinde nasıl bir rol oynayabileceği konusunda farklı ülkelerde çeşitli çalışmalar ve senaryolar geliştirildiği söylenebilir.

BBT sistemi altı ana bileşenden oluşur: madde tepki kuramı modeli, madde havuzu, başlama kuralı, madde seçme kuralı, yetenek tahmin yöntemi ve sonlandırma kuralı (Thompson & Weiss, 2011; Weiss & Kingsbury, 1984). Bu altı bileşen aşağıda detaylı bir şekilde açıklanmaktadır. BBT uygulamalarında ayrıca temel bileşenlerden hariç iki bileşen daha yer alabilir.

Madde Tepki Kuramı Modeli

BBT uygulamalarının temeli MTK modeline dayanmaktadır. MTK'nin bir maddeye verilen cevaplar ve madde özellikleri ile bireylerin yetenek düzeyleri arasında matematiksel bir fonksiyon kurması, BBT uygulamalarında üst düzey istatistiksel analizler yapılmasına imkân tanır (van der Linden & Glas, 2010). Bireyin yeteneğinin hesaplanması ve bireyin kestirilen geçici yeteneğine uygun madde havuzundan madde seçilmesi ve madde havuzunun kalibrasyon işlemleri MTK modelleri ile yapılmaktadır (Magis ve diğerleri, 2017).

MTK'nin testten çok maddeye odaklanması birey madde etkileşiminin sağlanması açısından BBT uygulamalarında oldukça avantaj sağlamaktadır. MTK'nin parametre değişmezliği özelliği sayesinde bireylerin farklı maddelerden oluşan test formu almaları testten alınan puanlar üzerinde etkili olmayacaktır.

Veri yapısına uygun MTK modelinin seçilmesi oldukça önemlidir. Bu modeller ikili puanlanan maddeler için Rasch Modeli, 1PLM, 2PLM, 3PLM, 4PLM; çoklu puanlanan maddeler için ise ATM, KPM, GKPM, DÖM ve STM'dir (Hambleton & Swaminathon, 1985, Linden & Hambleton, 1997; Reckase, 2009). Veri yapısı çok boyutlu ise çok boyutlu Rasch, 1PL, 2PL, 3PL, 4PL modelleri (Embretson & Reise, 2000, Reckase, 2009); madde takımlarından oluşuyorsa Rasch, 1PL, 2PL, 3PL MTTK modelleri (Wainer ve diğerleri, 2000) kullanılabilir.

Madde Havuzu

BBT'nin en önemli bileşenlerinden olan madde havuzu sınava giren kişilere uygulanabilecek geniş bir madde grubunu içermektedir. Aslında geleneksel kâğıt kalem testleri de bir madde havuzundan oluşur. Ancak kâğıt kalem testlerindeki madde havuzu, buradan seçilen herhangi bir maddenin o testin verilen formu için sabit kalması açısından BBT'lerden farklıdır. Çünkü BBT uygulamalarında, sınava giren her kişi havuzdan alınan çeşitli madde gruplarından oluşan bireyselleştirilmiş bir teste tabi tutulur. Bu nedenle madde havuzunun büyüklüğü BBT uygulamalarında oldukça önemlidir. Stocking (1994), madde havuzunda madde takımı sayısının en az altı katı kadar madde bulunması gerektiğini belirtmiştir. Reckase (2010) ise, standart normal dağılan bir örneklem için yaklaşık 200 maddelik bir havuzun uygun olduğunu belirtmiştir. Fakat yine de alan yazında madde havuzunda kaç madde olması gerektiği ile ilgili sorulara, üzerinde uzlaşılan net bir cevap vermek oldukça zordur. Çünkü bir BBT madde havuzunun büyüklüğü; testin amacı, havuzda yer alacak madde türleri, uygulanacak MTK modeli ve en önemlisi maddelerin psikometrik özellikleri gibi birçok değişkene bağlı olarak değişebilmektedir (Weiss & Şahin, 2024).

Sadece madde havuzunun büyüklüğü değil madde havuzunda yer alan maddelerin kalitesi de BBT uygulamalarındaki uyarlanabilir algoritmanın performansı üzerinde önemli bir etkiye sahiptir (Flaughner, 2000; Stocking, 1994). Havuzda yer alan maddelerin güçlük parametresi, sağladıkları bilgi düzeyi, kapsamı temsil etme düzeyleri ve psikometrik özellikleri BBT uygulamaları ile kestirilen yeteneklerin bireyin gerçek yeteneğine yakın olmasını sağlayan faktörlerdir (Parshall ve diğerleri, 2002).

BBT madde havuzunun geliştirilmesine yönelik çalışmalar kâğıt kalem testindeki madde havuzu geliştirilmesine oldukça benzerdir. Çünkü her iki havuz türünde de maddelerin içerik özelliklerine göre yazılması, içerik kalitesi ve test duyarlılığı açısından incelenmesi ve psikometrik özelliklerinin değerlendirilebilmesi için ön teste tabi tutulması gerekmektedir (Flaughner, 2000). Bir maddenin kalitesini değerlendirmek için kullanılan

yöntemler, geleneksel madde istatistiklerini, madde parametreleri ve madde bilgileri gibi MTK veya MTTK temelli kriterleri içerebilmektedir (Wainer, 1989). İncelenmesi gereken bir diğer önemli nokta, her bir maddenin MTK veya MTTK gibi temel ölçüm modeline uyup uymadığıdır (Kingston & Dorans, 1985). Maddeler, sınava giren her bireyin tahmini yeteneğine uyacak şekilde uyarlanabilir şekilde seçildiğinden, madde havuzunun birkaç farklı yetenek düzeyi için yüksek kaliteli maddeler içermesi gerekir. Bu nedenle, bir BBT madde havuzu, kâğıt kalem testlerindeki madde havuzundan daha büyük olma eğilimindedir, böylece sınava giren kişinin çok çeşitli yetenekleri için farklı test maddesi kombinasyonları oluşturulabilir (Davey & Nering, 2002). Bu kadar çok sayıda maddeye sahip bir testi, çoğu zaman tek bir örneklem grubuna aynı zamanda uygulamak güç hatta imkânsız olabilir. Bu durumda farklı ölçekleme ve bağlama yöntemleri kullanılarak geniş madde havuzları literatüre uygun bir şekilde oluşturulabilir (Hambleton ve diğerleri, 1991; Vale, 1986). Özet olarak optimal düzeyde bir madde havuzunda, istenilen içerik kategorilerinde yeterli sayıda madde ve istenen madde parametreleri ve madde analiz istatistikleri dâhil olmak üzere yüksek kalitede, farklı yetenek düzeylerine uygun yeterli sayıda madde bulunmalıdır (Flaughner, 2000). Ancak bu şartlar sağlandığında, BBT uygulamalarının avantajlarından en etkili şekilde yararlanılabilir.

Başlama Kuralı

BBT uygulamalarında bireyin cevaplandıracağı maddeler, her bir maddeye vermiş olduğu cevaplara göre anlık/geçici yetenek kestirimi yapılarak belirlenmektedir. BBT uygulamalarına bireyler teste her zaman aynı madde ile başlamaz. Teste başlanılacak ilk maddenin anlık yetenek kestirimi olmadığından teste nasıl başlanacağını belirlenmesi gerekmektedir (de Ayala, 2009). Bireyin yeteneği hakkında ön bilgi mevcut ise bu bilgi başlama kuralı olarak kullanılabilir (Parshall ve diğerleri, 2002, Thompson & Weiss, 2011). Bu ön bilgi bireye ait diploma puanı, hazır bulunuşluk sınav puanı ya da herhangi bir akademik sınav ya da puan ortalaması olabilir. Genellikle birey hakkında başlangıçta böyle bir ön bilgi olmadığından BBT uygulamalarında başlama kuralı değeri θ ölçeğinde ifade

edilir. İlk madde olarak tüm bireylere aynı maddelerin uygulanması yöntemi, ilk maddenin kullanım sıklığı açısından sorun olabileceğinden pek önerilmemektedir. Alan yazında da en çok kullanılan yöntem ise yetenek dağılımının normal olduğu durumda bireylerin başlangıç yetenek düzeyini sıfıra eşitleyerek madde havuzundan ilk olarak ortalama güçlükteki bir maddeyi seçmektir (Keng, 2009; Murphy ve diğerleri, 2010; Yao, 2019). Diğer yöntemler ise maddelerin aşırı kullanımını engellemek amacıyla bireylerin başlangıç yetenek düzeyini $(-0.5, 0.5)$, $(-1, 0)$ gibi belli bir aralıktan rastgele seçilen bir değere eşitlenmesi yöntemlerini içerir (Hambleton & Xing, 2006; Thompson & Weiss, 2011). Başlama kuralı olarak tek bir madde seçmek yerine tek bir madde takımının seçilmesi de benzer mantığa dayanır. Sadece bir madde takımı birden fazla madde içerdiğinden burada uygulanması gereken algoritma ya da model farklı olabilir. Belirlenen ilk maddeye göre yapılan yetenek kestirimi BBT uygulamalarının verimliliği ve ölçme kesinliği açısından oldukça önemlidir (van der Linden & Pashley, 2010). Örneğin testin başlangıcında yapılan yanlış yetenek kestirimi testin gereksiz uzamasına sebep olabilir ve bu da BBT uygulamasının verimliliğini düşürebilir. Bu nedenle madde havuzunun genişliği, testin amacı, testin uzunluğu, seçilen MTK modeli başlama kuralının belirlenmesinde etkili olduğundan tüm değişkenlerin birlikte ele alınıp dikkatli bir şekilde incelenmesi önemlidir.

Madde Seçim Kuralı

Geleneksel kâğıt kalem testlerinde maddeler; testin uygulanmasından önce testi oluşturan kişiler tarafından seçilirken, BBT uygulamalarında bireylerin cevaplarına göre test sırasında en uygun maddeler seçilir. Böylece test maddelerini bireye özgü uyarlanabilir bir test algoritmasına dayanan ölçme sistemine ihtiyaç duyulmuştur. Madde seçim kuralı ile sınava giren bireyin geçici/anlık θ kestiriminde hatayı minimize eden ve maksimum bilgi sağlayan bir madde seçilir. Başlangıçta bu sistem karmaşık ölçme modeline sahip olmayan temel ve esnek düzeyde bir sistem iken (Lord, 1971) bilgisayar teknolojinin de ilerlemesiyle karmaşık test algoritmaları geliştirilerek MTK, MTTK gibi kuramlara temel alan sistemler oluşturulmuştur. Yaygın olarak kullanılan iki madde seçme algoritması şunlardır: Maksimum

bilgi yöntemi ve Bayes yöntemleri (Kingsbury & Zara, 1989; Thissen & Mislevy, 2000). Madde bilgi yöntemi, Fisher bilgisini ya da Kullback-Leibler bilgisini kullanmakta iken Bayes yaklaşımı yöntemi; log-olabilirlik fonksiyonuna yetenek seviyesine ait önsel ve sonsal dağılımları dâhil eder (Magis ve diğerleri, 2018).

Çoğu madde seçim yönteminin amacı, en çok bilgi veren maddeleri verimli şekilde toplamaktır (Parshall ve diğerleri, 2002). Madde bilgisi arttıkça test bilgi miktarı da artar ve buna bağlı olarak yetenek kestiriminin ölçüm kesinliği de artış gösterir. En çok kullanılan madde seçim yöntemi ise maksimum Fisher bilgi yöntemidir (van der Linden & Pashley, 2010). Bu yöntemde, bireyin kestirilen anlık yetenek düzeyinde maksimum bilgi veren madde seçilmektedir. BBT uygulamalarında madde seçiminin temelinde bilgi fonksiyonu kullanılarak gerçekleştirilmektedir (Dodd ve diğerleri, 1995). Örneğin Maksimum Fisher Bilgi (MFB) yöntemi kullanıldığında, belli bir yetenek düzeyine sahip birey için $I_i(\theta)$ fonksiyonunun maksimum değeri aldığı madde, bireye uygulanacak sıradaki madde olur. Bu yöntemin amacı, bireyin kestirilen geçici yeteneğine en uygun güçlük değerine sahip bir madde seçmektir.

Murphy ve diğerleri (2010) MFB, maksimum sonsal ağırlıklandırılmış bilgi ve minimum beklenen sonsal varyans yöntemlerini tek boyutlu BBT ve madde takımı tabanlı BBT uygulamalarında karşılaştırmış ve çok benzer sonuçlar elde etmiştir. MFB dışındaki yöntemler çok boyutlu modeller için çok daha uzun hesaplama süresi gerektirmiştir (Murphy, 2010; Segall, 1996; Veldkamp & van der Linden, 2002; Yao, 2012). BBT uygulamalarında madde seçim kuralları içinde MFB, en kolay ve pratik olandır. Fakat ayırt ediciliği yüksek olan maddeleri seçme eğiliminde olduğu için madde kullanım sıklığı oranı ile ilgili sorunlara sebep olabilmektedir (Chang & Ying, 1996). Bu nedenle madde havuzu 1PLM veya Rasch model gibi MTK modelleri ile kalibre edilen BBT uygulamalarında kullanılması önerilir.

Yetenek Kestirimi Yöntemi

BBT uygulamalarında, madde havuzundaki madde veya madde takımlarının parametre değerleri önceden kestirilip kalibre edildiği için uygulama sırasında kestirilmesi gereken bireylerin yetenek düzeyidir (θ). Yetenek kestirim süreci ise cevaplanan ilk maddeye dayalı ilk yetenek kestiriminin yapılması ile başlar. Başlangıçtaki yetenek tahminini belirlemenin bir yolu, bireyler hakkında ön bilgileri kullanmak veya θ 'nın normal dağılıma sahip olduğu varsayımı altında basitçe dağılımın ortalaması olan sıfıra eşitlemek/ayarlamaktır. Her bir madde veya madde takımı verildikten sonra, BBT algoritması tarafından bir sonraki madde veya madde takımının seçilmesi için yetenek kestirimlerinin bilinmesi gerekir. Nihâi yetenek kestirimi ise testin sonunda, sınava giren kişinin tüm yanıtlarına dayalı olarak gerçekleştirilir. Geçici ve nihai yetenek kestirimlerinin aynı yöntem kullanılarak elde edilmesi zorunlu değildir (Chang ve diğerleri, 2000). Nihai yetenek tahmini aynı zamanda farklı bir yetenek ölçüsüne de dönüştürülebilir (Parshall ve diğerleri, 2002). BBT uygulamalarında yetenek tahminine yönelik iki yaygın yaklaşım, maksimum olabilirlik ve Bayes yaklaşımına dayalı kestirim yöntemleridir. Maksimum olabilirlik kestirimi (MLE) yansız ve etkili bir kestirim yöntemi olmasına rağmen cevap kategorilerinde varyans olmadığı kestirim yapamaz (Hambleton ve diğerleri, 1991). Bunun yerine alan yazında BBT uygulamalarında ağırlıklandırılmış maksimum olabilirlik kestirim yönteminin (WMLE) kullanılması önerilir (Wang & Wang, 2001).

Maksimum sonsal (MAP) ve beklenen sonsal (EAP) Bayes yaklaşımına dayalı yöntemleridir ve gizil özellik üzerinde bir ön dağılımın tanımlanmasını gerektirir. İkisi de önsel bilgi kullanıldığı için standart hataları benzerdir fakat EAP yetenek sonsal dağılımın ortalamasını kullanırken MAP dağılımın modunu kullanır. EAP iteratif olmadığından hızlı hesaplanır, MAP ise iteratiftir ve hesaplanması zaman alır. MAP madde sayısı fazla olduğunda (20'den çok) yanlışlık göstermezken EAP madde sayısı çok olduğunda bile ortalamaya doğru yanlışlık gösterebilir. Bayes kestirim yöntemleri tüm yanıt örüntüleri için

kestirim yapabildiğinden EAP ve MAP yöntemlerinin kullanılması önerilmektedir (Hambleton & Swaminathon, 1985; Reckase, 2010).

Sonlandırma Kuralı

BBT uygulamalarında madde yönetiminin ne zaman sona ermesi gerektiğini belirleyen bir sonlandırma/durdurma kuralına ihtiyacı vardır ve bu kural BBT uygulamalarının etkililiği açısından oldukça önemlidir. Çünkü test olması gerekenden daha az sürede durdurulursa bireyin yeteneği yanlış kestirilir. Eğer test olması gerekenden daha uzun sürerse de BBT'nin ekonomiklik ilkesi ihlâl edilmiş olur. Aynı zamanda bireyde performans düşüklüğü meydana gelebilir ve bu durum da yanlış kestirimlere neden olabilir (Linacre, 2000).

BBT uygulamalarında sabit uzunluk ve değişken uzunluk olmak üzere iki temel sonlandırma/durdurma kuralı mevcuttur. Değişken uzunluktaki BBT uygulamalarında her sınava giren birey farklı sayıda maddeden oluşan test formları alır. Standart hata, θ ve minimum bilgi sonlandırma kuralları ise başlıca değişken uzunluklu sonlandırma kurallarındandır. Standart hata sonlandırma kuralına göre yetenek kestiriminin standart hatası belirlenen sabit değere geldiğinde test sonlandırılır (Thompson & Weiss, 2011). Değişken uzunluktaki BBT uygulamalarında, sınava girenler için uygulanan minimum ve maksimum sayıda madde olacak şekilde bazı kısıtlamalar eklemek de mümkündür. Sabit uzunluk durdurma kuralında ise bireylere önceden belirlenen sayıda madde uygulanır ve sonra test sonlandırılır. Bu durumda tüm bireyler aynı sayıda maddeden oluşan test formları alır.

Sabit uzunluktaki BBT'lerin uygulama sürecinin daha kolay olması ve madde havuzu kullanım oranlarının daha iyi tahmin edilebilmesi gibi avantajlara sahiptir (Thissen & Mislevy, 2000). Ayrıca sabit uzunluktaki testler adalet algısı yaratır ve sınava girenlere açıklanması daha kolaydır. Eğer sınava giren bireyler, farklı sayıda madde alırlarsa, bu durumda bazı adaylar yeteneklerini kanıtlamak için diğerleriyle aynı fırsata sahip

olmadıklarını iddia edebilir (Bergstrom & Lunz, 1999). Fakat yine de sabit uzunluktaki BBT uygulamaları uluslararası alanda da kullanılmaktadır (Örn; GRE, ASCB).

Sabit ve değişken uzunluk kuralı dışında bazı durumlarla karşılaşıldığı zaman da test durdurulur. Örneğin madde havuzundaki tüm maddeler kullanıldığında, ölçme kesinliği yeterli düzeye ulaştığında veya bireyde normal dışı davranışlarla (aynı cevap örüntüsünü işaretleme, kopya vb.) karşılaşıldığında da test sonlandırılabilir.

Yukarıda açıklanan BBT uygulamalarının altı bileşeni, uyarlanabilir bir testin ölçüm kesinliği ve verimliliği ile doğrudan ilgilidir. Fakat bir BBT uygulamasının etkili olması için genellikle kapsam dengeleme ve madde kullanım sıklığı olmak üzere iki önemli konunun da ele alınmasında fayda vardır.

Kapsam Dengeleme. Genel olarak eğitim alanında, içerikle ilgili geçerlik kanıtı için bir test geliştirilirken, testin ilgili içerik alanından maddelerin iyi bir temsili olmasına dikkat edilir. Kâğıt-kalem testlerinde, test formları önceden belli olduğundan ve her bireye aynı form uygulanacağından bu durumu karşılamak oldukça kolaydır. Ancak BBT uygulamalarında, önceden tanımlanmış test formları bulunmadığından bu tür özellikler, madde seçimi sırasında kapsam dengeleme stratejileriyle kontrol edilir. En yaygın kullanılan Kingsbury ve Zara (1989) tarafından geliştirilmiş kısıtlanmış BBT yönteminde, belirtke tablosundan yararlanılarak her bir konu alanına dayalı hedef ve gerçek yüzde değerlerine göre madde seçilir (Shin ve diğerleri, 2009). Bu yöntemin uygulanması da oldukça kolaydır. Madde takımlarının içeriği önceden oluşturulduğundan kapsam dengeleme, madde takımı oluşturma aşamasında dikkate alınarak da kontrol sağlanabilir.

Madde Kullanım Sıklığı Kontrol Yöntemi. Bazı maddelerin daha sık kullanılmasını önlemek ve dolayısıyla madde havuzunu daha etkili kullanmak amacıyla madde kullanım sıklığı yöntemleri kullanılmaktadır. Ayırt ediciliği yüksek olan maddeler; Maksimum Fisher Bilgi yöntemi ile daha sık seçilebilir ya da benzer yeteneklere sahip adaylar testin tamamı boyunca kendilerine verilen madde gruplarında önemli ölçüde örtüşmelerle karşılaşabilirler. Ayrıca madde havuzunun dengesiz bir şekilde kullanımı da BBT uygulamalarında

istenmeyen bir durumdur. Çünkü maddelerin önemli bir kısmının nadiren veya hiç uygulanmadığı büyük bir madde havuzu oluşturmak anlamlı değildir (Revuelta & Ponsoda, 1998). Ayrıca BBT uygulamalarının en önemli avantajlarından biri sınava girenlerin teste çeşitli yerlerde ve farklı zamanlarda girmelerine olanak tanıyan sınava erişim esnekliğidir (Wainer ve diğerleri, 2000). Bu durumları önleyerek madde ve test güvenliğini artırmak amacıyla madde kullanım sıklığı yöntemleri kullanılır. Bu tür test güvenliği endişeleri, yalnızca büyük ve sık sık yenilenen bir madde havuzuna sahip olmanın önemini değil (Way, 1998), aynı zamanda havuzdaki maddenin açığa çıkmasını kontrol edecek test algoritmasına olan ihtiyacı da vurgulamaktadır. Bu nedenle, bir maruz kalma kontrol prosedürünün iki ana hedefi olmalıdır: sıklıkla uygulanan maddelerin aşırı kullanımını önlemek ve nadiren veya hiç seçilmeyen maddelerin kullanım oranını arttırmak (Revuelta & Ponsoda, 1998). Literatürde çeşitli madde kullanım sıklığı stratejileri önerilmiş ve bu yaklaşımlar üç genel kategoriye ayrılmıştır: (i) rastgele seçme yöntemleri (ii) koşullu yöntemler (iii) tabakalı yöntemler (Davis, 2004; Davis & Dodd, 2003; Way, 1998). Kullanım pratikliği açısından rastgele seçme yöntemlerinden bilgi koşullu tesadüfi yöntemi oldukça yaygın olarak kullanılır (Kingsbury & Zara, 1989). Bu yöntemde, bireyin kestirilen geçici yetenek düzeyinde maksimum bilgi sağlayan beş madde belirlendikten sonra bunlar arasından rastgele bir madde seçilir.

Madde Takımı Tabanlı BBT

Tek bir uyarana bağlı birden fazla maddelerin oluşturduğu madde takımlarından oluşan testlerin BBT uygulamaları farklı yöntemlerle yapılabilmektedir. Bu madde takımlarının birer bağımsız maddeler gibi düşünülerek BBT uygulamalarının yapıldığı çalışmalar (Alkan, 2023; Kaya, 2022; Yao, 2019) olmasına rağmen son zamanlarda madde takımlarının bütün olarak ele alındığı BBT çalışmaları da sağladığı avantajlardan dolayı yaygınlaşmaktadır (Ersan-Cinar 2022; Keng, 2009; Murphy ve diğerleri, 2010; Yao; 2019).

Madde takımlarının çok kategorili olarak puanlandığı ve çok kategorili MTK modellerinin temel alındığı yöntemlerde, BBT uygulamalarının bazı sınırlılıkları vardır (Davis & Dodd 2003; Boyd ve diğerleri, 2013). Ayrıca çok kategorili puanlama yaklaşımı, ölçü birimi olarak maddeleri değil madde takımlarını aldığından madde takımında yer alan maddeler üzerinde araştırma yapılmasına imkân sağlamaz (Boyd ve diğerleri, 2013; Wainer ve diğerleri, 2007). Madde takımlarının kendi içinde bir bütün olarak ele alındığı desenlerden biri OM- BBT (passage model)'dir. Bu desende bir madde takımı seçildiğinde, bireylerin madde takımı içindeki her bir madde üzerinde nasıl performans gösterdiğine bakılmaksızın madde takımının tamamına verilen cevaplarla ilgilenilir ve tek boyutlu MTK modeli temel alınır. Bu sebepten oldukça temel düzeydedir ve madde takımı etki derecesini de hesaba katmadığından yerel bağımsızlık varsayımının ihmaline sebebiyet verir. Yetenek kestirimini yanlış hesaplamasından dolayı yüksek yerel bağımlılık derecesine sahip madde takımlarında kullanılması pek önerilmez. Alan yazında bu deseni kullanırken yine Maksimum Fisher Bilgi yöntemine göre madde seçme gerçekleştirilmiştir (Murphy ve diğerleri, 2010; Yao, 2019).

MTTK modeli kullanılarak madde takımları oluşturulup puanlandığında, madde takımı parametresini de hesaba katan BBT uygulamalarının, madde takımı düzeyinde (madde takımları arasında), madde düzeyinde (madde takımları içinde ve madde takımları arasında) uyarlama gibi türleri söz konusu olabilmektedir (Boyd ve diğerleri, 2004; Ersan-Cinar, 2022; Keng, 2008; Wainer ve diğerleri, 2007). Madde takımı tabanlı BBT uygulamalarından en çok kullanılan ise yine madde takımlarının bir bütün olarak seçilmesine imkân sunan fakat madde takımı MTK modellerini ele alan sadece madde takımları arasında uyarlanabilen (between testlet/testlet level) desenlerdir. Bu yöntem, madde takımı tabanlı gerçek veri ile yapılan BBT uygulamalarında en yaygını olanıdır. Bu yöntemde belli bir kritere göre bir madde havuzundan rastgele bir madde takımı seçilir (Wainer & Kiely, 1987). Başka bir deyişle, böyle bir algoritma madde takımı düzeyinde uyarlanabilir, ancak madde düzeyinde doğrusal ve sabittir. Bu yöntem ile, uygulama için bir

madde takımı seçildiğinde, sınava giren bireylerin madde takımı içindeki her bir madde üzerinde nasıl performans gösterdiğine bakılmaksızın, madde takımıyla ilişkili belirli bir madde kümesine verilen performansı ile ilgilenilir. Bu BBT deseninde de OM-BBT deseninde olduğu gibi madde takımlarının her biri bir bütün olarak ele alınıp bu madde takımları arasında seçim yapılmıştır fakat madde takımı modelleri temel alınmıştır. Tek boyutlu MTK modelleri yerine MTTK modellerinin temel alınmasının sebebi madde takımlarındaki yerel bağımlılık derecesinin de dikkate alınmasıdır. Bu sebeple özellikle yüksek yerel bağımlılık derecesine sahip madde takımları içeren testlerde oldukça etkili yetenek kestirimlerinin yapılmasına olanak sağlamaktadır.

Madde takımına dayalı BBT uygulamalarında alternatif bir diğer seçim düzeyi, madde takımları arasında uyarlamaya ek olarak madde takımları içinde de seçme yöntemi uygulamaktır. Böylece, seçilen bir madde takımında yer alan maddeler, bireylerin farklı bir madde takımı alabileceği şekilde uyarlanabilir. Hem madde takımları arasında hem de madde takımı içinde uyarlamaya imkân veren madde takımı tabanlı BBT desenleri özellikle simülasyon çalışmalarında oldukça fazla kullanılmaktadır. Özellikle madde havuzunun oluşturulması gerçek veri uygulamalarında oldukça zor olduğundan genellikle simülasyon çalışmalarında yaygındır. Bu çalışmalarda etkili sonuçlar elde etmek için havuzdaki madde takımıyla ilişkili maddelerin birden çok versiyonu veya permütasyonu oluşturulur. Permütasyonlar iki amaç göz önünde bulundurularak oluşturulur: İlki permütasyonların paralel formlar olması, ikincisi ise madde takımıyla ilişkili her maddenin, permütasyonlardan en az birine atanmasıdır. Tabi ki bu uyarlama yaklaşımı hem madde takımları arasında hem de madde takımının içindeki maddeler arasında olduğu için BBT uygulamalarının doğruluğu ve verimliliğinin daha yüksek çıkması beklenir. Hatta alan yazında sadece madde takımları arasında değil, aynı zamanda madde takımları içindeki uyarlamanın, gizil özelliklerin daha doğru ve kesin bir şekilde tahmin edilmesi için BBT performansını artırdığı savunulmaktadır (Ersan-Cinar, 2022; Keng, 2008). Otomatik madde oluşturmadaki gelişmelerle birlikte, bir madde takımı için birçok madde geliştirilebilir (Gierl & Haladyna, 2012). Böylece daha fazla

uyarlanabilirlik, ölçümün verimliliğini, hassasiyetini ve test puanlarının doğruluğunu da arttırmaktadır.

İlgili Araştırmalar

Alan yazında madde takımı içeren testlerin farklı BBT yaklaşımlarına göre incelendiği çalışmaların sayısı oldukça azdır. Aşağıda BBT, OM-BBT ve MT-BBT uygulamalarının kullanıldığı bu çalışmalar özetlenmiştir:

Wainer ve Kiely (1987), işlevsel bir BBT geliştirmeye yönelik bazı sorunlara ve bu sorunların çözüm yollarına değinmiştir. Madde takımlarının BBT bağlamında ele alındığı ilk makalelerden biri olan bu çalışma ile madde takımları hakkında detaylı bilgi verilmiştir. Hiyerarşik ya da doğrusal olarak geliştirilebilen madde takımları, uyarlanabilir bir testin verimliliğini korurken, titiz bir geçerlik çalışmasının yapılmasını da gerektirmektedir. Aynı madde takımları üzerindeki performansın, yetenek tahmini elde etmek için toplanabileceği sonucuna ulaşılmıştır. Fakat varılan sonuç şu anda bilgisayar tarafından oluşturulan testlerin uygulanabilirliğine güvenmenin pratik olmadığı ve bu konu ile ilgili daha çok araştırmaya ihtiyaç duyulduğudur.

Davis ve Dodd (2003), yaptıkları simülasyon çalışmasında dört BBT sistemi ile okuma metinleri içeren gerçek veri seti (Tıp Fakültesi Giriş Testinin Sözel Muhakeme bölümündeki madde takımları) kullanarak madde kullanım sıklığı prosedürlerini incelemiştir. Birincisi maksimum bilgi madde seçimi kullanan geleneksel bir CAT, ikincisi optimal maruz kalma oranları için bir temel sağlamak üzere rastgele madde seçimi kullanmıştır. Üçüncüsü Lunz ve Stahl'ın randomizasyon prosedürünün bir varyasyonunu, dördüncüsü ise Luecht ve Nungester'in bilgisayarlı uyarlanabilir sıralı test (CAST) sistemini kullanmıştır. Bu test sistemlerinden elde edilen sonuçları ölçme kesinliği, madde havuzu kullanım sıklığı ve test güvenliği açısından karşılaştırmıştır. BBT deseninde madde seçim yöntemi olarak MFB ve 0,10 logit içi yöntemlerini kullanmışlardır. Madde havuzunda yer alan madde takımlarının parametrelerini çok kategorili MTK modellerinden KPM ile kestirmişlerdir. Araştırma

sonucunda optimum düzeyde madde seçim prosedürünü belirlemek için bir dizi simüle edilmiş sabit uzunlukta BBT çalıştırılmıştır. Optimum madde seçim prosedürünü belirlemek için bir dizi simüle edilmiş sabit uzunlukta BBT uygulaması çalıştırılmıştır. Sonuçlar hem randomizasyon prosedürünün hem de sıralı BBT sisteminin maruz kalma kontrolü ve ölçüm hassasiyeti açısından iyi performans gösterdiğini ve tüm değişkenler dikkate alındığında sıralı BBT sisteminin en iyi genel çözümü sağladığını göstermiştir. Sonuçlar hem randomizasyon prosedürünün hem de CAST'ın maruz kalma kontrolü ve ölçüm hassasiyeti açısından iyi performans sergilediğini göstermiştir. Sonuç olarak madde güvenliği ve madde havuzu kullanımı açısından CAST deseni iyi sonuçlar vermiştir.

Boyd (2003), ölçüm hassasiyeti, test güvenliği ve uygulama kolaylığı açısından üç farklı çok kategorili MTK modeli altında altı farklı maruz kalma kontrol mekanizmasının performansını araştırmıştır. Mevcut çalışmada incelenen üç model, kısmi kredi, genelleştirilmiş kısmi kredi ve kademeli yanıt modelleridir. Madde maruz kalma oranlarını kontrol etmek için 0,10 logit dahilinde rastgele, Sympson-Hetter, koşullu Sympson-Hetter, a-Tabakalı ve geliştirilmiş a-Tabakalı prosedürler uygulandı. Her üç MTK modeli için de benzer sonuçlar elde edilmiştir ve 0,10 logit dahilinde prosedürlerin, ölçüm hassasiyeti ve uygulama kolaylığı üzerindeki etkisi göz önüne alındığında maruz kalma oranlarını kontrol etmek için en iyi seçenek olduğu bulunmuştur.

Keng (2008), MTTK modelini temel alan madde takımı tabanlı BBT ve ÇAT desenlerinin performansını incelemiştir. Çalışmadaki madde takımı tabanlı BBT deseni; (i) madde takımları arasında uyarlama-madde takımı düzeyinde (ii) madde takımları arasında ve içinde uyarlama- madde düzeyinde olmak üzere iki yöntemi içermektedir. Bu durumda aslında bu çalışma üç farklı deseni (madde takımları arasında uyarlama, madde takımları içinde ve arasında uyarlama, çok aşamalı test desenleri) karşılaştırmıştır. Manipüle edilen test koşulları arasında madde takımı sayısı, madde havuzu boyutu ve sınava girenlerin yetenek dağılımı yer almaktadır. Madde havuzu, geniş ölçekli bir okuma metinleri içeren elde edilen verilere dayalı olarak MTTK modeli tarafından kalibre edilmiş madde

parametreleri kullanılarak oluşturulmuştur. Üç test tasarımı, ölçme etkililiği ve madde kullanım sıklığı özelliklerine göre değerlendirilmiş ve üç uyarlanabilir test tasarımının da benzer ve iyi ölçüm doğruluğu sağladığı, kabul edilebilir madde kullanım sıklığı oranlarına sahip olduğu sonucuna ulaşılmıştır. Fakat genel olarak madde düzeyindeki BBT daha iyi ölçüm hassasiyeti sağlamış sonra ÇAT deseni ve madde takımı düzeyindeki BBT deseni izlemiştir. Özetle, üç uyarlanabilir test tasarımının da ölçüm doğruluğunun hem genel hem de θ ölçeğinin farklı noktalarında oldukça benzer olduğu sonucuna ulaşılmıştır.

Murphy ve diğerleri (2010), gerçek veriye dayalı simülasyon çalışmasında farklı yerel madde bağımlılığına sahip koşullarda madde seçme yöntemlerinin yetenek tahmini üzerindeki etkisini incelemiştir. Buradaki diğer önemli amaç ise madde takımlarındaki bağımlılık MTTK altında doğru şekilde modellendiğinde ve madde yanıt teorisi (MTK) altında yanlış modellendiğinde madde seçme yöntemleri ile MTK modelleri arasındaki ilişkiyi incelemektir. Genel olarak 3PL MTK ve MTTK modelleri, MTK modeli kapsamındaki standart hataların aşırı iyimser olması dışında, koşullar genelinde benzer şekilde performans gösterdi. Diğer önemli bir bulgu ise kullanılan MTK modelleri ile madde seçim yöntemleri (MFI, MPWI ve MEPV) arasında herhangi bir etkileşim olmadığı yani MTK ve MTTK modellerinin benzer koşullar arasında her seçim yönteminde de benzer şekilde performans gösterdiği anlamına gelir. Madde takımı sayısı arttıkça yetenek kestirimlerinin standart hatası, gerçek ve kestirilen yetenekler arasındaki korelasyon ve yanlılık neredeyse aynı çıkmıştır. Tabi ki MTK doğru modellendiğinde yani yerel bağımsızlık sağlandığında en etkili sonuç elde edilmiştir. Standart hata tahminlerinin genel ortalaması, MTK modelleri altında tahmin edildiğinde 0,29 civarında, MTTK modelleri altında tahmin edildiğinde ise 0,31 ile 0,32 civarındaydı. Ayrıca Chen ve diğerleri (2000) bulgularıyla tutarlı olarak madde seçim yönteminin standart hata tahminlerinin büyüklüğünü etkilediğine dair bir belirtiye rastlanmamıştır.

Yao (2019); okuma metninden oluşan madde takımı tabanlı üç BBT desenini, madde seçme tekniğini araştırmak ve daha iyi yetenek kestirimi için öneriler geliştirmek amacıyla

incelemiştir. BBT madde takımı seçiminde üç farklı desen geliştirmiştir: (a) madde takımı etki modeli (b) okuma metni modeli (P) (c) tek boyutlu MTK modeli. Madde takımı etki modelinin okuma metni veya tek boyutlu modellerden daha iyi performans gösterdiği bulunmuştur. Madde takımlarındaki maddeler yerel bağımlı olduğu için okuma metni modeli ve tek boyutlu MTK modelleri, yetenek kestirimini şişirme eğilimi göstermiş dolayısıyla daha küçük standart hata ve daha büyük yanlılık elde edilmiştir. Madde takımı etkisinin dikkate alındığı MTTK modeli en iyi güvenilirliğe ve en küçük yanlılık değerlerine sahipti. Ayrıca havuzda daha fazla madde bulunduğunda MTTK modellerinin madde havuzu kullanımı tek boyutlu MTK modelleri kadar iyi olabiliyor. Gerçek uygulamada, eğer tek boyutlu MTK modeli, veri setine en iyi şekilde uyuyorsa, o zaman madde takımı etkisi güçlü değildir ve tek boyutlu MTK modeli kullanılmalıdır.

Ersan-Cinar (2022), madde takımı tabanlı BBT uygulamalarında yetenek kestiriminin doğruluğu ve verimliliği artırmak için (i) madde takımları arasında (ii) madde takımı içinde ve arasında uyarlanabilirlik üzerinde çalışmıştır. Simülasyon çalışmasında manipüle edilen diğer değişkenler, farklı madde takımı/madde seçim yaklaşımları, yetenek kestirim yöntemleri, madde takımı sayısı ve yerel bağımlılık derecesidir. Çalışmanın önemli bulgularından biri, sadece madde takımları arasında değil, aynı zamanda madde takımları içindeki uyarlamanın, yetenek düzeylerinin daha doğru ve kesin bir şekilde tahmin edilmesi için BBT performansını artırmasıdır. Bu çalışmanın sonuçları, önceki araştırma bulgularına paralel olarak madde parametre iyileşmesinin arttığını göstermiştir (Keng, 2008); bu nedenle, eğer madde takımlarında kişi parametrelerinin kestirilebilmesi için madde takımı boyutuna göre yeterli miktarda madde varsa hem madde takımı içi hem de madde takımları arası uyarlama önerilir. Ek olarak madde havuzu kullanımında da farklı BBT desenleri arasında tutarlı bir farklılık gözlenmemiştir. Yeterli madde havuzuna sahip bir BBT uygulamasında daha uzun testler, kişi parametrelerinin tahmini açısından daha iyi performans gösterdiği sonucuyla tutarlılık sağlanmıştır (Weiss, 1982). Örneğin, bu çalışmada 40 maddelik bir test, 20 maddelik bir testle karşılaştırıldığında önemli ölçüde

daha iyi birey parametreleri üretmiştir. Ayrıca madde takımı sayısı aynı olduğunda, seçilen madde takımı içindeki madde sayısını artırmak yerine madde takımı sayısını artırmak daha iyi performans göstermiştir. Yetenek kestirim yöntemi EAP ve MAP yöntemleri, birey yetenek kestiriminde ve havuz kullanımı açısından oldukça benzer performans göstermiştir. Bununla birlikte, gizil özellik; sürekliliğinin aşırı uçları dışında, MAP, özellikle madde takımı sayısının kısa olduğu durumlarda, EAP'nin ürettiği hatalardan daha düşük bir hata üretti. Bu bulgu, Seo ve Weiss'in (2015) iki faktörlü BBT deseninde kullandığı MAP yönteminin daha yüksek hassasiyet ürettiğini gözlemlediği önceki araştırma bulgularıyla tutarlıdır.

Frey ve diğerleri (2016), çok boyutlu uyarlanabilir test (MAT) ile rastgele madde takımı etkisine dayalı madde takımı tabanlı BBT uygulamasını bazı koşullar altında karşılaştırmıştır. Madde takımı etki büyüklüğü (0, 0,5, 1.0, 1,5) ve madde takımı uzunluğu (3, 6, 9) koşulları altında yapılan simülasyon çalışmalarının sunucuna göre madde takımı tabanlı BBT yönteminden elde edilen sonuçlar daha güvenilir bulunmuştur. Tabi ki madde takımı etki derecesi arttıkça ölçüm etkililiği azalmaktadır. Ayrıca elde edilen diğer bir önemli bulgu ise madde takımı uzunluğunun belli bir aralıkta olması gerektiğidir.

Kang ve diğerleri (2022) madde takımlarının rastgele madde takımı etkileri gösterdiği durumlarda çeşitli puanlama modellerinin performansını incelenmiştir. Araştırma için dört model dikkate alınmıştır: (i) kısmi puan modeli (ii) çok kategorili madde olarak madde takımı modeli (iii) rastgele etki madde takımı modeli (iv) sabit etki madde takımı modeli. Modellerin performansı, madde takımlarının sıfırdan farklı rastgele etkilere sahip olduğu iki uyarlanabilir testte değerlendirilmiştir. Çalışmanın genel sonuçları, belirgin rastgele madde takımı etkilerine rağmen tüm desenlerde yetenek kestirimi ve sınıflama doğruları karşılaştırılabilir nitelikteydi. Bu çalışmanın amacı teknolojiyle güçlendirilmiş yenilikçi maddelere yönelik pratik puanlama modellerini araştırmaktır. Madde takımı etkilerini modellemek için en mantıklı yaklaşım, rastgele etki modeli olsa da pratikte çok fazla etki etmemiştir. Yetenek kestiriminde kısmi puan modeli ve sabit etki madde takımı modeli rastgele etki modelinden daha iyi performans gösterdiği sonucuna ulaşılmıştır. Yanıt

verilerinin rastgele etki modelinden alınmış olmasına rağmen, kısmi puan modeli ve sabit etki modeli en küçük yanlılığı, RMSE'yi ve gerçek yeteneklerle en yüksek korelasyonu göstermiştir. Test takımlarının büyük etkileri olduğunda rastgele etki modeli alternatif puanlama modellerine göre açıkça üstün performans göstermiştir. Bu simülasyon çalışmasında rastgele madde takımı etkilerine rağmen kısmi puan modeli veya sabit etki modeli tercih edilirse, yetenek tahminindeki yanlılığı ve bunun sonucunda sınıflandırma üzerindeki etkiyi en aza indirmek için yeterince fazla sayıda madde takımının kullanılması gerektiğini göstermektedir. Özet olarak daha basit ve pratik olarak daha uygulanabilir modellerin, yanıt verileri rastgele etki test takımı modelini takip ettiğinde bile rastgele etki modeli kadar iyi performans gösterdiğini veya ondan daha iyi performans gösterebileceği sonucuna varılmıştır.

Pak (2017), simülasyon çalışmasında BBT uygulamalarının madde takımlarıyla performansını incelemiştir. Madde takımı tabanlı BBT'ler için yaygın uygulama, uyarlanabilir bir madde takımı seçmek ve madde takımında bir araya getirilen tüm maddeleri yönetmektir. Ancak BBT uygulamalarının doğruluk ve etkililiği, bu uyarlanabilirlik yaklaşımıyla tam olarak gerçekleştirilemeyebilir. Bu çalışmada madde takımı tabanlı BBT uygulamalarında doğruluğu ve verimliliği artırmak için madde takımı arasında ve içindeki uyarlanabilirlik üzerinde çalışmıştır. Manipüle edilen diğer değişkenler, farklı madde takımı/madde seçim yaklaşımları, yetenek tahmin yöntemleri, madde takımı sayısı/test takımı sayısı ve madde takımı bağımlılığının büyüklüğü olarak belirlenmiştir. Genel olarak Rasch madde takımı modellerinde madde takımı etkisinin (γ) iyi işlendiği sonucuna ulaşılmıştır. Çalışmanın diğer bir önemli bulgusu, sadece madde takımları arasında değil, aynı zamanda madde takımları içindeki uyarlamanın, bireyin yeteneğinin daha doğru ve etkili tahmini için BBT performansını artırmasıdır. Genel olarak, kısa bir madde takımı uzunluğunun daha iyi sonuçlar verdiği ve bir madde takımı uzunluğunun fazla olduğu durumlarda madde takımlarının karma formatlı bir sınava dönüştürülebileceği sonucuna varılmıştır.

İlgili Araştırmalar Özeti

Madde takımlarından oluşan testlerde BBT desenleri ile ilgili çalışmaların sayısının uluslararası alan yazında oldukça az, ulusal alan yazında ise hiç olmadığı görülmektedir. Yapılan çoğu çalışma simülasyon çalışması olup gerçek veri setine dayalı simülasyon çalışmalarında ise gerçek madde parametreleri kullanılmıştır. MTTK temelli BBT uygulamaları ile MTK temelli BBT uygulamalarına dayalı oluşturulan BBT desenlerinin ölçme kesinliği ve madde havuzu kullanımına ilişkin sonuçların benzer olduğu çalışmaların yanı sıra farklı olduğu çalışmalar da bulunmaktadır. Bunun nedeni olarak veri setindeki madde takımlarının yerel bağımlılık dereceleri, madde takımı uzunlukları ve madde havuzunun büyüklükleri gösterilebilir. Ayrıca MTTK temelli BBT desenlerinde hem madde takımları arası hem de madde takımı içinde uyarlamaya dayalı desen en etkili sonucu vermiştir. Ek olarak yetenek kestirim yöntemi EAP ve MAP yöntemleri, birey yetenek kestiriminde ve havuz kullanımı açısından oldukça benzer performans göstermiştir. Bununla birlikte, gizil özellik; sürekliliğinin aşırı uçları dışında, MAP EAP'nin ürettiği hatalardan daha düşük bir hata üretmiştir.

Bölüm 3

Yöntem

Bu bölümde araştırmanın türü, çalışma grubu, veri toplama süreci, veri toplama araçları ve verilerin analizi hakkında bilgiler verilmiştir.

Araştırmanın Türü

Araştırmada, madde takımlarından oluşan farklı BBT yaklaşımlarının gerçek veriye dayalı madde havuzu kullanılarak çeşitli simülasyon koşulları altında etkililiğinin incelenmesi amaçlanmıştır. Ayrıca madde havuzu oluşturulurken her bir form, bilgisayar tabanlı test olarak uygulanmıştır. Bu çalışma ile literatürde yer alan farklı MTK modellerine dayalı olarak geliştirilen farklı BBT desenlerini karşılaştırma amaçlandığından bu yönüyle araştırma bir yöntem karşılaştırmasıdır.

Verilerin Elde edilmesi

Araştırma kapsamında kullanılan veri seti, 220K161 numaralı “İngilizce Hazırlık Sınıflarında Bilgisayarda Bireyselleştirilmiş Test Kullanımı” başlıklı Tübitak projesi kapsamında Bilkent Üniversitesi YDYO’da İngilizce hazırlık eğitimi alan öğrencilere 2022-2023 Güz döneminde bilgisayar tabanlı olarak Uygulanan Yabancı Dil Seviye Belirleme Sınavları (YDSBS)’ndan elde edilmiştir. Bu sınavlardan elde edilen veriler kullanılarak madde havuzu oluşturulmuş ve gerçek veriye dayalı simülasyon uygulaması gerçekleştirilmiştir.

Madde havuzunun oluşturulmasında ve simülasyon uygulamasında verileri kullanılan YDSBS, öğrencilerin İngilizce seviyelerini belirlemek amacıyla okuduğunu anlama ve dinleme becerilerini ölçen çoğunluğu tek bir uyarana bağlı olan okuma ve dinleme bölümünden oluşmaktadır. Bağımsız madde sayısının çok az olması, çalışmanın madde takımı tabanlı BBT uygulamasını içermesi ve sadece tek boyutluluğa dayalı okuduğunu anlama becerisinin ölçülmek istenmesinden dolayı araştırma kapsamında

sadece okuduğunu anlama becerilerini ölçen aynı paragrafa bağlı üç maddenin yanıtlanmasını gerektiren madde takımları kullanılmıştır. Dolayısıyla çalışma kapsamında, 77 madde takımından oluşan toplam 231 maddeden elde edilen veriler kullanılmıştır.

Madde havuzunun özellikleri ve büyüklüğü BBT'lerde etkili kestirimler yapılmasında oldukça önemlidir (Davey & Nering, 2002; Stocking, 1994; Wainer, 1990). Fakat tek bir okuma parçasına bağlı okuduğunu anlama maddelerinden oluşan madde havuzunun bir öğrenciye tek seferde uygulanması neredeyse imkânsızdır. BBT uygulamalarının etkililiği açısından geniş madde havuzu oluşturmak için zaman ve motivasyon gibi önemli değişkenler nedeniyle madde havuzunda yer alan maddelerin tamamı aynı gruba aynı zamanda uygulanamamıştır. MTK'nın değişmezlik özelliği varsayımına göre testteki maddelerin sayısı fazla olduğunda aynı evren içerisinde yer alan farklı gruplara alt testler farklı zaman dilimlerinde uygulanarak ölçekleme çalışmaları yapılabilir (Kolen & Brennan, 2004). Fakat bu maddelerin tek bir madde havuzunda birleştirilebilmesi için ortak maddeler ya da ortak bireyler yoluyla ölçekleme çalışmalarının yapılması gerekir (Nydick & Weiss, 2009). Bu nedenle bu çalışmada dokuz farklı set hazırlanarak ölçekleme yapılabilmesi için ortak maddelerden oluşan bir madde havuzu oluşturulmuştur. Her bir sette dokuz madde (üç madde takımı) ortak olup setler minimum 30 maksimum 36 maddeden oluşmaktadır. Her bir veri setini ise minimum 101 maksimum 104 bireyin cevapları oluşturmuştur.

Tablo 1

Formlara Göre Madde ve Cevaplayan Sayısı

	form1	form2	form3	form4	form5	form6	form7	form8	form9
Madde Takımı	10	10	12	12	12	11	12	11	11
Madde	30	30	36	36	36	33	36	33	33
Cevaplayan	100	103	100	101	100	102	103	102	104

Madde sayısının yanında kaliteli maddelerden oluşan madde havuzunun kullanılması BBT'lerde etkili kestirimler yapılmasında oldukça önemlidir (Choi & van der Linden, 2018; Hendrickson, 2007; Flaughner, 2000; Wainer, 1990). Havuzun kaliteli ve iyi

hazırlanmış olması; havuzdaki maddelerin yüksek düzeyde bilgi vermesi, ölçmek istediği amaca uygun olması ve farklı yetenek düzeylerindeki adaylara hitap etmesi ile ilgilidir (Parshall ve diğerleri, 2002; Reckase, 2010). Bu nedenle bu çalışmadaki maddeler proje kapsamında İngilizce alanında uzman kişiler tarafından yazılmış ve farklı dil uzmanları tarafından kontrol edilerek uzman görüşlerine dayalı düzeltmeler yapılmıştır. Daha sonra 60 kişilik bir grup ile pilot çalışma yapılarak madde istatistikleri hesaplanmış, gerekli düzeltmeler yapılmıştır.

Veri Analizi

Araştırmanın amacına uygun olarak veri analizi iki aşamada gerçekleştirilmiştir. İlk aşama BBT desenleri için madde havuzunun geliştirilme süreci boyunca yapılan tüm analizleri; ikinci aşama araştırma sorularının detaylı cevaplarını içermektedir. Tüm analizler R yazılım programı kullanılarak gerçekleştirilmiştir. İlk aşamada R programında yer alan mirt (Chalmers, 2023), psych (Revelle, 2022) paketleri; ikinci aşamada ise yine mirt paketi ile ek algoritmalar yazılarak kullanılan MASS (Ripley, 2023) paketinden faydalanılmıştır.

Madde Havuzunun Oluşturulması

Madde havuzundaki maddelerin kalitesi BBT çalışmalarında oldukça önemlidir (Choi & van der Linden, 2018; Hendrickson, 2007; Flaughner, 2000; Wainer, 1990). Bu nedenle madde havuzundaki maddeler proje kapsamında uzmanlar tarafından hazırlanmış, dil ve teknik uygunluk açısından da maddeleri hazırlayan uzmanlardan farklı uzmanlarından görüşler alınmıştır. Bu görüşler doğrultusunda, gerekli düzeltmeler yapıp madde havuzunda kalan tüm maddelerin yer aldığı form hazırlanmıştır. Ön pilot uygulamasından elde edilen bilgiler doğrultusunda taslak formda gerekli düzeltmeler yapılmış ve test pilot uygulama için 60 kişilik bir gruba uygulanmıştır. Pilot uygulamadan elde edilen veriler kullanılarak madde ve test istatistikleri incelenmiş ve testin okuduğunu anlama bölümü çoktan seçmeli 256 madde ile tamamlanmıştır. Madde istatistikleri, konu kapsamı ve uzman görüşlerine göre taslak form, dokuz farklı paralel forma ayrılmıştır. Bilkent Üniversitesi

Yabancı Diller Yüksekokulu Hazırlık sınıfında okumakta olan her seviyeden öğrenci grubuna uygulanmıştır. Fakat bu çalışmada her bir formdaki madde takımlarında üçer tane madde olmasına dikkat edilmiş ve 77 madde takımı toplam 231 maddeden oluşan madde havuzu kullanılmıştır.

Model Veri Uyumunun İncelenmesi

BBT uygulamalarında gerçekleştirilen madde kalibrasyonunda kullanılan en yaygın yöntem model uyum indekslerini istatistiksel olarak değerlendirmektir. Ancak, bir modelin verilere uyup uymadığını değerlendirmek için örneklem büyüklüğü dikkate alınmalıdır (Hambleton ve diğerleri, 1991). MTK modellerinde etkili bir kalibrasyon için en az 500 test katılımcısına ihtiyaç vardır (DeMars, 2010; Han ve diğerleri, 2019; Harwell & Janosky, 1991; Reise & Yu, 1990; Stone, 1992). Fakat Rasch modeli az sayıda örneklem için bile güvenilir sonuçlar sağlayabilir (Lord, 1980). Bu çalışmada, her bir formda minimum 100, maksimum 104 cevap örüntüsü bulunduğu için Rasch modelleri kullanılmıştır. Chen ve diğerleri (2014) çeşitli örneklemelerde (10, 50, 100, 250) parametre kestirimlerinin Rasch analizleri üzerindeki etkisini değerlendirdiği çalışmada 100 ve üzeri örneklemelerde tutarlı sonuçlar elde etmiştir. Hatta kayıp verilerin olmadığı 50 kişilik örneklemelerde bile sonuçlar kabul edilebilir düzeyde çıkmıştır. BBT ve OM-BBT desenleri için tek boyutlu Rasch Modeli; MT-BBT için Rasch MTTK modeli uyum indeksleri ve ANOVA ile olabilirlik oranı testi ile karşılaştırılmış ve χ^2 -fark testi ile test edilmiştir (du Toit, 2003). Özetle, okuma bölümlerindeki maddeler Rasch MTK ve Rasch MTTK modelleri ile ayrı ayrı iki şekilde kalibre edilmiştir ve model uyum indeksleri Ek A'da verilmiştir. Her bir test formu için iki MTK modelinin uyum indekslerini görülmektedir. Düşük AIC, SABIC ve BIC değerlerine sahip modelin veriye daha iyi uyduğu söylenebilir (Rijmen, 2010) Uyum indeksleri, Rasch MTTK için AIC, SABIC ve BIC model indekslerinin değerlerinde düşüşler olduğunu göstermektedir. Ancak bu değerler çok yüksek değildir. Olabilirlik testinin *p-değeri* sadece Form 1, 2, 4 ve 9 için anlamlıdır. Bu sonuç da bu dört form için Rasch MTTK modelin verilere Rasch MTK modeline göre daha iyi uyum sağladığını göstermektedir. Bu sonuç alan yazında madde

takımları içeren testlerde yerel bağımlılığı dikkate alan MTTK modelinin tek boyutlu MTK modellere göre daha iyi model veri uyumu sağladığı sonucuna ulaşan çalışmalarla tutarlıdır (Demars, 2012; Yılmaz-Koğar & Kelecioğlu, 2017). Fakat bu çalışmada p değeri anlamlı olmasına rağmen ki-kare fark değerleri de oldukça düşüktür. Ayrıca logLik değerlerindeki artışların tüm test formları için fazla yüksek olmadığı görülmektedir. Bu test formları kullanılarak bir madde havuzu oluşturulduğunda, amacımıza uygun olarak iki farklı MTK modelinden biri ya da ikisi de kullanılarak kalibrasyon işlemi gerçekleştirilebilir ve kalibrasyon sonrası da tüm madde havuzu için model veri uyum indeksleri incelenerek veriye uygun model seçilebilir.

Yerel Bağımsızlık Varsayımının İncelenmesi

Bu çalışmada, dört yerel bağımlılık türünden aynı madde takımındaki yerel bağımlılık çiftleri türü ele alınmıştır. Çünkü yerel bağımlılığın aynı madde takımı içinde yer alan madde çiftleri arasında olması önemlidir (Kim ve diğerleri, 2007). Bunun nedeni, madde takımları içindeki yerel bağımlılığının kaynağının, yapıyla ilgisiz varyansların kaynağını oluşturan ve nüans faktör olarak kabul edilen madde takımı faktörüdür. Her bir sete ait yerel bağımlılık dereceleri YB χ^2 ve Q3 istatistiklerine göre Tablo 2'de hesaplanmıştır.

Tablo 2

Yerel Bağımlılık İçeren Madde Takımlarının Yüzdesi

Form	YB- $\chi^2 > 3$	YB-Q3 > 0,2	Form	YB- $\chi^2 > 3$	YB-Q3 > 0,2
1	6 (%60)	4 (%40)	6	5 (%45,5)	4 (%36,4)
2	5 (%50)	5 (%50)	7	7 (%63,5)	6 (%54,5)
3	2 (%16,7)	2 (%16,7)	8	7 (%63,5)	6 (%54,5)
4	7 (%58,3)	6 (%50)	9	5 (%45,5)	4 (%36,4)
5	7 (%58,3)	5 (%42,7)			

*YB:Yerel bağımlılık

Tablo 2 incelendiğinde her bir formda yer alan madde takımlarının içinde yer alan madde çiftleri arasındaki yerel bağımlı madde takımı sayısı 2 (%16,7) ile 7 (%63,5) arasında değişmektedir. Kesme puanları χ^2 ve Q3 için sırasıyla 3 ve 0,2 olarak alınmıştır ve sadece

aynı madde takımı içindeki madde çiftlerinde yerel bağımlılık hesaplanmıştır. Örneğin, Form 1'deki 10 testin 6'sında yerel bağımlılık gözlemlenmiştir. Bu yerel bağımlılık aynı madde takımı içindeki herhangi bir madde çifti arasında olabilir. Form 7 ve 8 en yüksek, Form 3 ise en düşük yerel bağımlılığa sahip madde takımı oranlarına sahiptir. Bu sonuç, yerel bağımsızlık varsayımının bu test formlarında sağlanamayabileceğini göstermektedir. Bu nedenle, yerel bağımlılık derecesini dikkate alan uygun MTK modelleri seçilebilir.

Madde havuzu kalibrasyonu sonrasında Rasch MTTK modele göre madde takımı etki parametreleri elde edilmelidir. Bu çalışmada madde takımı etkisinin (γ) sırasıyla 0, 0.5 ve 1 değerlerini madde takımı etkisi yok, orta ve büyük derecede madde takımı etkisi olarak yorumlanmıştır (Wang ve diğerleri, 2002; Wang & Wilson, 2005). Bu veri setinde ise MTTK modeline göre madde takımı etki parametreleri incelendiğinde; maddelerin yarısından fazlası (%56) düşük düzeyde yerel bağımlılığa sahiptir. Yüksek düzeyde yerel bağımlılığa sahip sadece 14 (%18,2) madde takımı vardır.

Tek Boyutluluk Varsayımının İncelenmesi

Açımlayıcı faktör analizi (AFA), tek boyutluluk varsayımı altında her bir test formuna uygulanmıştır. Bunun için öncelikle KMO ölçümü ve Bartlett'in küresellik testi dokuz test formu için hesaplanmıştır. KMO değerleri tüm test formlarında 0,78 ile 0,85 arasında değişmektedir. Ayrıca Bartlett testi de anlamlı bulunmuştur ($p < 0,05$). Sonuçlar verilerin AFA için uygun olduğunu desteklemektedir. Öz değerler, açıklanan varyans oranları ve hiyerarşik omega sonuçları tüm test formları için ayrı olarak hesaplanmış ve Tablo 3'te verilmiştir.

Tablo 3**Boyutluluk Analizi Sonuçları**

Form	Boyut 1 (öz değer-açıklanan varyans)	Boyut 2 (öz değer-açıklanan varyans)	Boyut 3 (öz değer -açıklanan varyans)	$R_{1/2}$	H_w
1	5,44 (%18,1)	1,65 (%5,5)	1,10 (%3,7)	3,29	0,47
2	4,99 (%16,6)	1,54 (%5,1)	1,11 (%3,7)	3,24	0,48
3	5,26 (%17,4)	1,72 (%4,8)	1,33 (%4,3)	3,06	0,46
4	4,11 (%22,5)	1,34 (%3,7)	1,21 (%3,4)	4,55	0,66
5	6,55 (%18,2)	1,99 (%5,4)	1,17 (%3,3)	3,29	0,61
6	6,11 (%18,5)	1,23 (%3,74)	1,00 (%3,0)	2,65	0,54
7	6,79 (%13,75)	2,40 (%6,7)	1,34 (%3,7)	2,82	0,48
8	6,04 (%18,2)	2,38 (%7,2)	1,20 (%3,6)	2,54	0,63
9	5,99 (%18,1)	2,34 (%4,1)	1,14 (%3,4)	2,56	0,56

* $R_{1/2}$: Birinci özdeğerin ikinciye oranı; H_w : Hiyerarşik omega

Tablo 3 incelendiğinde, 1'den büyük öz değerlerin sayısı dört ile altı arasında değişmektedir ancak ilk üç faktöre ilişkin öz değerler ve açıklanan varyans oranları incelenmiştir. Birinci faktör tarafından açıklanan varyansların hiçbirisi %22,5'i aşmamaktadır, bu sonuç test formlarındaki maddelerin muhtemelen birden fazla örtük özelliği ölçtüğünü göstermektedir. Bununla birlikte, birinci ve ikinci öz değerlerin oranları 2,54 ile 4,55 arasında değişmektedir ve test formları genelinde ikinci öz değerler üçüncü öz değerden çok daha büyük değildir. Ayrıca hiyerarşik omega değeri de tüm formlar için ölçüt değeri olan 0,70'in altında çıkmıştır. Elde edilen sonuçlara göre verinin tek boyutlu olduğunu söylemek güçtür. Aslında elde edilen bu sonuç madde takımlarının yapısı ile de tutarlıdır. Çünkü madde takımlarındaki her bir madde takımı genel bir boyut dışında özel bir boyut ile de ilişkilendirilmektedir. Her bir madde takımının ilişkilendirildiği bu faktörler alan yazında da nüans faktörler olarak değerlendirilmektedir (Ersan-Cinar, 2022; Murphy ve diğerleri, 2010). Bu nedenle bu çalışmada da benzer yaklaşım izlenmiştir.

Madde Havuzu Kalibrasyonu

BBT çalışmalarında madde havuzunu geniş tutmak amacıyla ortak maddeler içeren ve paralel olarak oluşturulan testler bireylere uygulanır. Madde havuzundaki maddelerin tamamının aynı anda aynı gruba uygulanması neredeyse imkansızdır. Böylece dokuz farklı

bağlantı maddesiyle dokuz farklı paralel form oluşturulmuştur. MTK'nın değişmezlik özelliği varsayımına göre madde sayısı fazla olduğunda bu maddeler aynı evrendeki farklı gruplara farklı zamanlarda uygulanabilmektedir (Kolen, 2004). Ancak ayrı test formlarından farklı kalibrasyonlara ilişkin parametre tahminleri aynı ölçekte olamaz (Laurier, 1999; Zeng, 1996). Bu tasarımda testlerin ve grupların eşdeğer olmasına gerek olmamasına rağmen (Vale, 1986) iki önemli nokta şudur: Ortak maddelerin orta güçlükte olması gerektiği ve ortak madde sayısının her bir formda yer alan toplam madde sayısının yaklaşık yüzde 20-25'i kadar olmasıdır (Hambleton ve diğerleri, 1991).

Farklı maddelerden oluşan testler ortak maddeler kullanılarak birbiriyle bağlanır ve bu işlem sonucundaki puanlar tek bir ölçeğe getirilir. (Nydick & Weiss, 2009). Kalibrasyon işlemi, parametrelere belli bir dönüşüm uygulayarak tek bir ölçeğe yerleştirilme sürecini içerir (Kolen, 2004). Farklı gruplara uygulanan testlerdeki maddeler, çeşitli kalibrasyon yöntemleriyle tek bir madde havuzunda toplanabilir. Bu yöntemlerden en yaygın kullanılanları, eş zamanlı ve ayrı kalibrasyonlardır. Eş zamanlı kalibrasyon adından da anlaşılacağı gibi farklı alt gruplardan elde edilen verileri aynı anda birlikte kalibre ederek dikey ölçekleme oluşturur (Meng, 2007). Özetle ayrı kalibrasyonda her bir test formlarından elde edilen madde ve yetenek parametreleri her bir form için ayrı kestirilirken eş zamanlı kalibrasyonda aynı anda bir kez kestirilmektedir. Tüm parametrelerin aynı anda kestirilme yöntemi bu parametrelerin aynı ölçeğe getirilmelerini garantilemektedir (Hanson & Beguin, 2002). Ayrı kalibrasyon sonucu elde edilen madde ve yetenek parametreleri ölçek dönüştürme işlemi ile aynı ölçeğe getirilebilir. Eş zamanlı kalibrasyonda ortak maddeler ya da ortak bireyler kullanılarak madde ve yetenek parametreleri aynı ölçeğe getirilebilir (Kim ve diğerleri, 2009, Nydick & Weiss, 2009). Test eşitleme sürecinin de ilk basamağı olan ayrı ya da eş zamanlı (concurrent) kalibrasyon ile madde ve yetenek parametreleri kestirilebilir. Bu çalışmada eş zamanlı kalibrasyon kullanılarak dokuz farklı formdan bağlantı maddeleri (ankor/ortak maddeler) sayesinde tek bir madde havuzu oluşturulmuş, madde ve yetenek

parametreleri ortak maddeler aracılığıyla aynı ölçek düzeyine getirilmiştir. Bu madde havuzuna ilişkin veri toplama deseni Tablo 4'te verilmiştir.

Tablo 4

Madde Havuzuna İlişkin Veri Toplama Deseni

Toplam 915 birey	9 madde	21 madde	21 madde	27 madde	27 madde	27 madde	24 madde	27 madde	24 madde	24 madde	Toplam 231 madde
Grup1 (100)	Ortak maddeler	Form1									30
Grup2 (103)		Form2									30
Grup 3 (100)		Form3									36
Grup 4 (101)		Form4									36
Grup 5 (100)		Form5									36
Grup 6 (102)		Form6									33
Grup 7 (103)		Form7									36
Grup 8 (102)		Form8									33
Grup 9 (104)		Form9									33

Tablo 4'te satırlarda maddeler, sütunlarda ise bireyler yer almaktadır. Ortak maddeler kullanılarak farklı gruplara uygulanan dokuz farklı form ile ölçekleme işlemi gerçekleştirilmiş ve toplam 915 bireyin 231 maddeye ilişkin madde ve yetenek parametreleri belirlenmiştir.

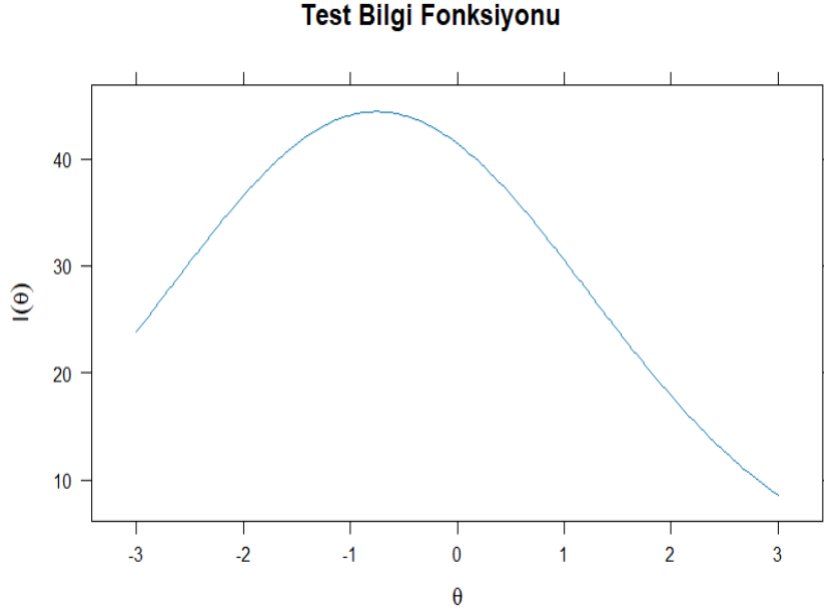
Bu çalışma üç farklı BBT deseninde kullanılacak iki farklı madde havuzunu içermektedir. Birinci madde havuzu MT-BBT deseninde kullanılacak olan MTTK modelini içermektedir. Her bir formda yaklaşık 100 kişi olması sebebiyle kalibrasyon işlemi için en iyi modelin Rasch MTTK modeli olduğuna karar verilmiştir. Ayrıca 2PL MTTK modeline göre kalibrasyon yapıldığında veri setinde yakınsama problemi olduğu görülmüştür. Bu nedenle Rasch MTTK modeline göre kalibrasyon işlemi yapılmıştır; madde, yetenek ve madde takımı etki parametreleri kestirilmiştir. İkinci madde havuzu ise BBT ve OM-BBT desenleri için Rasch MTK modeline göre kalibre edilmiştir. Alan yazında farklı MTK modelleri temeline

dayanan BBT desenleri ile yapılan çalışmalarda bu desenlere uygun kalibre edilmiş farklı madde havuzlarının kullanıldığı görülmektedir (Keng, 2008; Wang 2017; Yao, 2019).

Sonuç olarak bu çalışmada kullanılan madde takımı ve madde parametreleri, bir İngilizce seviye belirleme sınavından elde edilen 915 gerçek öğrenci yanıt verileri kalibre edilerek kestirilmiştir. Gerçek veri okuduğunu anlama metinleri değerlendirmelerine dayalı olarak iki madde havuzu üzerinden kestirilen madde parametreleri ile oluşturulmuştur. Gerçek veri 915 bireyin yanıtladığı 77 madde takımı toplam 231 maddeyi içermektedir. Madde takımlarının tamamı üçer maddeden oluşmaktadır. Verilere MTTK ve tek boyutlu MTK modelleri uygulanmıştır. Kestirilen 77 madde takımı etki parametresi 0,13 ile 1,79 arasında değişmektedir. Madde havuzu 1’de 77 madde takımı etki parametresi vardır. Madde Havuzu 2’de ise 231 tek boyutlu MTK madde parametresi vardır. Her iki model de verilere eşit derecede uyum sağlamaktadır. Güçlük parametreleri için iki modele ilişkin tahminler arasındaki korelasyonlar 0,98’dir; AIC değerleri Rasch MTK ve MTTK modelleri için sırasıyla 32140,24 ve 32134,57’dir. İki model arasındaki loglik değerlerine göre yapılan kıkare fark testi de anlamlı çıkmamıştır ($\chi^2=79,83$; $sd=77$). Havuz 1, 77 madde takımı etki parametresi içeren MT-BBT deseni için kullanılmıştır. Havuz 2, tek boyutlu Rasch modele göre kestirilen 231 maddeye ait parametre değerlerini içeren BBT deseni için kullanılmıştır. Havuz 2, ayrıca OM-BBT deseninde de kullanılmıştır. Ayrıca test bilgi fonksiyonları da madde bilgi fonksiyonlarının toplamı olduğundan BBT uygulamalarında madde havuzu hakkında bilgi vermektedir (de Ayala, 2009). Test bilgi fonksiyonu grafiği Şekil 4’te verilmiştir.

Şekil 4

Test Bilgi Fonksiyonu Grafiği



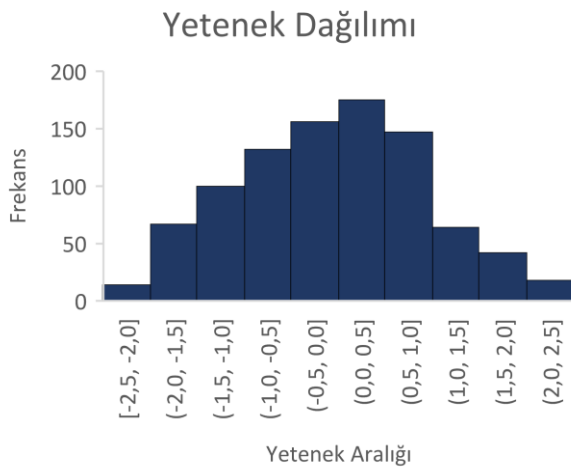
Şekil 4 incelendiğinde madde havuzundaki çoğu maddenin orta güçlükte veya kolay olduğu yani bazı maddelerin bireylere kolay geldiği görülmektedir. Bu durum da test bilgisi fonksiyonunun yetenek ölçeğinin $(-1,0)$ aralığında zirve yapmasına neden olmuştur. Bu çalışmada madde takımlarından oluşan madde havuzu, alan yazında gerçek veri seti ile oluşturulan madde takımlarının yer aldığı madde havuzu ile de benzerlik göstermektedir (Keng, 2008). Ek olarak madde havuzunun etkililiğini değerlendirmek amacıyla, test maddelerinin güçlük parametreleri ile cevaplayıcıların yetenek parametrelerinin aynı ölçekte yer aldığı birey-madde haritası (wright-map) çizilmiştir (Ek-B). Hem maddelerin hem de bireylerin dağılımlarının yer aldığı bu haritada sağa çarpık bir dağılım gözlenmiştir. Bunun anlamı uygulanan örnekleme tüm maddelerin güçlük düzeyine denk gelebilecek bireylerin yer almamasıdır. Fakat MTK'nın parametre değişmezliği özelliği sayesinde bu herhangi bir sorun teşkil etmemektedir. Sonuç olarak İngilizce okuma becerisini ölçmek için madde havuzunda farklı yetenek düzeylerine uygun yeterli sayıda maddenin bulunduğu sonucuna varılmıştır.

Yetenek Parametrelerinin Elde Edilmesi

BBT, MT-BBT ve OM-BBT desenlerinde kullanılan bireylere ait yetenek parametreleri gerçek veri setinden elde edilen kovaryans matrisi kullanılarak çok değişkenli normal dağılımdan 78 boyutlu olacak şekilde türetilmiştir (Keng, 2008; Murphy, 2010). Analiz sürecinde gerçek madde parametrelerine bağlı olarak bireylere ilişkin yetenek parametreleri üretilmiş ve her bir analiz için aynı yetenek parametreleri kullanılmıştır (Özdemir, 2015). BBT ve OM-BBT desenlerinde tek boyutlu MTK modeli kullanıldığından genel (birinci-genel) yetenek parametresi kullanılmıştır. MT-BBT deseninde ise her bir madde takımına ait yetenek parametresi ve bir genel yetenek parametresi olmak üzere 78 yetenek parametresi kullanılmıştır. Gerçek veri setinden elde edilen yetenek parametrelerinin dağılımları Şekil 5'te verilmiştir.

Şekil 5

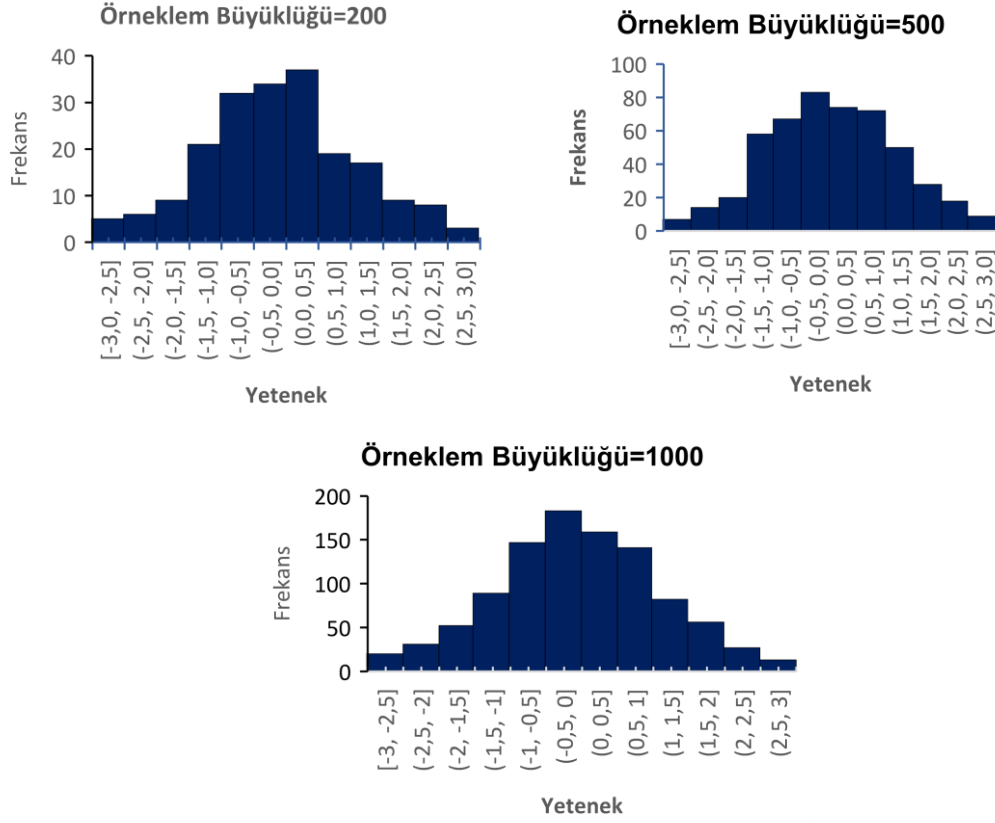
Gerçek Veri Setine ait Yetenek Dağılımı



Şekil 5 incelendiğinde, veri setinden madde parametrelerine uygun elde edilen genel yeteneklerin dağılımları yaklaşık olarak normal dağıldığı görülmektedir. Bu çalışmada örneklem büyüklüğü bir koşul olarak belirlendiğinden 200, 500 ve 1000 örneklem büyüklüklerine göre elde edilen yetenek parametre dağılımlarının histogram grafikleri ise Şekil 6'da sunulmuştur.

Şekil 6

Yetenek Parametrelerinin Dağılımı



Örnekleme hatasını minimize etmek amacıyla BBT uygulamalarındaki simülasyon çalışmalarında en az 10 tekrarlar yapılmıştır (Cheng ve diğerleri, 2015; Keng, 2010; Kim, 2010; Murphy ve diğerleri, 2010; Özdemir & Gelbal, 2022; Pak, 2017; Yao, 2013; Yao, 2019). Örnekleme hatasını azaltmak için en az 25 tekrar yapılması gerektiği de önerilmektedir (Harwell ve diğerleri, 1996). Bu çalışmada ise verilerin analizi sürecinde 30 tekrar yapılarak elde edilen değerlerin ortalaması kullanılmıştır.

Araştırma Deseni

Bu çalışmada farklı BBT yaklaşımlarında gerçek veri seti üzerinden elde edilen farklı simülasyon koşullarından elde edilen sonuçlar karşılaştırılmıştır. Araştırmanın odak noktasını madde takımlarından oluşan testlerde farklı MTK modelleri kullanılarak

oluşturulan BBT desenlerinin incelenmesi oluşturmaktadır. Bu amaçla araştırma kapsamında yürütülen simülasyon koşullarına ait ayrıntılar Tablo 5'te yer almaktadır.

Tablo 5

Araştırma Kapsamında Manipüle Edilen Değişkenler

Değişken	Koşul	Koşul Sayısı
Desen	BBT, OM-BBT, MT-BBT	3
Yetenek kestirim yöntemi	EAP, MAP	2
Madde takımı sayısı	6,9,12	3
Örneklem büyüklüğü	200,500,1000	3

Tablo 5'te görüldüğü üzere üç farklı BBT deseni (BBT, MT-BBT, OM-BBT), üç farklı madde takımı uzunluğu (6,9,12), iki farklı yetenek kestirimi (EAP, MAP) üç farklı örneklem büyüklüğü (200, 500,1000) koşulları ele alınarak gerçek veri seti parametreleri kullanılarak simülasyon çalışması gerçekleştirilmiştir.

Test uzunluğu arttıkça ölçme sonuçlarının güvenilirliği artmaktadır (Crocker & Algina, 1986). Böylelikle bireyler hakkında daha geçerli ve güvenilir bilgiler elde edilebilmektedir. Bu nedenle test uzunluğu arttırıldıkça BBT desenlerinin kestirimleri hakkında bilgi edinmek ve en optimal madde takımı sayısını belirlemek amacıyla madde takımı sayısı bir değişken olarak manipüle edilmiştir. Yabancı dil sınavlarında okuduğunu anlama metinlerinde oluşan madde takımları fazla tercih edilmektedir. BBT uygulamaları kullanıldığında geleneksel kâğıt-kalem testlerinin en az yarı uzunluğundaki testlerle aynı ölçme kesinliğine ulaşılmaktadır (Weiss, 1982). Aynı zamanda alan yazında yaklaşık 20 maddeden oluşan testler kısa, 40 maddeden oluşan testler uzun olarak ele alındığından bu çalışmada kısa, orta ve uzun olmak üzere üç kategorili madde takımı sayısı değişken olarak belirlenmiştir (Demars, 2010; Glas ve diğerleri, 2000). Bu durum dikkate alınarak madde takımı sayısı 6, 9 ve 12 olarak belirlenmiştir. BBT simülasyonlarında eşit uzunlukta (üçer maddeden oluşan) madde takımları olacağı için test uzunlukları 3'ün katları olacak şekilde seçilmiştir.

MTK'ye dayalı yöntemlerden, Maksimum Olabilirlik Kestirim Yöntemi (ML- Maximum Likelihood), Maksimum Sonsal Kestirim Yöntemi (MAP) ve Beklenen Sonsal Kestirim Yöntemi (EAP) alan yazında sıklıkla kullanılmaktadır. MAP yöntemi, iki faktörlü BBT uygulamalarında EAP yönteminden daha yüksek hassasiyet üretir (Özdemir, 2015; Seo & Weiss, 2015). Üstelik MAP, EAP yönteminden daha az hesaplama süresi gerektirmektedir (Yao, 2012). Öte yandan, herhangi bir yinelemeli arama süreci gerektirmeyen ve madde takımları ve madde takımı tabanlı BBT ile gizli özellik tahmininde uzun süreli bir geçmişe sahip olan EAP'nin hesaplanması daha kolaydır (Boyd ve diğerleri, 2013; Glas ve diğerleri, 2000; Frey ve diğerleri, 2016; Keng, 2008; Murphy ve diğerleri, 2010). Benzer şekilde Yao (2019), BBT çalışmasında sınırlı iki faktörlü modelde yetenekler için MAP tahmin yöntemini kullanmıştır. Bu çalışmada hem EAP hem de MAP, yetenek kestirimlerinde kullanılarak sonuçlar karşılaştırılmıştır. Yetenek kestirimleri için “mirt” R paketi (Chalmers, 2022) kullanılmıştır.

MTTK tabanlı çalışmaların çoğunda örneklem büyüklüğü 1000 olarak sınırlandırılmıştır (Bradlow ve diğerleri, 1999; Wainer ve diğerleri, 1992). Fakat alan yazında 1000'den küçük örneklemle MTK modellerinin çalışıldığı araştırmalar da bulunmaktadır (Wang ve diğerleri, 2022) Bu amaçla özellikle madde takımlarından oluşan BBT çalışmalarında analiz sürecinin kullanılan modelden kaynaklı fazla zaman almasından dolayı 200 ve 500 kişilik örneklem büyüklükleri seçilmiş ve 1000 örneklem ile birlikte koşul olarak eklenmiştir.

Özet olarak 3 (BBT deseni) x 3 (madde takımı sayısı) x 3 (örneklem büyüklüğü) x 2 (yetenek kestirim yöntemi) olmak üzere toplam 54 koşul incelenmiş ve örneklem hatasını minimum düzeye indirmek için her bir koşul için 30 replikasyon (tekrar) yapılarak analizler tamamlanmıştır. Analizlerin sonucunda 30 tekrar sonucu elde edilen değerlerin ortalaması kullanılmıştır. Üç test deseninin her birinde uygulanacak maddeleri ve madde takımlarını uyarlanabilir bir şekilde seçmek için Maksimum Fisher Bilgi fonksiyonu kullanılmıştır. Bireylerin geçici/anlık yetenek kestirimleri için EAP (Bock & Mislevy, 1982) ve MAP yetenek

kestirim yöntemleri kullanılmıştır. Sabit uzunlukta bir durdurma kuralı tüm test tasarımları için kullanılmış ve testteki madde sayısı toplam madde takımı sayısı koşuluna göre belirlenmiştir.

BBT Desenlerinin Oluşturulması

Araştırmada BBT deseni olarak BBT, MT-BBT VE OM-BBT desenleri ele alınmıştır. Bu desenlerin oluşturulmasında “MASS” ve “mirt” paketleri kullanılarak yeni kodlar yazılmıştır. Üç farklı BBT deseni, iki farklı yetenek kestirim yöntemi, üç farklı madde takımı sayısı ve üç farklı örneklem büyüklüğü koşulları altında incelenmiştir.

BBT ile her bir madde bağımsız olarak ele alındığı MTK modeli olarak tek boyutlu Rasch MTK modeli kullanılmıştır. Bu durumda madde havuzu 231 maddeden oluşmakta olup madde takımı etkisi yani madde takımlarındaki yerel bağımlılık derecesi göz ardı edilmiştir. Bireylerin başlangıç yetenek kestirim değeri sıfır olarak belirlenmiş ve geçici yetenek kestirim değerinde en fazla bilgi veren 5 madde arasından rastgele seçilen madde bireye uygulanmıştır. Sonlandırma kuralı olarak sabit uzunluk yöntemi kullanılmıştır. Madde takımı sayısı araştırmanın bağımsız değişkenlerinden birisi olduğu için testi sonlandırmada üç farklı madde takımı sayısı (6, 9, 12) kullanılmıştır. Aynı şekilde üç farklı örneklem büyüklüğü de bağımsız değişkenlerden biri olarak kullanılmıştır.

OM-BBT deseni ile madde takımlarının ele alındığı MTK modeli olarak yine tek boyutlu Rasch MTK modeli kullanılmıştır. Bu durumda madde havuzu 77 madde takımından oluşmakta olup madde takımı etkisi yani madde takımlarındaki yerel bağımlılık derecesi göz ardı edilmiştir. Bireylerin başlangıç yetenek kestirim değeri sıfır olarak belirlenmiş ve geçici yetenek kestirim değerinde en fazla bilgi veren beş madde takımı arasından rastgele seçilen madde bireye uygulanmıştır. Burada madde takımları üçerli maddelerden oluştuğu için Maksimum Fisher Bilgi fonksiyonu her üç maddenin toplamı olarak hesaplanmıştır. Tüm madde takımlarındaki madde sayısı birbirlerine eşit olduğu için madde bilgilerinin toplamı yerine ortalaması da aynı sonucu vermektedir. Sonlandırma kuralı olarak yine sabit uzunluk yöntemi kullanılmıştır. Madde takımı sayısı araştırmanın bağımsız değişkenlerinden birisi

olduğu için testi sonlandırmada üç farklı madde takımı sayısı (6, 9, 12) kullanılmıştır. Aynı şekilde üç farklı örneklem büyüklüğü de bağımsız değişkenlerden biri olarak kullanılmıştır.

MT-BBT deseni ile madde takımlarının ele alındığı MTK modeli olarak Rasch MTTK modeli kullanılmıştır. Bu durumda madde havuzu 77 madde takımından oluşmakta olup madde takımı etkisi yani madde takımlarındaki yerel bağımlılık derecesi de dikkate alınmıştır. Bireylerin başlangıç yetenek kestirim değeri sıfır olarak belirlenmiş ve geçici yetenek kestirim değerinde en fazla bilgi veren beş madde takımı arasından rastgele seçilen madde bireye uygulanmıştır. Burada madde takımları üçerli maddelerden oluştuğu için Maksimum Fisher Bilgi yöntemi her üç maddenin toplamı olarak hesaplanmıştır. Tüm madde takımlarındaki madde sayısı birbirlerine eşit olduğu için madde bilgilerinin toplamı yerine ortalaması da aynı sonucu verecektir. Sonlandırma kuralı olarak yine sabit uzunluk yöntemi kullanılmıştır. Madde takımı sayısı araştırmanın bağımsız değişkenlerinden birisi olduğu için testi sonlandırmada üç farklı madde takımı sayısı (6, 9, 12) kullanılmıştır. Aynı şekilde üç farklı örneklem büyüklüğü de bağımsız değişkenlerden biri olarak kullanılmıştır.

Madde ve Madde Takımı Seçme Yöntemi

Tüm BBT desenlerindeki MTK ve MTTK modelleri için madde seçim yöntemi olarak Maksimum Fisher Bilgi yöntemi kullanılmıştır. Bu yöntemin seçilmesinin temel nedeni, birçok araştırmacı tarafından madde takımı temelli BBT araştırmalarında etkin bir şekilde kullanılmasıdır (Boyd ve diğerleri, 2013; Glas ve diğerleri, 2000; Keng, 2008; Murphy ve diğerleri, 2010). Alan yazında da bu şekilde madde takımı/madde seçimi sırasında madde bağımlılığı göz ardı edildiği çalışmalar yer almaktadır (Boyd ve diğerleri, 2013; Glas ve diğerleri, 2000; Keng, 2008; Murphy ve diğerleri, 2010). Bu nedenle, uyarlanabilirliğin yalnızca madde takımları arasında olduğu durumlarda, madde takımları, Rasch MTK modeli altında hesaplanan her bir madde takımının toplam madde bilgisine dayalı olarak seçilmiştir. Fakat bir testte yer alan madde takımlarındaki madde sayıları farklı olabilir ve bu durumda madde takımı seçimi için madde takımı bilgi fonksiyonunun kullanılması, daha fazla madde içeren madde takımını daha avantajlı hale getirecektir. Bu nedenle Frey ve diğerleri (2016),

madde takımı bilgi fonksiyonunun kendisi yerine madde takımının ortalama madde bilgisi fonksiyonunun kullanılması önermiştir. Bu çalışmada her madde takımında aynı sayıda madde olduğundan madde bilgi fonksiyonlarının toplamı ve ortalaması aynı sonuca sebep olacaktır. Özetle, madde takımı modelleme yaklaşımlarına ilişkin karar; madde, madde takımı ve test bilgi fonksiyonlarını etkilemekte ve aynı şekilde BBT uygulamalarında madde ve madde takımı seçimi bu bilgi fonksiyonlarına bağlı gerçekleşmektedir. Bu durum da BBT performansını önemli derecede etkilemektedir. Çünkü tüm madde takımlarında aynı sayıda madde yer almaktadır. BBT'lerde MTTK için de madde seçiminde tek boyutlu MTK modelinin kullanılmasının nedeni, bireye özgü madde takımı etkisinin önceden bilinmemesidir (Glas ve diğerleri, 2000). Bu durumda MT-BBT desenlerinde MFB yöntemi kullanıldığında, madde takımı etkileri sistematik olmayan hata olarak kabul edilir. Ancak bu yöntem madde takımı/madde seçimi sırasında madde takımı yapısının göz ardı edilmesi nedeniyle de eleştirilmektedir (Frey ve diğerleri, 2016; Yao, 2019). Fakat yine de bu çalışmada kullanılabilirliği sebebiyle bu yöntem tercih edilmiştir.

Değerlendirme Kriteri

Her bir BBT yaklaşımı için farklı test koşullarını içeren gerçek veriye dayalı simülasyon çalışmaları ise çalışmanın ikinci aşamasını oluşturmaktadır. Bu aşamada her bir koşul sonucunda araştırmanın bağımlı değişkenleri olan gerçek (θ) ve kestirilen yetenek ($\hat{\theta}$) değerleri arasındaki korelasyon (r), standart hata (SH), hata kareleri ortalamasının karekökü (Root Mean Square Error – RMSE), gerçek ve kestirilen θ değerleri arasındaki farkın ortalaması olan yanlılık (BIAS), bu farkın mutlak değeri olan mutlak yanlılık (Mean Absolute BIAS-MAB) değerleri, madde kullanım sıklığı, kullanılmayan madde sayısı ve test çakışma oranları hesaplanmıştır.

Her bir BBT desenine ait yapılan ölçmenin doğruluğunu incelemek amacıyla elde edilen gerçek ve kestirilen yetenekler arasındaki Pearson Momentler Korelasyon Katsayısı

$r(\theta, \hat{\theta})$ değerleri karşılaştırılmıştır. Bu korelasyon katsayısı yetenek kestirimleri arasındaki uyum katsayısı olarak da adlandırılmaktadır.

BBT, MT-BBT ve OM-BBT desenlerini ölçme kesinliği açısından karşılaştırmak için RMSE, BIAS ve MAB değerleri denklem (12)'de yer alan formüller kullanılarak hesaplanmıştır. Yetenek kestirimine ait standart hata değerleri ise mirt paketi yardımıyla yetenek kestirimi yapılırken hesaplanmıştır. θ_i gerçek yetenek, $\hat{\theta}_i$ kestirilen yetenek, N kişi sayısı olmak üzere;

$$\begin{aligned} RMSE &= \sqrt{\frac{\sum_{i=1}^N (\hat{\theta}_i - \theta_i)^2}{N}} \\ BIAS &= \frac{\sum_{i=1}^N (\hat{\theta}_i - \theta_i)}{N} \\ MAB &= \frac{\sum_{i=1}^N |\hat{\theta}_i - \theta_i|}{N} \end{aligned} \quad (12)$$

Ayrıca BBT desenlerinin madde havuzu kullanımına ilişkin her madde için madde kullanım sıklığı, kullanılmayan madde sayısı ve test çakışma oranları hesaplanarak incelenmiştir. Bu değerler test güvenliği değişkenlerini temsil etmede de kullanılmaktadır (Boyd, 2003).

Madde kullanım sıklığı oranı yani madde ifşa/maruziyet oranı aşağıdaki denklem (13) ile hesaplanmıştır.

$$\text{Madde kullanım sıklığı oranı}(r_j) = \frac{n_m}{n_T} \quad (13)$$

Madde kullanım sıklığı oranı, maddenin uygulandığı birey sayısının (n_m) testi alan toplam birey sayısına (n_T) bölünmesiyle hesaplanmaktadır (He ve diğerleri, 2014). Üç test yaklaşımını madde havuzu kullanımı açısından karşılaştırmak için ise hiç kullanılmayan madde sayısının oranları incelenmiştir.

Test çakışma oranı ise aşağıdaki denklem (14) ile hesaplanmıştır (Chen ve diğerleri, 2003).

$$Test\ çakışma\ oranı = \frac{n_T \sum_{j=1}^J r_j^2}{k(n_T-1)} - \frac{1}{n_T-1} \quad (14)$$

Denklemden yer alan k test uzunluğunu, r_j ise madde kullanım sıklığı oranını ifade etmektedir.

Bölüm 4

Bulgular, Yorumlar ve Tartışma

Bu bölümde öncelikle araştırma problemine dair tüm koşullardan elde edilen bulgular sunulmuştur. Ardından sırasıyla her bir alt probleme göre bulgulara yer verilmiştir.

Araştırma Probleminin Ölçme Kesinliği ve Doğruluğuna İlişkin Bulgular ve Yorumlar

Çalışmanın araştırma problemi olan “BBT, MT-BBT ve OM-BBT desenlerinin ölçme kesinliği ve doğruluğu farklı koşullar altında ne düzeyde değişim göstermektedir?” sorusuna cevap vermek üç farklı BBT desenlerinden hesaplanan yetenek kestirimleri karşılaştırılmıştır. Tablo 6’da yer alan sütunlar sırasıyla yetenek kestirimi, örneklem büyüklüğü ve madde takımı sayısı koşullarını içermektedir. Bu koşullara göre farklı BBT desenlerinden elde edilen yetenek kestirimlerine ilişkin RMSE, MAB, BIAS, SH ve r değerleri hesaplanmış ve Tablo 6’da verilmiştir.

Tablo 6*Farklı BBT Desenlerine Göre Tüm Koşullardan Elde Edilen Bulgular*

Yetenek Kestirimi	Örneklem Büyüküğü	Madde Takımı Sayısı	BBT					MT-BBT					OM-BBT				
			RMSE	BIAS	MAB	SH	r	RMSE	BIAS	MAB	SH	r	RMSE	BIAS	MAB	SH	r
EAP	200	6	0,494	-0,041	0,397	0,480	0,910	0,598	-0,041	0,470	0,484	0,870	0,509	-0,011	0,406	0,482	0,902
		9	0,427	-0,028	0,335	0,404	0,927	0,497	-0,054	0,384	0,428	0,915	0,426	-0,005	0,337	0,408	0,933
		12	0,385	-0,019	0,302	0,354	0,949	0,484	-0,027	0,382	0,380	0,920	0,364	-0,004	0,287	0,361	0,952
	500	6	0,474	0,005	0,392	0,483	0,907	0,547	-0,026	0,442	0,482	0,897	0,471	0,019	0,378	0,488	0,916
		9	0,425	0,002	0,342	0,406	0,931	0,476	-0,038	0,383	0,428	0,916	0,415	0,015	0,324	0,414	0,938
		12	0,377	0,007	0,302	0,357	0,947	0,438	0,045	0,342	0,380	0,929	0,358	0,011	0,285	0,366	0,952
	1000	6	0,482	-0,004	0,385	0,482	0,909	0,587	-0,031	0,468	0,482	0,889	0,488	-0,012	0,393	0,480	0,918
		9	0,407	0,004	0,318	0,404	0,939	0,502	-0,056	0,397	0,428	0,907	0,432	-0,024	0,330	0,407	0,936
		12	0,350	0,008	0,275	0,354	0,955	0,441	-0,046	0,346	0,381	0,929	0,377	-0,011	0,313	0,360	0,952
MAP	200	6	0,489	-0,039	0,384	0,481	0,922	0,587	-0,018	0,449	0,460	0,875	0,494	-0,034	0,402	0,483	0,915
		9	0,431	-0,024	0,341	0,405	0,932	0,497	-0,025	0,396	0,395	0,919	0,424	-0,016	0,354	0,410	0,938
		12	0,392	-0,029	0,303	0,356	0,949	0,444	-0,020	0,350	0,349	0,935	0,382	-0,019	0,315	0,364	0,949
	500	6	0,471	-0,003	0,380	0,484	0,923	0,524	-0,007	0,415	0,465	0,894	0,469	0,022	0,376	0,491	0,916
		9	0,419	0,006	0,332	0,408	0,933	0,458	-0,009	0,368	0,398	0,921	0,403	0,019	0,326	0,418	0,939
		12	0,363	0,005	0,292	0,360	0,951	0,408	-0,007	0,327	0,351	0,938	0,357	0,011	0,285	0,369	0,952
	1000	6	0,488	0,001	0,381	0,481	0,907	0,544	-0,012	0,436	0,466	0,885	0,496	-0,014	0,395	0,483	0,907
		9	0,420	0,007	0,331	0,406	0,935	0,451	-0,029	0,357	0,394	0,924	0,415	-0,010	0,329	0,410	0,936
		12	0,368	0,008	0,291	0,357	0,951	0,411	-0,027	0,323	0,349	0,938	0,359	-0,014	0,285	0,364	0,953

Yetenek kestirimi, örneklem büyüklüğü ve madde takımı sayısı değişkenlerine göre BBT desenlerinin ölçme kesinliği ve doğruluğu incelendiğinde, madde takımı sayısının artmasının ölçme etkililiği ve kestirimini kesin olarak arttırdığı söylenebilir. Yine benzer şekilde genel olarak örneklem büyüklüğü de ölçme kesinliğini artırmaktadır. Özellikle 200 örneklem büyüklüğünden 500 örneklem büyüklüğüne çıkıldığında bu artış görülmektedir. Fakat 500 ile 1000 örneklem büyüklüklerinde ise bu artış miktarı çok az olup hatta tüm BBT yöntemlerinde ölçme kesinliğinin neredeyse aynı olduğu söylenebilir. Yetenek kestirimi yöntemine BBT yöntemlerinin ölçme kesinliğinin de değiştiği gözlenmektedir. Örneğin genel olarak BBT yönteminde EAP, MT-BBT ve OM-BBT yöntemlerinde ise MAP yetenek kestirimi yönteminin ölçme kesinliği daha yüksektir. Yorumlama kolaylığı açısından üç farklı BBT deseni için genel olarak tüm koşullara ait RMSE, BIAS, MAB, SH ve r değerlerinin minimum (min) ve max (maksimum) değerleri Tablo 7’de verilmiştir.

Tablo 7

BBT Desenlerine Göre RMSE, BIAS, MAB ve r Değerleri

BBT Desenleri	RMSE (min, max)	BIAS (min, max)	MAB (min, max)	SH (min, max)	$r(\theta_i, \hat{\theta}_i)$ (min, max)
BBT	(0.350, 0.494)	(-0.041, 0.021)	(0.275, 0.392)	(0.354, 0.483)	(0.907, 0.955)
MT-BBT	(0.398, 0.598)	(-0.056, 0.007)	(0.324, 0.490)	(0.381, 0.481)	(0.869, 0.941)
OM-BBT	(0.357, 0.509)	(-0.034, 0.022)	(0.285, 0.406)	(0.361, 0.491)	(0.902, 0.953)

Tablo 7 incelendiğinde, tüm koşullarda BBT deseninden hesaplanan RMSE değerlerinin 0,350 ile 0,494 aralığında, MT-BBT’den elde edilen RMSE değerlerinin 0,398 ile 0,598 aralığında, OM-BBT’den elde edilen RMSE değerlerinin ise 0,357 ile 0,509 aralığında değiştiği görülmektedir. Tüm koşullarda BBT deseninden elde edilen RMSE değerleri, diğer desenlerden elde edilen RMSE değerinden daha düşüktür. Fakat OM-BBT desenine ait RMSE değer aralığı ise BBT deseninden elde edilen RMSE değer aralığı ile oldukça benzerdir ve MT-BBT yönteminden düşüktür. Bu durum, BBT ve OM-BBT desenlerinden tüm koşullarda daha etkili yetenek kestirimleri elde edildiğini göstermektedir.

Tüm koşullarda elde edilen BIAS değerleri üç farklı BBT desenine göre incelendiğinde, yanlılık değerlerinin çok düşük ve tüm koşullarda benzer değerler olduğu görülmektedir. Yine RMSE değerlerinde olduğu gibi BIAS değer aralıklarında da BBT ve OM-BBT değerlerinin min ve max değer aralıklarının tüm koşullarda neredeyse aynı olduğu söylenebilir.

Tablo 7 incelendiğinde, tüm koşullarda BBT deseninden elde edilen MAB değerleri 0,275 ile 0,392 aralığında, MT-BBT deseninden elde edilen MAB değerleri 0,324 ile 0,490 aralığında, OM-BBT deseninden elde edilen MAB değerleri ise 0,285 ile 0,406 aralığında değişmektedir. Görüldüğü üzere tüm koşullarda BBT yönteminden elde edilen MAB değerleri, diğer yöntemlerden elde edilen MAB değerinden daha düşüktür. Bu durum, BBT'nin tüm koşullarda daha etkili yetenek kestirimi yaptığını göstermektedir. Fakat OM-BBT deseninden elde edilen MAB değerleri BBT deseninden elde edilenler ile oldukça benzerdir.

Standart hata değer aralıkları üç farklı desen için incelendiğinde, değer aralıklarının oldukça benzer olduğu görülmektedir. Fakat yine de en iyi sonucun BBT yöntemine ait olduğu görülmektedir. Kestirilen ve gerçek yetenek değerleri arasındaki korelasyon yani uyum değerleri incelendiğinde, öncelikle tüm desenlerde bu uyumun oldukça iyi düzeyde olduğu görülmektedir. Yine BBT ile OM-BBT yöntemlerinden elde edilen korelasyon değerleri oldukça benzer olmakla birlikte 0,90'dan yüksek bulunmuştur. Her bir alt problem madde takımı sayısı, örneklem büyüklüğü ve yetenek kestirim yöntemine göre BBT, OM-BBT ve MT-BBT desenlerinin ölçme kesinliğinin sonuçları karşılaştırılmış olup genel olarak üç desen için de ölçme kesinliğinin iyi derecede olduğu söylenebilir.

Alan yazın incelendiğinde, ölçme kesinliği açısından tek boyutlu MTK modellerinin standart hataları daha düşük çıksa da tüm koşullar genelinde benzer şekilde performans gösterdiği sonucuna ulaşılmıştır (Murphy ve diğerleri, 2010; Sireci ve diğerleri, 1991; Wainer ve diğerleri, 2007; Wainer & Thissen, 1996; Yen, 1993). Bu bulgunun da önceki araştırmalardan elde edilen sonuçlarla tutarlı olduğu söylenebilir.

Birinci Alt Araştırma Problemine İlişkin Bulgular ve Yorumlar

Birinci alt araştırma problemi “Madde takımı sayısına (6-9-12) göre BBT, MT-BBT ve OM-BBT yöntemlerinin gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, RMSE, BIAS, MAB ve SH değerleri nasıl değişmektedir?” Araştırmanın birinci alt problemine cevap bulabilmek için üç farklı desene ve üç farklı madde takımı sayısına göre RMSE, BIAS, MAB ve SH değerleri hesaplanmıştır. Elde edilen bu değerler Tablo 8’de verilmiştir.

Tablo 8

Madde takımı sayısına Göre RMSE, BIAS, MAB ve SH Değerleri

Madde takımı sayısı	BBT				MT-BBT				OM-BBT			
	RMSE	BIAS	MAB	SH	RMSE	BIAS	MAB	SH	RMSE	BIAS	MAB	SH
6	0,483	-0,014	0,387	0,482	0,565	-0,023	0,449	0,471	0,488	-0,005	0,392	0,485
9	0,422	-0,006	0,333	0,406	0,480	-0,035	0,381	0,412	0,419	-0,004	0,333	0,411
12	0,373	-0,003	0,294	0,356	0,438	-0,015	0,345	0,368	0,366	-0,004	0,295	0,364

Tablo 8’de ilk sütundaki madde takımı sayısı değişkeninde yer alan 6 madde takımı 18, 9 madde takımı 27 ve 12 madde takımı ise 36 maddeyi ifade etmektedir. Çünkü her bir madde takımı üçer maddeden oluşmaktadır. Madde takımı sayısı 6 için ortalama RMSE değeri sırasıyla en az BBT yöntemi için 0,483 iken, OM-BBT yöntemi için 0,488 ve MT-BBT yöntemi için 0,565’tir. Benzer şekilde madde takımı sayısı 6 için ortalama MAB değeri BBT için 0,387 iken OM-BBT için 0,392 ve MT-BBT için 0,449’dur. Madde takımı sayısı 6 için BBT deseni ile OM-BBT deseninin ölçme kesinliği oldukça benzer olup bu iki desenin MT-BBT’den daha iyi ölçme kesinliğine sahip oldukları görülmektedir.

Madde takımı sayısı 9 için ortalama RMSE sırasıyla en az BBT yöntemi için 0,422 iken, OM-BBT yöntemi için 0,419 ve MT-BBT yöntemi için 0,480’dır. Benzer şekilde madde takımı sayısı 9 uzunluğu için ortalama MAB değeri BBT için 0,333 iken OM-BBT için de 0,333 ve MT-BBT için 0,381’dir. Madde takımı sayısı 9 için BBT deseni ile OM-BBT deseninin ölçme kesinliği neredeyse aynı olup MT-BBT’den daha iyi ölçme kesinliğine sahip

oldukları görülmektedir. Madde takımı sayısı 6'ya benzer şekilde madde takımı sayısı 9 için de BBT ve OM-BBT desenlerinin benzer ve MT-BBT desenine göre daha iyi yetenek kestirimi yaptığı söylenebilir.

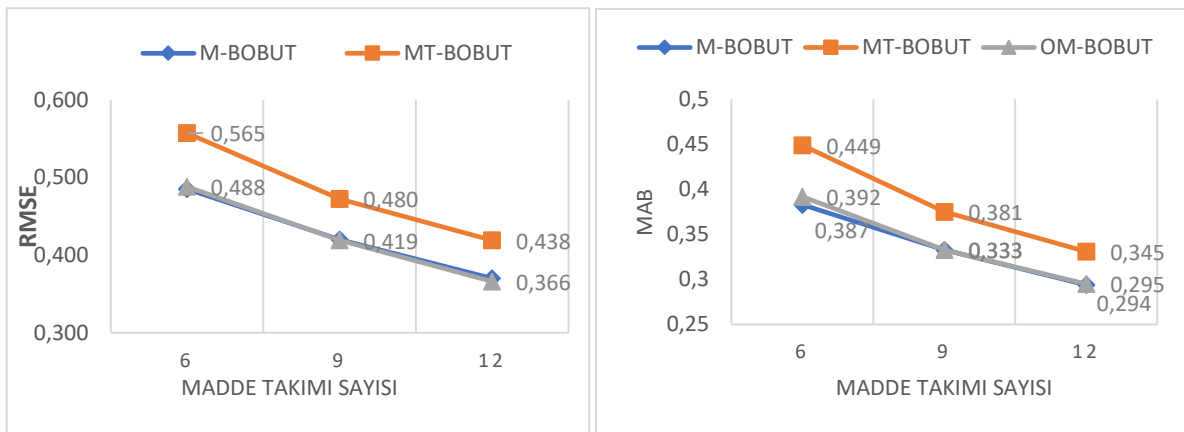
Madde takımı sayısı 12 için ortalama RMSE değeri sırasıyla en az OM-BBT yöntemi için 0,366 iken, BBT yöntemi için 0,373 ve MT-BBT yöntemi için 0,438'dir. Benzer şekilde madde takımı sayısı 12 için ortalama MAB değeri BBT için 0,294 iken OM-BBT için de 0,295 ve MT-BBT için 0,345'tir. Madde takımı sayısı 12 için BBT deseni ile OM-BBT deseninin ölçme kesinliği neredeyse aynı olup MT-BBT'den daha iyi ölçme kesinliğine sahip oldukları görülmektedir. Diğer test uzunluklarına benzer şekilde, madde takımı sayısı 12 için BBT ve OM-BBT desenlerinin MT-BBT desenine göre daha iyi yetenek kestirimi yaptığı söylenebilir.

Özet olarak her üç madde takımı sayısı (6, 9, 12) düzeyinde de RMSE, MAB ve SH değerlerine göre BBT deseninin OM-BBT deseni ile oldukça benzer MT-BBT desenine göre daha etkili yetenek kestirimi sunduğu görülmektedir.

Tablo 8'e göre RMSE ve MAB bulguları Şekil 7'de sunulmuştur.

Şekil 7

Madde Takımı Sayısına Göre RMSE ve MAB Değerleri



Şekil 7 incelendiğinde her üç madde takımı sayısı (6, 9, 12) düzeyinde de BBT deseninin OM-BBT ve MT-BBT desenlerinden daha düşük olduğu yani daha iyi ölçme kesinliği sunmaktadır. Madde takımı sayısı arttıkça RMSE değerlerinin üç desende de

azaldığı görülmektedir. BBT deseni için RMSE farkı 6-9, 9-12 için sırasıyla 0,049 ve 0,062; MT-BBT yöntemi için 0,085 ve 0,042; OM-BBT için ise 0,069 ve 0,053'tür. Madde takımı sayısı 6 ile 9 arasındaki RMSE farkı fazla iken, madde takımı sayısı 9 ile 12 arasındaki RMSE farkının azaldığı görülmektedir. Alan yazında da bu çalışmanın bulgularını destekleyecek şekilde madde takımı sayısı arttıkça RMSE değerlerinin azaldığı yani gerçek ve kestirilen yeteneklerin birbirlerine yaklaştığı bulgusunu desteklemektedir (Cheng ve diğerleri, 2015; Deng ve diğerleri, 2010; Erdem-Kara & Doğan, 2021; Keng, 2010;). Aslında sabit madde takımı sayısı sonlandırma kuralı olarak seçildiğinde, bu durum testin uygulanabilirliğini artırabilir. Fakat burada dikkat edilmesi gereken nokta en optimal madde takımı sayısına karar verebilmektir. Çünkü belli bir uzunluktan daha fazla olan testler hem BBT uygulamalarının amacından uzaklaşmasına hem de sınava giren bireylerde motivasyon düşüklüğüne neden olabilir. Bu nedenle alan yazında da genellikle orta uzunlukta (20 ile 30 arası) madde takımı sayısı önerilmektedir (Curi & Silvia, 2019; Tseng, 2016; van der Linden & Pashley, 2010).

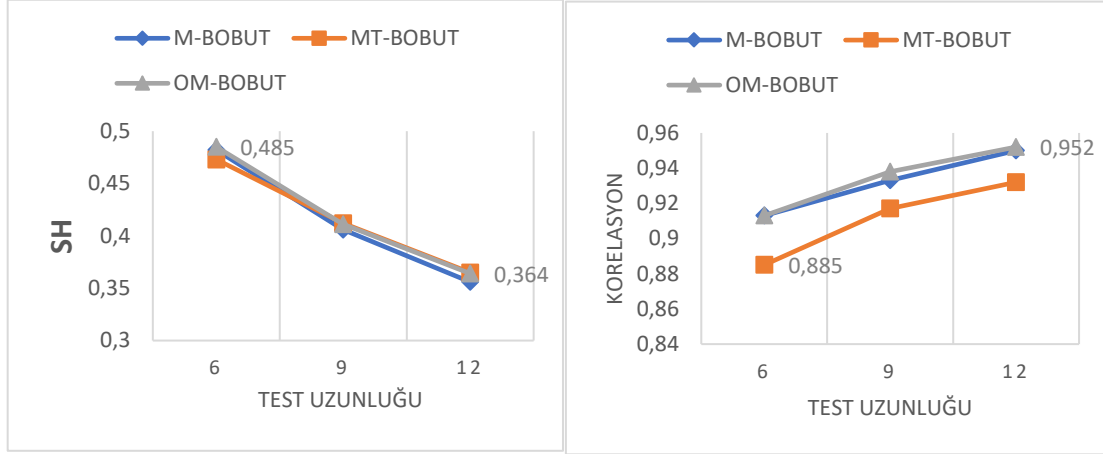
MAB değerleri incelendiğinde, madde takımı sayısı 6 ile madde takımı sayısı 9 arasındaki MAB değeri farkı BBT, MT-BBT ve OM-BBT için sırasıyla 0,054, 0,066, ve 0,059 iken, madde takımı sayısı 9 ile madde takımı sayısı 12 arasındaki MAB değeri farkı sırasıyla 0,039, 0,036 ve 0,038'dir. Test uzunlukları 6 ile 9 arasındaki MAB farkı fazla iken, test uzunlukları 9 ile 12 arasındaki MAB farkının azaldığı görülmektedir.

Madde takımı sayısına göre RMSE ve MAB farkları incelendiğinde, madde takımı sayısına bağlı olarak bu hata değerleri orantılı bir şekilde artmamıştır. Genel olarak uzunluk arttıkça hata değerlerinin azalması beklenen bir durum iken testin belli bir uzunluktan fazla olması ölçme kesinliği açısından olumlu karşılanmamaktadır. Çünkü fazla uzun testler adayın dikkat ve motivasyonunu olumsuz etkileyebilmekte bu da geçerli ve güvenilir sonuçlar almayı engellemektedir. Bu nedenle optimal madde takımı sayısı belirlenirken hata değerlerinin yanında bu durumun da dikkate alınması gerekir. Şekil 7 incelendiğinde,

madde takımı sayısı 9 ve 12 olanların madde takımı sayısı 6'ya göre yüksek ölçme kesinliği gösterdiği sonucuna ulaşılmaktadır.

Şekil 8

Madde Takımı Sayısına Göre SH ve r Değerleri Grafiği



Şekil 8 incelendiğinde, madde takımı sayısı arttıkça SH değerleri azalırken gerçek ve kestirilen yetenekler arasındaki korelasyonun ise arttığı görülmektedir. SH değerleri üç farklı BBT yönteminde de 0,49 altında değerler alırken gerçek ve kestirilen yetenekler arasındaki korelasyon da 0,89'dan yüksek bulunmuştur. Her ne kadar SH değeri en yüksek MT- BBT desende elde edilse de bu üç desende de oldukça benzer sonuçlar elde edilmiştir. Korelasyon yani kestirilen ve gerçek yetenekler arasındaki uyum katsayısı da BBT ile OM-BBT desenlerinde neredeyse birbirlerinin aynısı iken MT-BBT desende daha düşük değerler elde edilmiştir. Murphy ve diğerleri (2010), MT-BBT ile BBT ölçme kesinliği açısından benzer performans gösterdiğini bulmuştur. Ayrıca MTTK ile MTK modellerinden elde edilen yetenekler arasındaki uyumun oldukça benzer olduğu başka çalışmalar da mevcuttur (Wainer ve diğerleri, 2007; Wainer & Thissen, 1996; Yen, 1993). Fakat Yao (2019) MT-BBT performansını diğer iki yöntemden daha iyi bulmuştur. Tabi burada madde takımlarının etkisinin büyüklüğü ve madde havuzunun model veri uyumunu sağlaması da oldukça önemli rol oynamıştır. Özet olarak van der Linden (1998) tarafından belirtildiği gibi, test maddelerinin sayısı $k \rightarrow \infty$ arttıkça yetenek kestirimlerinin dağılımı gerçek yetenek

seviyesinde birbirine yaklaşılmaktadır. Bu nedenle, farklı yöntemler madde takımı sayısı arttıkça benzer sonuçlara ulaşacaktır. Madde takımı sayısı arttıkça yetenek kestirimlerinin standart hatası, gerçek ve kestirilen yetenekler arasındaki korelasyon ve yanlılık neredeyse aynı çıkması alan yazındaki çalışmaların sonuçlarıyla da tutarlıdır.

İkinci Alt Araştırma Problemine İlişkin Bulgular ve Yorumlar

“Örneklem büyüklüğüne (200-500-1000) göre BBT, MT-BBT ve OM-BBT yöntemlerinin gerçek ve kestirilen yetenek düzeyler arasındaki korelasyon, RMSE, BIAS, MAB, SH değerleri nasıl değişmektedir?” Araştırmanın ikinci alt problemine cevap bulabilmek için üç farklı desene ve üç farklı örneklem büyüklüğüne göre RMSE, BIAS, MAB ve SH değerleri hesaplanmıştır. Elde edilen bu değerler Tablo 9’da verilmiştir.

Tablo 9

Örneklem Büyüklüğüne Göre RMSE, BIAS, MAB ve SH Değerleri

Örneklem Büyüklüğü	BBT				MT-BBT				OM-BBT			
	RMSE	BIAS	MAB	SH	RMSE	BIAS	MAB	SH	RMSE	BIAS	MAB	SH
200	0,428	-0,025	0,339	0,413	0,509	-0,031	0,405	0,423	0,433	-0,015	0,350	0,418
500	0,427	0,006	0,340	0,416	0,475	-0,006	0,380	0,422	0,412	0,016	0,339	0,424
1000	0,421	0,004	0,332	0,414	0,489	-0,034	0,388	0,417	0,428	-0,014	0,341	0,417

Tablo 9 incelendiğinde, örneklem büyüklüğü 200 için ortalama RMSE değeri sırasıyla en az BBT deseni için 0,428 iken, OM-BBT deseni için 0,433 ve MT-BBT deseni için 0,509’dur. Benzer şekilde ortalama MAB değeri en az sırasıyla BBT için 0,340 iken OM-BBT için 0,339 ve MT-BBT için 0,380’dir. Örneklem büyüklüğü 200 için BBT deseni ile OM-BBT deseninin ölçme kesinliği neredeyse aynı olup MT-BBT deseninden daha iyi ölçme kesinliğine sahip oldukları görülmektedir.

Örneklem büyüklüğü 500 için ortalama RMSE sırasıyla en az OM-BBT yöntemi için 0,412 iken, BBT yöntemi için 0,427 ve MT-BBT yöntemi için 0,475’tir. Benzer şekilde ortalama MAB değeri OM-BBT için 0,339 iken BBT için de 0,340 ve MT-BBT için 0,380’dir.

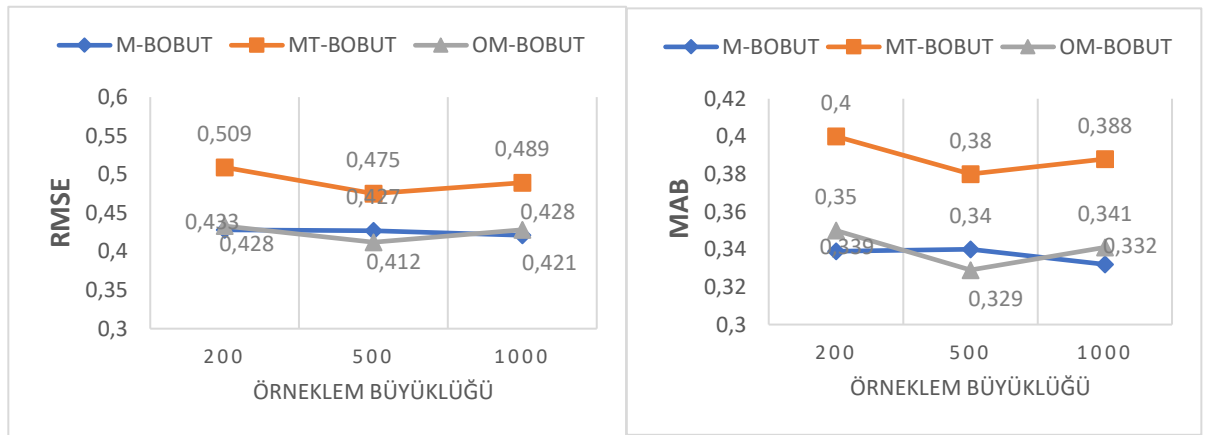
Örneklem büyüklüğü 500 için BBT deseni ile OM-BBT deseninin ölçme kesinliği neredeyse aynı olup MT-BBT deseninden daha iyi ölçme kesinliğine sahip oldukları görülmektedir. Örneklem büyüklüğü 200'e benzer şekilde örneklem büyüklüğü 500 için de BBT ve OM-BBT desenlerinden MT-BBT desenine göre daha etkili yetenek kestirim sonuçları elde edildiği görülmektedir.

Örneklem büyüklüğü 1000 için ortalama RMSE değeri sırasıyla en az BBT deseni için 0,421 iken, OM-BBT için 0,428 ve MT-BBT için 0,489'dur. Benzer şekilde ortalama MAB değeri BBT için 0,332 iken OM-BBT için de 0,341 ve MT-BBT için 0,388'dir. Örneklem büyüklüğü 1000 için BBT deseni ile OM-BBT deseninin ölçme kesinliği oldukça benzer olup MT-BBT'den daha iyi ölçme kesinliğine sahip oldukları görülmektedir. Diğer örneklem büyüklüklerine benzer şekilde, örneklem büyüklüğü 1000 için BBT ve OM-BBT desenlerinin MT-BBT yöntemine göre daha iyi yetenek kestirimi yaptığı söylenebilir.

Özet olarak her üç örneklem büyüklüğü (200, 500, 1000) düzeyinde de RMSE, MAB değerlerine göre BBT deseninin OM-BBT deseni ile benzer MT-BBT desenine göre daha etkili yetenek kestirimi sunduğu görülmektedir. Tablo 8'e göre RMSE ve MAB bulguları Şekil 9'da sunulmuştur.

Şekil 9

Örneklem Büyüklüğüne Göre RMSE ve MAB Değerleri

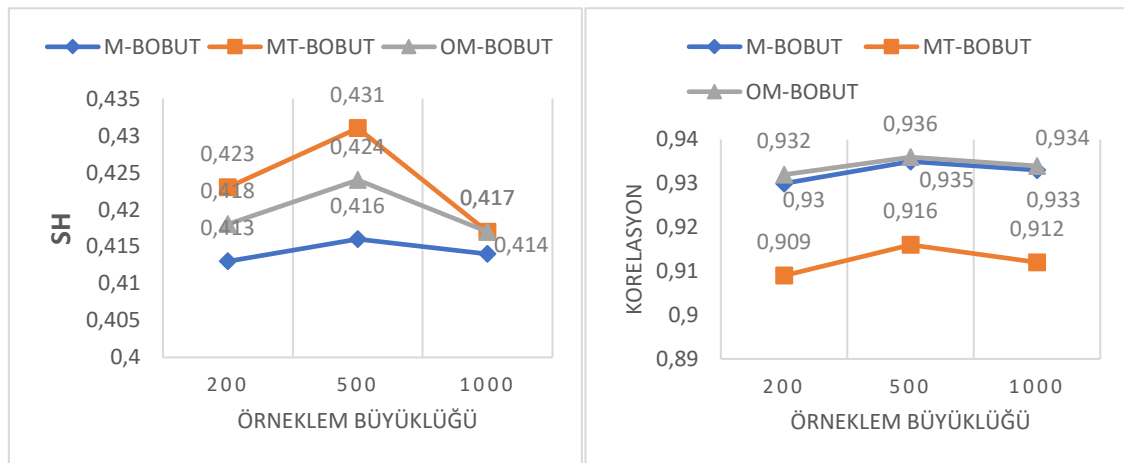


Şekil 9 incelendiğinde her örneklem büyüklüğü (200, 500, 1000) düzeyinde de BBT'nin OM-BBT ve MT-BBT desenlerinden daha düşük olduğu yani daha iyi ölçme kesinliği sunmaktadır. Örneklem büyüklüğü arttıkça RMSE değerlerinin genel olarak üç desende de azaldığı görülmektedir. BBT yöntemi için RMSE farkı 200-500, 500-1000 için sırasıyla 0,049 ve 0,062; MT-BBT deseni için 0,085 ve 0,042; OM-BBT için ise 0,069 ve 0,053'tür. Örneklem büyüklüğü 200 ile 500 arasındaki RMSE farkı, örneklem büyüklüğü 500 ile 1000 arasındaki RMSE farkından fazladır.

RMSE değerlerine benzer şekilde MAB değerleri incelendiğinde, örneklem büyüklüğü 200 ile 500 arasındaki MAB değeri farkı BBT, MT-BBT ve OM-BBT için sırasıyla 0,001, 0,025 ve 0,011 iken, örneklem büyüklüğü 500 ve 100 arasındaki MAB değeri farkı sırasıyla 0,001, 0,008 ve 0,008'dir. Tüm bu örneklem büyüklüklerinde MAB değerlerinin oldukça düşük olduğu dikkat çekmektedir. Özellikle örneklem büyüklüğü 500, örneklem büyüklüğü 1000'den çok az da olsa OM-BBT ve BBT yöntemlerinde daha düşük MAB değerleri vermiştir. Farklı örneklem büyüklüklerine göre elde edilen SH ve r değerleri Şekil 10'da sunulmuştur.

Şekil 10

Örneklem Büyüklüğüne Göre SH ve r Değerleri



Şekil 10 incelendiğinde, SH değerleri üç farklı BBT yöntemi için de oldukça benzerdir ve 0,413 ile 0,431 arasında değişmektedir. Yani onda birler basamağına yuvarlandığında tüm örneklem koşullarında bütün BBT yöntemlerinden elde edilen yetenek kestirimlerinin standart hatası birbirlerine eşit olmaktadır. Gerçek ve kestirilen yetenekler arasındaki korelasyonun ise BBT ile OM-BBT yöntemlerinde neredeyse aynı MT-BBT yönteminde ise en fazla 0,23'lük ($n=200$) bir azalış olduğu görülmektedir. Örneklem büyüklüğü arttıkça oluşan bu farklılık oldukça azalmıştır. Fakat yine de minimum hatanın ve maksimum uyumun 500 örneklem büyüklüğünde olduğu görülmektedir. Yani $n=500$ büyüklüğünde bir örneklem de ideal örneklem büyüklüğü olarak kabul edilebilir (Kim ve diğerleri, 2010). Özdemir (2015) çok boyutlu verilerle çalıştığından pratiklik ve kolaylık sağlaması adına veri setini 500 ile sınırlandırmıştır. Bu çalışma sonucuna göre de özellikle madde takımları gibi çok zaman alan BBT uygulamalarında örneklem büyüklüğü 500 ile sınırlandırılabilir. Bu çalışma için en basit model olan Rasch model temel alındığından küçük örneklemelerde de parametre tahminlerinde güvenilir sonuçlar elde edilmiştir. Fakat diğer MTK modelleri ile de çalışıldığında 500 örneklem sayısı gayet ideal olacaktır. Örneklem büyüklüğü arttıkça, ölçme ve sınıflama doğruluğunun arttığı bilinse de (Rudner & Guo, 2011) BBT uygulamalarında 150 bireyden oluşan bir kalibrasyon örneğiyle de doğru yetenek parametre kestirimlerinin yapılabilirdi bilinmektedir (Şahin & Weiss, 2015).

Üçüncü Alt Araştırma Problemine İlişkin Bulgular

“Yetenek kestirim yöntemine (EAP-MAP) göre BBT, MT-BBT ve OM-BBT yöntemlerinin gerçek ve kestirilen yetenek düzeyler arasındaki korelasyon, RMSE, BIAS, MAB, SH değerleri nasıl değişmektedir?” Araştırmanın üçüncü alt problemine cevap bulabilmek için üç farklı desene ve iki farklı yetenek kestirim yöntemine göre RMSE, BIAS, MAB ve SH değerleri hesaplanmıştır. Elde edilen bu değerler Tablo 10'da verilmiştir.

Tablo 10*Yetenek Kestirim Yöntemine Göre RMSE, BIAS, MAB ve SH Değerleri*

Kestirim Yöntemi	BBT				MT-BBT				OM-BBT			
	RMSE	BIAS	MAB	SH	RMSE	BIAS	MAB	SH	RMSE	BIAS	MAB	SH
EAP	0,425	-0,007	0,336	0,414	0,502	-0,031	0,398	0,430	0,427	-0,002	0,339	0,418
MAP	0,426	-0,002	0,337	0,415	0,480	-0,018	0,380	0,414	0,422	-0,006	0,341	0,421

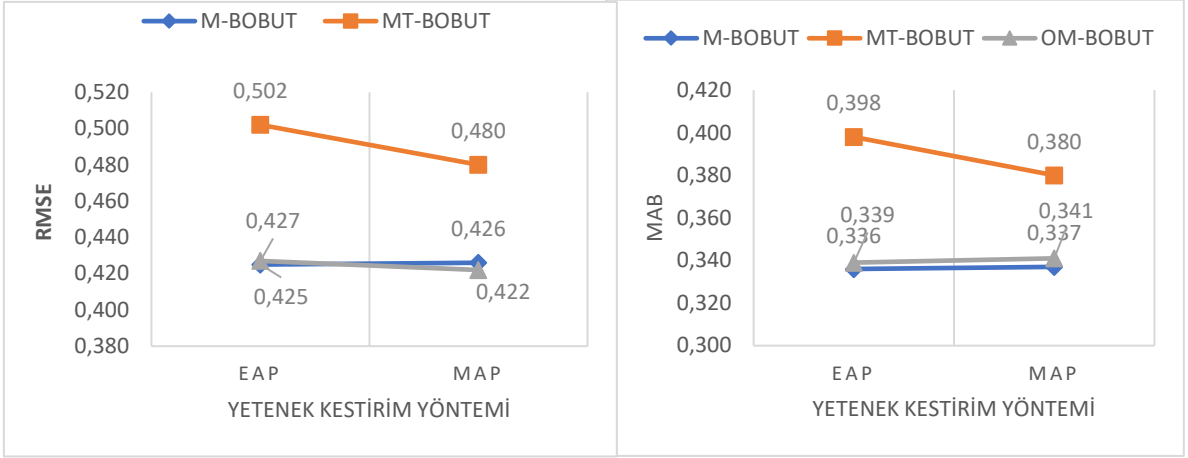
Tablo 10 incelendiğinde EAP yetenek kestirimi için ortalama RMSE değeri sırasıyla en az BBT deseni için 0,425 iken, OM-BBT yöntemi için 0,427 ve MT-BBT deseni için 0,502'dir. Benzer şekilde ortalama MAB değeri BBT için 0,336 iken OM-BBT için 0,339 ve MT-BBT için 0,398'dir. EAP için BBT ile OM-BBT desenlerinin ölçme kesinliği neredeyse aynı olup MT-BBT'den daha iyi ölçme kesinliğine sahip oldukları görülmektedir.

MAP yetenek kestirimi için ortalama RMSE değeri sırasıyla en az OM-BBT deseni için 0,422, BBT deseni için 0,426 ve MT-BBT deseni için 0,480'dir. Ortalama MAB değeri BBT için 0,337 iken OM-BBT için 0,341 ve MT-BBT için 0,380'dir. MAP için BBT deseni ile OM-BBT deseninin ölçme kesinliği neredeyse aynı olup MT-BBT deseninden daha iyi ölçme kesinliğine sahip oldukları görülmektedir.

Özet olarak her iki yetenek kestirim (EAP, MAP) yönteminde de RMSE, MAB değerleri açısından BBT deseninin OM-BBT deseni ile benzer MT-BBT desenine göre daha etkili yetenek kestirimi sunduğu görülmektedir. Farklı yetenek kestirim yöntemlerine göre elde edilen RMSE ve MAB değerleri Şekil 11'de sunulmuştur.

Şekil 11

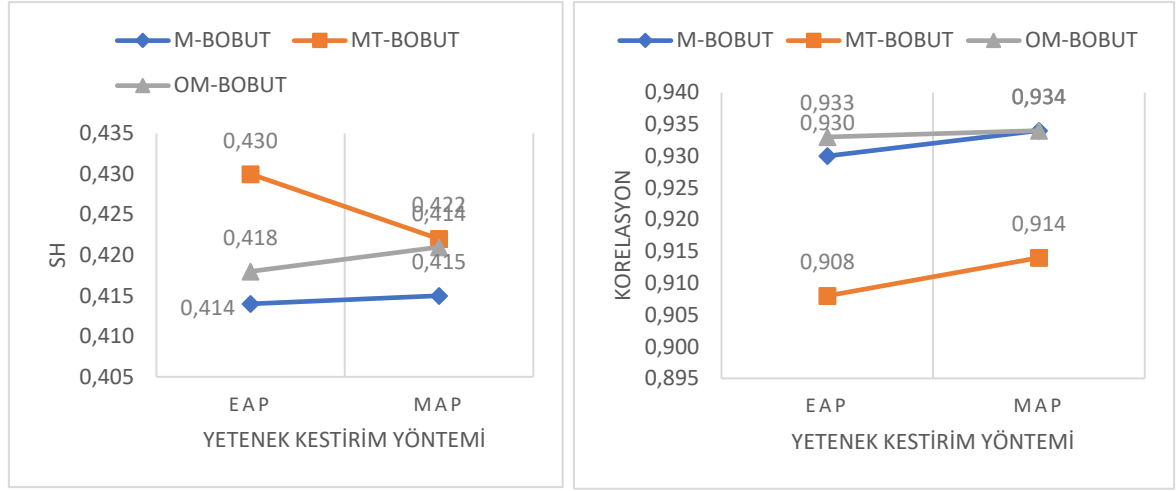
Yetenek Kestirim Yöntemine Göre RMSE ve MAB Değerleri



Şekil 11 incelendiğinde her iki kestirim yöntemi (EAP, MAP) BBT ve OM-BBT desenlerinin birbirinin aynısı olduğu ve MT-BBT yöntemlerinden daha düşük olduğu yani daha iyi ölçme kesinliği sunmaktadır. Üç farklı BBT deseninde de RMSE değerlerinin MAP yönteminde azaldığı görülmektedir. Bu fark MT-BBT deseninde daha çok belirgindir. BBT deseni için RMSE farkı 0,001; MT-BBT yöntemi için 0,02; OM-BBT için ise 0,003'tür. Alan yazında da MAP kestirim yönteminin madde seçme ve madde düzeyinde boyutluluk modellerinden minimum düzeyde etkilenmesinden dolayı kullanılması önerilmektedir (Özdemir, 2015). RMSE değerlerine MAB değerleri incelendiğinde, EAP-MAP yetenek kestirim yöntemleri arasındaki MAB değeri farkı BBT, MT-BBT ve OM-BBT için sırasıyla 0,001, 0,018, ve 0,002'dir. Farklı yetenek kestirim yöntemlerine göre elde edilen SH ve r değerleri Şekil 12'de sunulmuştur.

Şekil 12

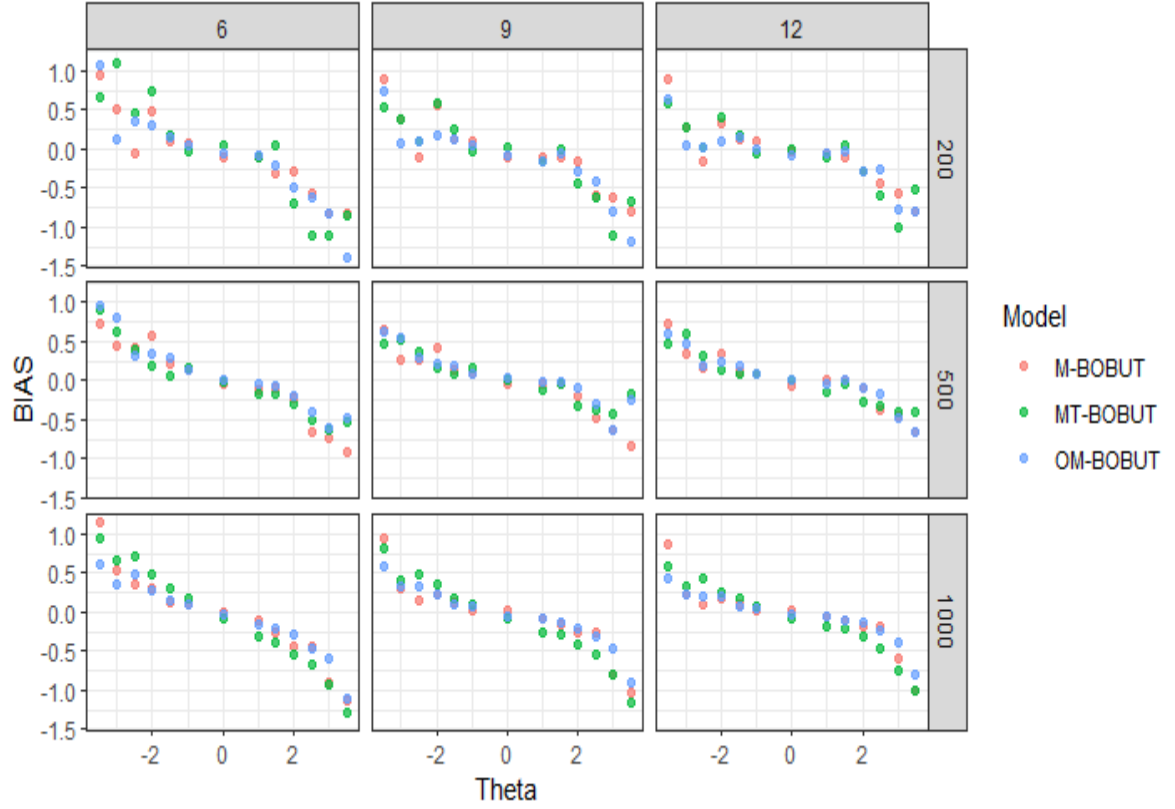
Yetenek Kestirim Yöntemine Göre SH ve r Değerleri



Şekil 12 incelendiğinde, farklı yetenek kestirim yöntemlerine göre SH değerleri üç farklı BBT deseni için de oldukça benzerdir ve 0,414 ile 0,430 arasında değişmektedir. Yani onda birler basamağına yuvarlandığında yetenek kestirimi koşullarında bütün BBT desenlerinden elde edilen yetenek kestirimlerinin standart hatası birbirlerine eşittir. Gerçek ve kestirilen yetenekler arasındaki korelasyonun ise BBT ile OM-BBT desenlerinde neredeyse aynı MT-BBT deseninde ise en fazla 0,026'lık bir azalış olduğu görülmektedir. MAP yetenek kestiriminde üç farklı BBT için de gerçek ve kestirilen yetenekler birbirlerine daha yakın çıkmıştır. MT-BBT deseninde MAP kestirim yöntemine göre SH değeri EAP kestirim yöntemine göre daha fazla azalış göstermiştir. Diao ve Reckase (2009) yapmış oldukları çalışmada MAP yönteminin daha güvenilir ve tutarlı verdiği sonucunu bulmuştur. Ayrıca, Bayes MAP yetenek kestiriminin MLE yetenek kestirim yöntemlerine göre gerçek θ değerine daha çabuk yakınsadığı belirtilmiştir (Özdemir, 2015).

Şekil 13

Madde Takımı Sayısı ve Örneklem Büyüklüklerine Göre Yetenek Ölçeği Boyunca Farklı BBT Desenlerinden Elde edilen BIAS Değerleri-MAP



Şekil 13 incelendiğinde, üç farklı desene göre madde takımı sayısı arttıkça yetenek ölçeği boyunca yeteneğin daha az hata ile kestirildiği görülmektedir. Örneklem büyüklüğü 500'e çıktığında BIAS değerleri düşerken 500 ve 1000 iken fark gözlenmemiştir. Aslında üç farklı desende de BIAS değerleri oldukça düşük ve birbirlerine benzerdir. Yani BBT ve OM-BBT MT-BBT desenlerinin iyi ölçme kesinliğine ulaştığı görülmektedir. Yetenek ölçeğinin uç noktalarına doğru gidildikçe ise BBT desenlerinin ölçme kesinliği arasındaki farkın BBT lehine arttığı görülmektedir. Özellikle MT-BBT deseni kullanıldığında örneklem büyüklüğü 500 civarında tutulmalıdır. Örneklem büyüklüğü farklı olduğundan bu grafiğin verilme amacı, örneklemeler arasında karşılaştırmadan ziyade betimleyici bir özellik taşıdığı söylenebilir.

Araştırma Probleminin Madde Havuzu Kullanımına İlişkin Bulgular ve Yorumlar

“BBT, OM-BBT ve MT-BBT desenlerinin madde kullanım sıklığı, kullanılmayan madde ve test çakışma oranları nasıl değişmektedir?” araştırma sorusu incelenmiştir. Maddelerin ortalama kullanım sıklığı oranları, kullanılmayan madde oranları ve son olarak test çakışma oranları her bir koşul altında değerlendirilerek Tablo 11’de verilmiştir.

Tablo 11 incelendiğinde, ortalama madde kullanım sıklığı oranı tüm desenlerde sabit madde takımı sayısı kullanıldığı için aynı olduğu görülmektedir. Fakat madde havuzunun etkin kullanımı için kullanılmayan madde oranları incelendiğinde; tüm koşullarda BBT deseninde 0,012 ile 0,117 aralığında, MT-BBT deseninde 0,376 ile 0,610 aralığında, OM-BBT deseninde 0,285 ile 0,610 aralığında değiştiği görülmektedir. Dolayısıyla BBT deseninde diğer desenlere göre madde havuzunun oldukça etkili ve efektif kullanıldığı sonucuna varılır. Çünkü madde takımı sayısının fazla olduğu koşullarda madde havuzundaki maddelerin neredeyse tamamına yakını kullanmaktadır. Fakat MT-BBT ve OM-BBT desenleri BBT desenine göre her ne kadar düşük madde güvenliğine sahip gibi görünse de aslında bu desenlerdeki madde havuzunu etkin bir şekilde kullanmaktadır. Çünkü madde takımı sayısının fazla olduğu koşullarda madde havuzundaki maddelerin yarıdan fazlasını kullanmaktadır ve bu sonuç da madde güvenliği açısından oldukça önemlidir. Madde havuzundaki maddelerin kullanma oranının azlığı aslında havuzdan daha fazla madde kullanıldığını ifade etmektedir. Bu durum ise hem madde kullanım sıklığını düşürmekte hem de madde ifşa oranını azaltması açısından madde güvenliğini oldukça olumlu etkilemektedir.

Benzer şekilde test çakışma oranı incelendiğinde; tüm koşullarda BBT deseninde 0,128 ile 0,235 aralığında, MT-BBT deseninde 0,245 ile 0,525 aralığında, OM-BBT deseninde 0,338 ile 0,421 aralığında değiştiği görülmektedir. MT-BBT ve OM-BBT test çakışma oranı değerleri oldukça benzerdir. Test çakışma oranının azalması bireylere uygulanan ortak madde sayısının az olduğunu gösterir ve bu durum da test güvenliğinin yüksek olmasına neden olur.

Tablo 11*Madde Kullanım Sıklığı, Kullanılmayan Madde ve Test Çakışma Oranları*

Yetenek Kestirimi	Örneklem Büyüküğü	Madde takımı sayısı	BBT			MT-BBT			OM-BBT		
			Madde Kullanım Sıklığı	Kullanılmayan Madde	Test Çakışma	Madde Kullanım Sıklığı	Kullanılmayan Madde	Test Çakışma	Madde Kullanım Sıklığı	Kullanılmayan Madde	Test Çakışma
EAP	200	6	0,078	0,113	0,133	0,078	0,61	0,376	0,078	0,584	0,363
		9	0,116	0,056	0,179	0,117	0,506	0,456	0,116	0,429	0,393
		12	0,156	0,048	0,221	0,156	0,441	0,518	0,156	0,298	0,418
	500	6	0,078	0,082	0,138	0,078	0,61	0,367	0,078	0,61	0,385
		9	0,116	0,034	0,189	0,117	0,519	0,441	0,116	0,403	0,405
		12	0,155	0,030	0,235	0,156	0,415	0,509	0,156	0,298	0,421
	1000	6	0,078	0,074	0,138	0,078	0,597	0,363	0,078	0,571	0,377
		9	0,117	0,035	0,182	0,117	0,506	0,449	0,116	0,403	0,41
		12	0,156	0,017	0,228	0,156	0,415	0,517	0,156	0,299	0,43
MAP	200	6	0,078	0,117	0,128	0,078	0,585	0,38	0,078	0,558	0,338
		9	0,116	0,061	0,173	0,117	0,519	0,458	0,116	0,429	0,387
		12	0,156	0,012	0,216	0,156	0,416	0,525	0,156	0,312	0,401
	500	6	0,078	0,039	0,136	0,078	0,584	0,366	0,078	0,519	0,336
		9	0,116	0,013	0,184	0,117	0,506	0,452	0,116	0,377	0,379
		12	0,156	0,013	0,233	0,156	0,415	0,52	0,156	0,286	0,403
	1000	6	0,078	0,013	0,138	0,078	0,571	0,245	0,078	0,532	0,338
		9	0,117	0,013	0,184	0,117	0,467	0,455	0,116	0,376	0,38
		12	0,156	0,013	0,230	0,156	0,376	0,526	0,156	0,285	0,408

Dördüncü Alt Araştırma Problemine İlişkin Bulgular ve Yorumlar

“Farklı test uzunluklarında BBT, MT-BBT ve OM-BBT desenlerinin madde kullanım sıklığı, kullanılmayan madde sayısı ve test çakışma oranları nasıl değişmektedir?” Araştırmanın dördüncü alt problemine cevap bulabilmek için üç ayrı desene ve üç ayrı madde takımı sayısına göre madde kullanım sıklığı, kullanılmayan madde ve test çakışma oranları hesaplanmıştır. Elde edilen bu değerler Tablo 12’de verilmiştir.

Tablo 12

Madde Takımı Sayısına Göre Madde Kullanım Sıklığı, Kullanılmayan Madde ve Test Çakışma Oranları

Madde Takımı Sayısı	BBT			MT-BBT			OM-BBT		
	Madde Kullanım Sıklığı	Kullanılmayan Madde	Test Çakışma	Madde Kullanım Sıklığı	Kullanılmayan Madde	Test Çakışma	Madde Kullanım Sıklığı	Kullanılmayan Madde	Test Çakışma
6	0,078	0,073	0,135	0,078	0,593	0,350	0,078	0,562	0,356
9	0,116	0,035	0,182	0,117	0,504	0,452	0,116	0,403	0,392
12	0,156	0,022	0,227	0,156	0,413	0,519	0,156	0,296	0,414

Tablo 12 incelendiğinde tüm BBT desenlerinde test sabit uzunluklu olduğu için ortalama madde kullanım sıklığı oranlarının aynı olduğu görülmektedir. Fakat madde takımı sayısı arttıkça bu oranın arttığı görülmektedir. Çünkü madde kullanım sıklığı testte kullanılan madde sayısının testi alan birey sayısına oranı olduğundan madde sayısı arttıkça maddenin kullanım oranı da artacaktır. Örneğin madde takımı sayısı 6 olduğunda ortalama madde kullanım sıklığı oranı 0,078 iken madde takımı sayısı 12 olduğunda bu oran 0,156’ya çıkmıştır. Bu artış çok az olduğundan madde güvenliği açısından iyi sonuçlar elde edildiği söylenebilir. Benzer şekilde madde havuzundaki kullanılmayan madde oranları incelendiğinde, en verimli madde havuzu kullanımının BBT deseni olduğu ve bu desende 6,9,12 madde takımı uzunlukları için sırasıyla 0,073; 0,035 ve 0,022 değerlerinin elde edildiği görülmektedir. Aslında bu oran en fazla madde takımı sayısının 6’dan 9’a çıkarıldığında değişmiştir. MT-BBT ve OM-BBT desenlerinde ise kullanılmayan madde

oranları oldukça benzer olmakla birlikte OM-BBT daha iyi sonuçlar vermiştir. Örneğin madde takımı sayısı 12 için MT-BBT için kullanılmayan madde oranı 0,413 iken OM-BBT için bu oran 0,296'ya düşmüştür. Madde takımlarının birlikte ele alındığını düşünürsek madde havuzunun yarıdan fazlasının kullanıldığı ve madde güvenliği açısından iyi sonuçların elde edildiği sonucuna varılabilir. Testte yer alan madde takımı sayısı arttıkça bireylere uygulanan ortak madde sayısı arttığından test çakışma oranlarının da arttığı gözlenmiştir (Deng ve diğerleri, 2010). Test çakışma oranları yine en düşük BBT desenlerinden elde edilmiştir. Test güvenliği açısından da en iyi desenin sırasıyla BBT, OM-BBT ve MT-BBT olduğu söylenebilir.

Beşinci Alt Araştırma Problemine İlişkin Bulgular ve Yorumlar

“Farklı örneklem büyüklüklerinde BBT, MT-BBT ve OM-BBT desenlerinin madde kullanım sıklığı, kullanılmayan madde sayısı ve test çakışma oranları nasıl değişmektedir?” Araştırmanın beşinci alt problemine cevap bulabilmek için üç ayrı desene ve üç ayrı örneklem büyüklüğüne göre madde kullanım sıklığı, kullanılmayan madde ve test çakışma oranları hesaplanmıştır. Elde edilen bu değerler Tablo 13'te verilmiştir.

Tablo 13

Örneklem Büyüklüğüne Göre Madde Kullanım Sıklığı, Kullanılmayan Madde ve Test Çakışma Oranları

Örneklem Büyüklüğü	BBT			MT-BBT			OM-BBT		
	Madde Kullanım Sıklığı	Kullanılmayan Madde	Test Çakışma	Madde Kullanım Sıklığı	Kullanılmayan Madde	Test Çakışma	Madde Kullanım Sıklığı	Kullanılmayan Madde	Test Çakışma
200	0,117	0,068	0,175	0,117	0,513	0,452	0,117	0,435	0,383
500	0,117	0,035	0,186	0,117	0,508	0,443	0,117	0,416	0,388
1000	0,117	0,028	0,183	0,117	0,489	0,426	0,117	0,411	0,391

Tablo 13 incelendiğinde tüm BBT desenlerinde test sabit uzunluklu olduğu için ortalama madde kullanım sıklığı oranlarının örneklem büyüklüğüne dayalı değişmediği görülmektedir. Aynı zamanda örneklem büyüklüğü arttıkça da bu oran sabit kalmaktadır. BBT deseninde kullanılmayan madde oranı tüm örneklem koşullarında (200-500-1000)

oldukça düşüktür. 200 örneklem büyüklüğünde kullanılmayan madde sayısı oranı 0,068 iken MT-BBT deseninde 0,513 OM-BBT deseninde ise 0,435'tir. Örneklem büyüklüğü 500 iken kullanılmayan madde oranı, BBT deseninde 0,035, MT-BBT deseninde 0,508 ve OM-BBT deseninde 0,416'dır. Örneklem büyüklüğü 1000 iken bu oranlar daha azalarak BBT için 0,028, MT-BBT için 0,489 ve OM-BBT için 0,411 elde edilmiştir. Genel olarak örneklem büyüklüğünün farklı BBT desenlerinde madde güvenliğini çok fazla etkilemediği sonucuna ulaşılmıştır.

Altıncı Alt Araştırma Problemine İlişkin Bulgular ve Yorumlar

“Yetenek kestirim yöntemine göre BBT, MT-BBT ve OM-BBT desenlerinin madde kullanım sıklığı, kullanılmayan madde sayısı ve test çakışma oranları nasıl değişmektedir?” Araştırmanın altıncı alt problemine cevap bulabilmek için üç farklı desene ve iki farklı yetenek kestirim yöntemine göre madde kullanım sıklığı, kullanılmayan madde ve test çakışma oranları hesaplanmıştır. Elde edilen bu değerler Tablo 14'te verilmiştir.

Tablo 14

Yetenek Kestirim Yöntemine Göre Madde Kullanım Sıklığı, Kullanılmayan Madde ve Test Çakışma Oranları

Yetenek Kestirimi	BBT			MT-BBT			OM-BBT		
	Madde Kullanım Sıklığı	Kullanılmayan Madde	Test Çakışma	Madde Kullanım Sıklığı	Kullanılmayan Madde	Test Çakışma	Madde Kullanım Sıklığı	Kullanılmayan Madde	Test Çakışma
EAP	0,117	0,054	0,183	0,117	0,513	0,444	0,117	0,433	0,400
MAP	0,117	0,033	0,180	0,117	0,493	0,436	0,117	0,408	0,374

Tablo 14 incelendiğinde, tüm BBT desenlerinde test sabit uzunluklu olduğu için ortalama madde kullanım sıklığı oranlarının yetenek kestirim yöntemine göre değişmediği görülmektedir. BBT deseninde kullanılmayan madde oranı EAP ve MAP yöntemlerinde oldukça düşüktür. EAP kestirim yönteminde kullanılmayan madde sayısı oranı BBT, MT-BBT ve OM-BBT desenlerinde sırasıyla 0,054, 0,513 ve 0,433'tür. Test çakışma oranı da benzer şekilde en düşük BBT deseninde 0,183 en yüksek MT-BBT deseninde 0,513 olarak bulunmuştur. OM-BBT deseninde ise test çakışma oranı 0,400 olarak hesaplanarak MT-

BBT değerine yakın bulunmuştur. MAP kestirim yönteminde ise kullanılmayan madde sayısı oranı BBT, MT-BBT ve OM-BBT desenlerinde sırasıyla 0,033, 0,495 ve 0,408'dir. Test çakışma oranı ise bu üç farklı BBT deseninde sırasıyla 0,180, 0,436 ve 0,374 bulunmuştur. Genel olarak yetenek kestirim yönteminin farklı BBT desenlerinde madde güvenliğini çok az etkilediği görülmektedir. MAP kestirim yönteminde çok az da olsa madde havuzu daha etkili kullanılmıştır. Bu fark kullanılmayan madde sayısı oranında en fazla 0,025 iken test çakışma oranında 0,026'dır. Alan yazın incelendiğinde de yetenek kestirim yönteminin BBT madde havuzunun etkili kullanımına etkisinin pek olmadığı çalışmalar mevcuttur (Ersan-Cinar, 2022; Seo & Weiss, 2015). Alan yazına göre sonsal dağılım simetrik olduğunda her iki yöntem de aynı tahmin değerini vermektedir (Reckase, 2010). Fakat çeşitli sebeplerden dolayı alan yazında en çok tercih edilen EAP ve MAP kestirim yöntemi olmuştur. Örneğin; iteratif olmayan ve hesaplanması kolay olan EAP kestirim yöntemi madde takımı tabanlı BBT uygulamalarında sıklıkla tercih edilmektedir (Boyd ve diğerleri, 2013; Frey ve diğerleri, 2016; Glas ve diğerleri, 2000; Keng, 2008; Murphy ve diğerleri, 2010). Fakat MAP yöntemi de MT-BBT uygulamalarında yüksek ölçme kesinliği sağlamasından ve madde sayısı çok olduğunda yanlışlık göstermemesinden dolayı son yıllarda tercih edilmektedir (Seo & Weiss, 2015; Yao, 2012).

Bölüm 5

Sonuç ve Öneriler

Bu bölümde çalışmanın bulgularından elde edilenler doğrultusunda sonuç ve öneriler başlıkları yer almaktadır.

Sonuç

Bu çalışmada madde takımlarından oluşan testlerin BBT uygulamaları farklı desenler oluşturularak incelenmiş ve sonuçlar karşılaştırılmıştır. Bu bağlamda öncelikle bilgisayar tabanlı olarak özel bir üniversitenin hazırlık sınıfında okuyan öğrencilere uygulanan ve madde takımlarından oluşan İngilizce seviye belirleme sınavından elde edilen gerçek veri seti kullanılmıştır. Madde takımları öğrencilerin okuduğunu anlama becerilerini ölçmek amacıyla ortak metinler içerdiğinden test, dokuz farklı set şeklinde öğrencilere uygulanmıştır. Her bir set ölçekleme yapmak amacıyla dokuz ortak madde içerdiğinden eş zamanlı kalibrasyon yöntemiyle madde havuzu oluşturulmuştur. Madde takımlarının bağımsız birer madde olarak ele alındığı BBT, madde takımlarının birlikte ve madde takımı tepki modelinin temel alındığı MT-BBT ve son olarak madde takımlarının yani okuma metinlerinin yine birlikte ele alındığı fakat tek boyutlu MTK modeli temelli OM-BBT olmak üzere üç farklı BBT deseni oluşturulmuştur. Bu test desenleri gerçek veri setinden oluşan madde havuzu kullanılarak madde takımı sayısı, örneklem büyüklüğü ve yetenek kestirim yönteminin değiştirildiği koşullarda ölçme kesinliği ve doğruluğu, madde havuzu kullanımı açısından karşılaştırılmıştır. Araştırma sonucunda aşağıdaki sonuçlara ulaşılmıştır:

- 1) Araştırma kapsamında incelenen tüm koşullarda BBT ve OM-BBT desenlerinde MT-BBT desenine göre daha yüksek ölçme kesinliği ve doğruluğu elde edilmiştir. Fakat genel olarak bu üç desenin de yetenek kestirimi açısından benzer etkili kestirimler yapabildiği sonucuna ulaşılmıştır.
- 2) Bu araştırmada yer alan veri setindeki çoğu madde takımının yerel bağımlılık derecesi düşük çıktığından MT-BBT deseninden beklenen performans elde

edilememiştir. MTTK modelinin karmaşık olması ve analiz sürecinin oldukça zaman almasından dolayı düşük veya orta madde takımı etkisine sahip maddelerin çoğunlukta olduğu durumlarda bu desen yerine OM-BBT desenin seçilmesi uygun olacaktır. Çünkü OM-BBT deseni sonuçları genel olarak BBT deseniyle oldukça benzer çıkmıştır.

- 3) BBT deseni her bir maddeye göre uyarlama yaptığından özellikle madde ve test güvenliği sonuçlarının diğer iki desene göre daha iyi çıkması beklenen bir durumdur. Fakat madde takımında yer alan bir madde kökünün madde takımındaki her bir maddede ayrı olarak ele alınması geçerlik sorununa neden olabileceğinden dolayı BBT desenin bir ölçüt olarak değerlendirilip bunun yerine OM-BBT deseni tercih edilmelidir. Ayrıca OM-BBT deseninde kapsam dengelemesi açısından da daha fazla avantaj sağlamaktadır.
- 4) BBT, MT-BBT ve OM-BBT desenlerinin karşılaştırılmasında ele alınan madde takımı sayısı koşuluna göre 6, 9 ve 12 madde takımından oluşan test uzunlukları karşılaştırılmıştır. Madde takımı sayısı arttıkça ölçme kesinliğinin arttığı yani kestirilen ve gerçek yetenek kestirimi değerlerinin oldukça benzer olduğu sonucuna ulaşılmıştır. Fakat bu artış madde takımı sayısının 9'dan 12'ye çıkarılmasının 6'dan 9'a çıkarılmasına göre görece daha azdır.
- 5) Örneklem büyüklüğünün 200, 500 ve 1000 olarak ele alındığı koşullarda üç farklı desende de 200 örneklem büyüklüğünün ölçme kesinliği 500 ve 1000 örneklem büyüklüklerine göre daha düşük olduğu bulunmuştur. 500 ve 1000 örneklem büyüklükleri oldukça benzer olsa da $n=500$ büyüklüğünde bir örneklem de ideal örneklem büyüklüğü olarak kabul edilebilir. Sonuç olarak örneklem büyüklüğünün BBT ölçme kesinliğine önemli bir etkisi olmadığı sonucuna ulaşılmıştır.
- 6) EAP ve MAP yetenek kestirim yöntemlerine göre BBT desenin OM-BBT desenlerinin benzer ve bu desenlerin MT-BBT desenine göre daha etkili yetenek kestirimi sunduğu görülmüştür. Ayrıca BBT ve OM-BBT desenlerinde farklı yetenek kestirim yöntemlerinde de benzer ölçme kesinliği sağlandığı bulgusu elde edilmiş

fakat MT-BBT yönteminde MAP kestirim yönteminin daha etkili yetenek kestiriminde bulunduğu sonucuna ulaşılmıştır.

- 7) Bu çalışmada incelenen tüm koşullarda BBT deseninin diğer desenlerden daha az kullanılmayan madde sayısı ve test çakışma oranlarına sahip olduğu bulunmuştur. Yani madde ve test güvenliği açısından en iyi desen BBT'dir. Fakat tüm BBT desenlerinde test sabit uzunluklu olduğu için ortalama madde kullanım sıklığı oranları aynıdır. Kullanılmayan madde sayısının daha az olması BBT deseninin diğer desenlere göre madde havuzunu daha etkili kullandığını göstermektedir. Benzer şekilde madde kullanım sıklığı ve test çakışma oranlarının düşük olması, maddelerin ifşa/ortaya çıkma oranlarını azaltması nedeniyle test güvenliği açısından daha olumlu olarak yorumlanabilir.
- 8) BBT desenlerinde madde takımı sayısı arttıkça ortalama madde kullanım sıklığı oranında artış gözlenmiştir. Fakat yine de madde güvenliği açısından iyi sonuçlar elde edilmiştir. Benzer şekilde madde havuzundaki kullanılmayan madde oranları incelendiğinde en verimli madde havuzu kullanımının BBT deseni olduğu, MT-BBT ve OM-BBT desenlerinde ise kullanılmayan madde oranları oldukça benzer olmakla birlikte OM-BBT daha iyi sonuçlar verdiği bulunmuştur. Madde takımı sayısı arttıkça bireylerin aldığı ortak madde sayısı artmış böylece test çakışma oranlarının da artmasına neden olmuştur.
- 9) Tüm BBT desenlerinde test sabit uzunluklu olduğu için ortalama madde kullanım sıklığı oranlarının örneklem büyüklüğüne dayalı değişmediği gözlenmiştir. BBT deseninde kullanılmayan madde oranı tüm örneklem koşullarında (200-500-1000) oldukça düşüktür. Genel olarak örneklem büyüklüğünün farklı BBT desenlerinde test çakışma oranını çok fazla etkilemediği sonucuna ulaşılmıştır.
- 10) BBT deseninde kullanılmayan madde oranı EAP ve MAP yöntemlerinde oldukça düşüktür. Genel olarak yetenek kestirim yönteminin farklı BBT desenlerinde madde

havuzu kullanımını çok az etkilediği görülmektedir. Fakat madde havuzunun etkili kullanımı açısından MAP yöntemi çok az da olsa daha iyi sonuçlar üretmiştir.

Öneriler

Bu bölümde elde edilen sonuçlara dayalı olarak önerilere yer verilmiştir. Öneriler başlığı, uygulayıcılara ve araştırmacılara yönelik öneriler olmak üzere iki bölümden oluşmaktadır.

Uygulayıcılara Yönelik Öneriler

- 1) Bu araştırmadan elde edilen bulgular doğrultusunda madde takımlarından oluşan BBT uygulamalarında uygun MTK modeli temel alınarak OM-BBT veya MT-BBT desenlerinin kullanılması önerilmektedir.
- 2) Bu araştırmada elde edilen sonuçlar incelendiğinde, OM-BBT ve MT-BBT desenleri benzer ölçme kesinliği ve madde havuzu kullanım değerlerine sahip olduğundan madde takımı etki derecesine göre bu desenlerden biri seçilebilir. Örneğin düşük ya da orta düzeyde madde takımı etkisine sahip veri setlerinde OM-BBT deseni kullanılabilirlik açısından tercih edilebilir. Çünkü MT-BBT deseni madde takımı MTK modelini temel aldığından analiz süreci oldukça vakit alabilmektedir. Zaman ve emek açısından kullanışlı olması sebebiyle OM-BBT deseni tercih edilebilir.
- 3) Bu araştırmada çoğu madde takımının yerel bağımlılık derecesi düşük çıktığından MTTK modelinden beklenen performans elde edilememiştir. Bu nedenle sadece yüksek derecede yerel bağımlılığa sahip madde takımlarında MTTK yerel bağımlılık derecesi dikkate alınıp diğer madde takımlarının yerel bağımsız ($Y=0$) olduğu varsayılarak gerekli modeller oluşturulmalıdır.
- 4) Daha etkili yetenek kestirimleri sunmasından dolayı madde takımı sayısının minimum 9, yani 27 madde olarak tercih edilmesi önerilmektedir. Testteki madde takımı sayısı arttıkça BBT, OM-BBT ve MT-BBT desenlerinin ölçme kesinliği de

birbirlerine oldukça yaklaşılmaktadır. Bu değişim en fazla kısa uzunlukta madde takımı sayısından orta uzunlukta madde takımı sayısına geçişte gözlenmektedir.

- 5) Bu çalışmada belirlenen örneklem büyüklüğü koşullarından elde edilen sonuçlar göz önüne alındığında, BBT uygulamalarının MTK temelli olmasından da kaynaklı minimum 500 örneklem ile çalışılması daha uygun olacaktır.

Araştırmacılara Yönelik Öneriler

- 1) Bu araştırmada ortak metinlere bağlı madde takımlarından oluşan bir İngilizce seviye belirleme testi kullanılmıştır. Bu nedenle MTK veya MTTK modellerini temel alan BBT, MT-BBT ve OM-BBT desenleri incelenmiştir. Gelecekte yapılacak çalışmalarda madde takımlarının çok kategorili olarak puanlandığı çok kategorili MTK modelleri de kullanılarak elde edilen sonuçlar karşılaştırılabilir.
- 2) BBT çalışmalarının etkili sonuçlar verebilmesi için madde havuzunun büyüklüğü oldukça önem arz etmektedir. Özellikle ortak bir metne dayalı madde takımlarından oluşan testlerin uzunluğu öğrencilerin yanıtlayabileceği sayıda olması gerektiğinden tek bir set ya da form yerine birden fazla formlar oluşturularak ölçekleme yöntemlerine başvurulmaktadır. Bu çalışmada ise yaklaşık 30-36 maddeden oluşan ve ortak maddeler içeren dokuz farklı formun her biri yaklaşık 100 kişiden oluşan bir gruba uygulanarak ortak maddeye dayalı ölçekleme yapılmıştır. Her bir form az sayıda öğrenciye uygulanabildiğinden Rasch modele göre kalibrasyon yapılmıştır. Daha çok örnekleme ulaşılarak 2PL veya 3PL modelleri de kullanılarak karşılaştırmalar yapılabilir.
- 3) Bu çalışmada madde havuzu özel bir üniversitenin hazırlık sınıflarına uygulanarak madde ve birey parametreleri elde edilmiştir. Bu gerçek veriden elde edilen madde ve bireye özgü parametreler kullanılarak farklı madde takımı sayısı, örneklem büyüklüğü ve yetenek kestirim yöntemi koşulları belirlenerek BBT uygulamaları gerçekleştirilmiştir. Bu koşullara madde seçim yöntemi,

madde takımı sayısı, madde takımı etki derecesi gibi koşullar da eklenerek benzer çalışmalar yürütülebilir.

- 4) Bu çalışmada madde takımı tabanlı madde takımı modelini temel alan MT-BBT deseninde madde havuzundaki maddelerin kısıtlı sayıda olmasından dolayı sadece madde takımları arasında uyarlama yapılabilmiştir. Fakat daha sonraki çalışmalarda hem maddeler arası hem de madde takımları içinde uyarlamaya imkân verecek BBT deseni de eklenerek sonuçlar arası karşılaştırmalar yapılabilir.
- 5) Bu çalışmada incelenen BBT desenleri ölçme kesinliği, madde havuzu kullanımı açısından karşılaştırılmıştır. Fakat daha sonraki çalışmalarda iki kategorili ya da çok kategorili sınıflamaları temel alan sınıflama doğruluğunu karşılaştıran çalışmalara da yer verilebilir.
- 6) Araştırmada madde takımı sayısı 6, 9 ve 12 olarak belirlenmiştir. Farklı madde takımı sayısı kullanılarak benzer çalışmalar yapılabilir. Madde takımı tabanlı BBT uygulamaları için en uygun (optimal) madde takımı sayısının belirlenmesi verimliliği artıracağından test uzunluğu ile ilgili çalışmalar alan yazına katkı sağlayacaktır.
- 7) Örneklem büyüklüğü ise BBT uygulamalarında ölçme kesinliğini fazla etkilememekle birlikte BBT uygulamalarının MTK temelli olmasından dolayı minimum 500 örneklem ile çalışılması doğru yetenek kestirimlerinin yapılması açısından daha uygun olacaktır. Fakat yine de örneklem büyüklüğünün 1000'den büyük olduğu durumlar için analizler yapılmalıdır.
- 8) Bayes yöntemleri (EAP-MAP) BBT uygulamalarında etkili sonuçlar verdiklerinden daha sık tercih edilmektedir. Gelecek çalışmalarda MLE yöntemi de incelenerek Bayes yöntemlerle karşılaştırılması, gerçek zamanlı BBT desenleri oluşturulurken oldukça fayda sağlayacaktır.

- 9) Araştırmada ele alınan BBT desenlerinde içerik dengelemesi ile ilgili bir kısıtlamaya yer verilmemiştir. Gelecekteki çalışmalarda bu kısıtlamalara da yer verilerek madde takımlarından oluşan BBT desenlerine etkileri araştırılabilir.
- 10) Araştırmada okuma ve dinleme bölümlerinden oluşan YDSBS'nin sadece okuma bölümünden elde edilen veriler madde havuzuna dâhil edilmiştir. Gelecek araştırmalar ise birden fazla yeterli ölçen okuma ve dinleme gibi farklı bölümleri ele alarak çok boyutlu BBT desenlerini inceleyebilir.
- 11) Çalışmada ele alınan okuma metinleri çoğunlukla düşük derecede madde takımı etkisine sahip olduğundan yabancı dil testlerinde yer alan cloze-testler gibi yüksek derecede yerel bağımlılığa sahip olabilecek madde türleri seçilerek yeni çalışmalar yapılabilir.
- 12) Bu çalışmada madde takımlarının birlikte seçilmesini sağlayan MT-BBT ve OM-BBT desenleri oluşturulmuştur. Gelecek çalışmalarda ise mantığı bu oluşturulan BBT desenlerine benzeyen fakat panel, modül vb. yönler açısından ayrışan çok aşamalı test desenleri de oluşturularak farklı BBT yaklaşımlarının karşılaştırılması sağlanabilir.

Kaynaklar

- Ackerman, T. A. (1987). *The robustness of LOGIST and BILOG IRT estimation programs to violations of local independence*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.
- Alkan, D. (2023). *İngilizce seviye belirleme sınav sonuçları üzerinde bilgisayarda bireyselleştirilmiş sınıflama testi yaklaşımının uygulanması*. [Yayımlanmamış Doktora Tezi]. Hacettepe Üniversitesi.
- Aybek, E. C., & Çıkrıkçı, R. N. (2018). Kendini değerlendirme envanteri'nin bilgisayar ortamında bireye uyarlanmış test olarak uygulanabilirliği. *Turkish Psychological Counseling and Guidance Journal*, 8(50), 117-141.
- Baldonado, A. A., Svetina, D., & Gorin, J. (2015). Using necessary information to identify item dependence in passage-based reading comprehension tests. *Applied Measurement in Education*, 28(3), 202–218. <https://doi.org/10.1080/08957347.2015.1042154>
- Bergstrom B. A., & Lunz, M. E. (1999). CAT for certification and licensure. In F. Drasgow & J. B. Olson-Bunchanan (Ed.), *Innovations in computerized assessment* (pp. 67-91). Lawrence Erlbaum Associates.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Ed.), *Statistical theories of mental test scores* (pp. 397–472). Addison- Wesley.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6(4), 431–444. <https://doi.org/10.1177/014662168200600405>
- Boyd, A. M. (2003). *Strategies for controlling testlet exposure rates in computerized adaptive testing systems* [Doctoral dissertation]. The University of Texas].

- Boyd, A. M., Dodd, B., & Fitzpatrick (2013). A comparison of exposure control procedures in CAT systems based on different measurement models for testlets. *Applied Measurement in Education*, 26(2), 113–135. <http://dx.doi.org/10.1080/08957347.2013.765434>
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlet. *Psychometrika* 64, 153–168.
- Braeken, J. A. (2011). Boundary mixture approach to violations of conditional independence. *Psychometrika* 76, 57–76 (2011). <https://doi.org/10.1007/s11336-010-9190-4>
- Chalhoub–Deville, M., & Deville, C. (1999). Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics*, 19, 273–299. <https://doi.org/10.1017/S0267190599190147>
- Chalmers, R. P. (2023). *mirt: Multidimensional item response theory*. R package version 1.41.
- Chang, H.-H. (2015). Psychometrics behind Computerized Adaptive Testing. *Psychometrika*, 80(1), 1–20. <https://doi.org/10.1007/s11336-014-9401-5>
- Chang, S. W., Ansley, T. N., & Lin, S. H. (2000). *Performance of item exposure control methods in computerized adaptive testing: Further explorations*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. <https://eric.ed.gov/?id=ED442837>
- Chang, H. H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20, 213–229.
- Chen, C.–T. & Wang, W.–C. (2007). Effects of ignoring item interaction on item parameter estimation and detection of interacting items. *Applied Psychological Measurement*, 31(5), 388–411. <https://doi.org/10.1177/0146621606297309>

- Chen, S.-Y., Ankenmann, R. D., & Chang, H.-H. (2000). A comparison of item selection rules at the early stages of computerized adaptive testing. *Applied Psychological Measurement*, 24(3), 241–255. <https://doi.org/10.1177/01466210022031705>
- Chen, S.-Y., Ankenmann, R. D., & Spray, J. A. (2003). The relationship between item exposure and test overlap in computerized adaptive testing. *Journal of Educational Measurement*, 40(2), 129–145. <https://doi.org/10.1111/j.1745-3984.2003.tb01100.x>
- Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265–289. <https://doi.org/10.2307/1165285>
- Chen, W.-H., Lenderking, W., Jin, Y. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Qual Life Res* 23, 485–493. <https://doi.org/10.1007/s11136-013-0487-5>
- Choi, S. W., & van der Linden, W. J. (2018). Ensuring content validity of patient-reported outcomes: a shadow-test approach to their adaptive measurement. *Quality of Life Research*, 27(7), 1683–1693. <https://doi.org/10.1007/s11136-017-1650-1>
- Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. Holt, Rinehart and Winston.
- Curi, M., & Silvia, V. (2019). Academic English proficiency assessment using a computerized adaptive test. *Tendencias Em Matematica Aplicada e Computacional*, 20(2), 381–401. <https://doi.org/10.5540/tema.2019.020.02.0381>
- Çoban, E. (2020). *Bilgisayar temelli bireyselleştirilmiş test yaklaşımlarının Türkiye'deki merkezi dil sınavlarında uygulanabilirliğinin araştırılması* [Yayımlanmamış Doktora Tezi]. Ankara Üniversitesi.
- Davey, T., & Nering, M. (2002). Controlling item exposure and maintaining item security. In C. Mills, M. T. Potenza, J. J. Fremer and W. C. Ward (Ed.), *Computer-Based*

Testing: Building the Foundation for Future Assessments. Lawrence Erlbaum Associates.

- Davis, L. L. (2004). Strategies for controlling item exposure in computerized adaptive testing with the generalized partial credit model. *Applied Psychological Measurement*, 28(3), 165-185. <https://doi.org/10.1177/0146621604264133>
- Davis, L. L. & Dodd, B. D. (2003). Item exposure constraints for testlets in the verbal reasoning section of the MCAT. *Applied Psychological Measurement*, 27(5), 335-356. <https://doi.org/10.1177/0146621603256804>
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. The Guilford.
- DeMars, C. (2010). *Item response theory*. Oxford University.
- DeMars, C. E. (2012). Confirming testlet effects. *Applied Psychological Measurement*, 36(2), 104-121. <https://doi.org/10.1177/0146621612437403>
- Deng, H., Ansley, T., & Chang, H. H. (2010). Stratified and maximum information item selection procedures in computer adaptive testing. *Journal of Educational Measurement*, 47(2), 202-226.
- Diao, Q., & Reckase, M. (2009). Comparison of ability estimation and item selection methods in multidimensional computerized adaptive testing. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*.
- Dodd, B. G., De Ayala, R. J., & Koch, W. R. (1995). Computerized adaptive testing with polytomous items. *Applied Psychological Measurement*, 19(1), 5-22. <https://doi.org/10.1177/014662169501900103>
- Dresher, A. R. (2002). *The examination of local item dependency of NAEP assessments using the testlet model* [Unpublished doctoral dissertation]. University of Pittsburgh.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates Publishers.

- Erdem Kara, B., & Doğan, N. (2022). The effect of ratio of items indicating differential item functioning on computer adaptive and multi-stage tests. *International Journal of Assessment Tools in Education*, 9(3), 682–696. <https://doi.org/10.21449/ijate.1105769>
- Ersan-Cinar, Ö. (2022). *The impact of local item dependence on computer adaptive testing given between and within testlet adaptivity* [Unpublished doctoral dissertation]. The University of Minnesota.
- Flaugher, R. (2000). Item pools. In H. Wainer (Eds.), *Computerized adaptive testing: A primer (2nd ed.)* (pp. 37–59). Lawrence Erlbaum Associates.
- Fliege, H., Becker, J., Walter, O. B., Bjorner, J. B., Klapp, B. F., & Rose, M. (2005). Development of a computer-adaptive test for depression (D-CAT). *Quality of life Research*, 14, 2277-2291. <https://doi.org/10.1007/s11136-005-6651-9>
- Frey, A., Seitz, N.–N., & Brandt, S. (2016). Testlet-based multidimensional adaptive testing. *Frontiers in Psychology*, 7, 1758. <https://doi.org/10.3389/fpsyg.2016.01758>
- Gierl, M.J., & Haladyna, T.M. (2012). *Automatic item generation: Theory and practice*. Routledge.
- Glas, C. A. W., Wainer, H., & Bradlow, E. T. (2000). MML and EAP estimation in testlet-based adaptive testing. In W. J van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (p. 271-287). Kluwer-Nijhoff.
- Hambleton, R. K., & Swaminathan, H. (1985). *Assumptions of item response theory*. In *Item Response Theory*. Springer, Dordrecht.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
- Hambleton, R. K., & Xing, D. (2006). Optimal and nonoptimal computer-based test designs for making pass-fail decisions. *Applied Measurement in Education*, 19(3), 221-239.

- Hanson, B. A., & Béguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, 26(1), 3–24. <https://doi.org/10.1177/0146621602026001001>
- 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/10.1177/014662169602000>
- He, L., & Min, S. (2024). Development and validation of a Computerized Adaptive EFL Test. *Zhejiang University Press*. <https://doi.org/10.1007/978-981-99-9987-3>
- Ho, T.-H. & Dodd, B. G. (2012). Item selection and ability estimation procedures for a mixed-format adaptive test. *Applied Measurement in Education*, 25(4), 305-326, <https://doi.org/10.1080/08957347.2012.714686>
- İnce-Aracı, G. (2022). *Psikolojik ölçeklerin çok boyutlu bobut olarak uyarlanması* [Adaptation of psychological scales as multidimensional cat] (Tez No. 741524) [Doktora Tezi, Gazi Üniversitesi]. YÖK. <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Jiao, H., Kamata, A., Wang, S., & Jin, Y. (2012). A multilevel testlet model for dual local dependence. *Journal of Educational Measurement*, 49(1), 82–100. <https://doi.org/10.1111/j.1745-3984.2011.00161.x>
- Jodoin, M. G., Zenisky, A., & Hambleton, R. K. (2006). Comparison of the psychometric properties of several computer-based test designs for credentialing exams with multiple purposes. *Applied Measurement in Education*, 19(3), 203–220. https://doi.org/10.1207/s15324818ame1903_3
- Kalender, İ. (2009). Başarı ve yetenek kestirimlerinde yeni bir yaklaşım: Bilgisayar ortamında bireyselleştirilmiş testler (Computerized adaptive tests-CAT). *CİTO Eğitim Kuram ve Uygulama*, 5, 39-48.

- Kang, H. A., Han, S., Betts, J., & Muntean, W. (2022). Computerized adaptive testing for testlet-based innovative items. *British Journal of Mathematical and Statistical Psychology*, 75(1), 136-157. <https://doi.org/10.1111/bmsp.12252>
- Katz, S., & Lautenschlager, G. (2001). The contribution of passage and no-passage factors to item performance on the SAT reading comprehension task. *Educational Assessment*, 7(2), 165–176.
- Kaya, E. (2022). *A comparability and classification analysis of computerized adaptive and conventional paper-based versions of an English language proficiency reading subtest [Bir İngilizce yeterlilik sınavı okuma alt testinin bilgisayar ortamında bireyselleştirilmiş test ve geleneksel kağıt kalem test uygulamalarının karşılaştırılabilirliği ve sınıflandırma analizi]* (Tez No. 709531) [Doktora Tezi, Bilkent Üniversitesi]. YÖK. <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Keller, L. A., Swaminathan, H., & Sireci, S. G. (2003). Evaluating scoring procedures for context-dependent item sets1. *Applied Measurement in Education*, 16(3), 207-222. https://doi.org/10.1207/S15324818AME1603_3
- Keng, L. (2008). *A Comparison of the performance of testlet-based computer adaptive tests and multistage tests* [Unpublished doctoral dissertation] The University of Texas.
- Kim, D., De Ayala, R. J., Ferdous, A. A., & Nering, M. L. (2007). *Assessing relative performance of local item dependence (LID) indexes*. Paper Presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.
- Kim, J. (2010). *A comparison of computer-based classification testing approaches using mixed-format tests with the generalized partial credit* [Unpublished doctoral dissertation]. The University of Minnesota.of Texas.
- Kim, J., Lee, W. C., Kim, D., & Kelley, K. (2009, April). *Investigation of vertical scaling using the rasch model*. Paper presented at the National Council on Measurement in Education, San Diego, CA.

- Kingston, N. M., & Dorans, N. J. (1985). The analysis of item-ability regressions: An exploratory IRT model fit tool. *Applied Psychological Measurement*, 9(3), 281–288. <https://doi.org/10.1177/014662168500900306>
- Kolen, M. J. (2004). Linking Assessments: Concept and History. *Applied Psychological Measurement*, 28(4), 219-226. <https://doi.org/10.1177/0146621604265030>
- Kolen, M. J., Hanson, B. A., & Brennan, R. L. (1992). Conditional standard errors of measurement for scale scores. *Journal of Educational Measurement*, 29, 285–307.
- Laurier, M. (1990, April). *What we can do with Computerized Adaptive Testing... and what we cannot do!* Paper presented at the Annual Meeting of the Regional Language Center Seminar.
- Lee, G., Kolen, M. J., Frisbie, D. A., & Ankenmann, R. D. (2001). Comparison of dichotomous and polytomous item response models in equating scores from tests composed of testlets. *Applied Psychological Measurement*, 25(4), 357–372. <https://doi.org/10.1177/01466210122032226>
- Li, Y., Bolt, D. M., & Fu, J. (2006). A Comparison of alternative models for testlets. *Applied Psychological Measurement*, 30(1), 3–21. <https://doi.org/10.1177/0146621605275414>
- Linacre, J. M. (2000). *Computer-Adaptive Testing: A methodology whose time has come* (No. 69, p. 58). MESA Memorandum.
https://www.cehd.umn.edu/EdPsych/C-Bas-R/Docs/Linacre2000_CAT.pdf
- Lord, F. M. (1971). A theoretical study of two-stage testing. *Psychometrika*, 36(3), 227–242. <https://doi.org/10.1007/BF02297844>
- Lord, F. M. (1980). *Applications of item response theory to practical problems*. Lawrence Erlbaum Associates.
- Lord, F.M., Novick, M.R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

- Luecht, R. M., & Sireci, S. G. (2011). A review of models for computer-based testing (Research Report RR-2011-12). *The College Board*.
- Luo, X., & Kim, D. (2018). A top-down approach to designing the computerized adaptive multistage test. *Journal of Educational Measurement*, 55(2), 243-263.
<https://doi.org/10.1111/jedm.12174>
- Magis, D., Yan, D., & Von Davier, A. A. (2017). *Computerized adaptive and multistage testing with R: Using packages catr and mstr*. Springer.
- Marais, I. D., & Andrich, D. (2008). Effects of varying magnitude and patterns of *local dependence in the unidimensional Rasch model*. *Journal of Applied Measurement*, 9(2), 105–124.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–173.
- Min, S., & He, L. (2014). Applying unidimensional and multidimensional item response theory models in testlet-based reading assessment. *Language Testing*, 31(4) 453–477. <https://doi.org/10.1177/0265532214527277>
- Murphy, D. L., Dodd, B. N., and Vaughn, B. K. (2010). A Comparison of item selection techniques for testlets. *Appl. Psychol. Measur.* 34, 424–437. <https://doi.org/doi:10.1177/0146621609349804>
- Nydyck, S. W., & Weiss, D. J. (2009). A hybrid simulation procedure for the development of CATs. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. www.psych.umn.edu/psylabs/CATCentral/
- Özdemir, B. (2015). *Madde düzeyinde boyutluluk modellerinin bilgisayar ortamında bireyselleştirilmiş test yöntemleri üzerindeki etkisinin incelenmesi. [Examining the effects of item level dimensionality models on multidimensional computerized adaptive testing methods]* (Tez No. 418187) [Doktora Tezi, Hacettepe Üniversitesi].
YÖK. <https://tez.yok.gov.tr/UlusalTezMerkezi/>

- Özdemir, B. & Gelbal, S. (2022). Ozdemir, B., Gelbal, S. Measuring language ability of students with compensatory multidimensional CAT: A post-hoc simulation study. *Educ Inf Technol* 27, 6273–6294. <https://doi.org/10.1007/s10639-021-10853-0>.
- Pak, S. (2017). *Ability parameter recovery of a computerized adaptive test based on rasch testlet models* [Unpublished doctoral dissertation]. University of Iowa.
- Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. Springer.
- Patsula, L. N. (1999). *Comparison of computerized adaptive testing and multi-stage testing*. (Publication No. 9950199) (Doctoral dissertation, University of Massachusetts-Amherst). ProQuest Dissertations & Theses Global.
- Penfield, R. D. (2006). Applying Bayesian item selection approaches to adaptive tests using polytomous items. 19(1), 1–20. *Applied Measurement in Education* https://doi.org/10.1207/s15324818ame1901_1.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press.
- Reckase, M. D. (2010). Designing item pools to optimize the functioning of a computerized adaptive test. *Psychological Test and Assessment Modeling*, 52(2), 127-141.
- Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(4), 311–327. <https://doi.org/10.1111/j.1745-3984.1998.tb00541.x>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological methods*, 21(2), 137.
- Rosenbaum, P. R. (1988). Item bundles. *Psychometrika*, 53(3), 349-360.

- Rudner, L. M., & Guo, F. (2011). Computer adaptive testing for small scale programs and instructional systems. *Journal of Applied Testing Technology*.
- Samejima, F. (1969). Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17
- Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61(2), 331-354. <https://doi.org/10.1007/BF02294343>
- Seo, D. G., & Weiss, D. J. (2015). Best design for multidimensional computerized adaptive testing with the bifactor model. *Educational and Psychological Measurement*, 75(6), 954-978. <https://doi.org/10.1177/0013164415575147>
- Shin, C. D., Chien, Y., Way, W. D., & Swanson, L. (2009). *Weighted penalty model for content balancing in CATs*. Pearson.
- Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28(3), 237-247. <http://www.jstor.org/stable/1434845>.
- Stocking, M. L. (1994). *Three practical issues for modern adaptive testing item pools* (ETS Research Report 94-5). Educational Testing Service.
- Şahin, M. D., & Gelbal, P. D. S. (2020). Development of a multidimensional computerized adaptive test based on the bifactor model. *International Journal of Assessment Tools in Education*, 323-342. <https://doi.org/10.21449/ijate.707199>
- Şenel, S. (2018). *Bilgisayar ortamında bireye uyarlanmış testlerin görme engelli öğrencilere uygunluğunun incelenmesi* [Investigation of the compatibility of computerized adaptive testing on students with visually impaired] (Tez No. 456700) [Doktora Tezi, Ankara Üniversitesi]. YÖK. <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Thissen, D., & Mislevy, R. J. (2000). Testing algorithms. In H. Wainer (Ed.), *Computerized adaptive testing: a primer* (pp. 101-133). Lawrence Erlbaum Associates.

- Thissen, D., Steinberg, L., & Mooney, J. A. (1989). Trace lines for testlets: A use of multiple-categorical-response models. *Journal of Educational Measurement*, 26(3), 247–260. <https://doi.org/10.1111/j.1745-3984.1989.tb00331.x>
- Thissen, D. & Wainer, H. (2001). *Test scoring*. Lawrence Erlbaum Associates.
- Thompson, N. A., & Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research, and Evaluation*, 16(1), 1-9. <http://pareonline.net/getvn.asp?v=16&n=1>
- Tseng, W.T. (2016). Measuring English vocabulary size via computerized adaptive testing. *Computers and Education*, 97, 69–85. <http://doi.org/10.1016/j.compedu.2016.02.018>
- Tuerlinckx, F., & De Boeck, P. (2001). The effect of ignoring item interactions on the estimated discrimination parameters in item response theory. *Psychological Methods*, 6(2), 181–195. <https://doi.org/10.1037/1082-989X.6.2.181>
- Vale, C. D. (1986). Linking item parameters onto a common scale. *Applied Psychological Measurement*, 10(4), 333-344. <https://doi.org/10.1177/014662168601000402>
- van der Linden, W. J. (2018). *Optimal test design*. Handbook of item response theory: Vol. 3. *Applications*, 167-195.
- van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. Springer.
- van der Linden, W. J. (1998). Optimal assembly of psychological and educational tests. *Applied Psychological Measurement*, 22(3), 195–211. <https://doi.org/10.1177/01466216980223001>
- van der Linden, W. J., & Pashley, P. J. (2010). Item selection and ability estimation in adaptive testing. In W. van der Linden & C. Glas (Ed.), *Elements of adaptive testing* (pp. 3– 30). Springer.

- van Rijn, P. W., Eggen, T. M., Hemker, B. T., & Sanders, P. F. (2002). Evaluation of selection procedures for computerized adaptive testing with polytomous items. *Applied Psychological Measurement*, 26, 393-411.
- Veldkamp, B. P., and van der Linden, W. J. (2002). Multidimensional adaptive testing with constraints on test content. *Psychometrika* 67, 575–588.
<https://doi.org/10.1007/BF02295132>
- Wainer, H. (1995). Precision and differential item functioning on a testlet-based test: The 1991 Law School Admissions Test as an example. *Applied Measurement in Education*, 8(2), 157–186. https://doi.org/10.1207/s15324818ame0802_4
- Wainer, H. (2000). *Computerized adaptive testing: A primer (2nd ed.)*. Lawrence Erlbaum Associates.
- Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. Cambridge University Press.
- Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., & Mislevy, R. J. (2000). *Computerized adaptive testing: A primer*. Routledge.
- Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24, 185–201.
- Wainer, H., & Lewis, C. (1990). Toward a psychometrics for testlets. *Journal of Educational Measurement*, 27(1), 1–14. <https://doi.org/10.1111/j.1745-3984.1990.tb00730.x>
- Wainer, H., & Thissen, D. (1996). How is reliability related to the quality of tests cores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice*, 15(1), 22–29.
- Wainer, H., & Wang, X. (2000). Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement*, 37(3), 203–220. <https://doi.org/10.1111/j.1745-3984.2000.tb01083.x>

- Wang, S., & Wang, T. (2001). Precision of Warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement*, 25(4), 317-331. <https://doi.org/10.1177/01466210122032163>
- Wang, T., Hanson, B. A., & Lau, C.-M. A. (1999). Reducing bias in CAT trait estimation: A comparison of approaches. *Applied Psychological Measurement*, 23(3), 263-278. <https://doi.org/10.1177/01466219922031383>
- Wang, X., Bradlow, E. T., & Wainer, H. (2002). A general Bayesian model for testlets: Theory and applications. *Applied Psychological Measurement*, 26(1), 109-128. <https://doi.org/10.1177/0146621602026001007>
- Wang, Z., Wang, C., & Weiss, D. J. (2022). Termination criteria for grid multiclassification adaptive testing with multidimensional polytomous items. *Applied Psychological Measurement*, 46(7), 551-570.
- Way, W. D. (1998). Protecting the integrity of computerized testing item pools. *Educational Measurement: Issues and Practice*, 17(4), 17-27. <https://doi.org/10.1111/j.1745-3992.1998.tb00632.x>
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive theory. *Applied Psychological Measurement*, 6(4), 473-492. <https://doi.org/10.1177/014662168200600408>
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361-375. <https://doi.org/10.1111/j.1745-3984.1984.tb01040.x>
- Weiss, D. J., & Sahin, A. (2024). *Computerized adaptive testing: From concept to implementation*. Guilford Publications.
- Yan, D., Von Davier, A. A., & Lewis, C. (2014). *Computerized multistage testing: Theory and applications*. Chapman and Hall/CRC.

- Yao, L. (2012). Multidimensional CAT item selection methods for domain scores and composite scores: Theory and applications. *Psychometrika* 77, 495–523. <https://doi.org/10.1007/s11336-012-9265-5>
- Yao, L. (2013). Comparing the performance of five multidimensional CAT selection procedures with different stopping rules. *Applied Psychological Measurement*, 37(1), 3–23. <https://doi.org/10.1177/0146621612455687>
- Yao, L. (2019). Item selection methods for computer adaptive testing with passages. *Frontiers in Psychology*, 10, 240. <https://doi.org/10.3389/fpsyg.2019.00240>
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30(3), 187–213. <https://doi.org/10.1111/j.1745-3984.1993.tb00423.x>
- Yılmaz-Koğar, E. & Kelecioğlu, H. (2017). Examination of different item response theory models on tests composed of testlets. *Journal of Education and Learning*, 6(4), <https://doi.org/113-126.10.5539/jel.v6n4p113>
- Zhang, O. (2010). *Polytomous IRT or testlet model: An evaluation of scoring models in small testlet size situations*. [Master's Dissertation, University of Florida] <http://ufdc.ufl.edu/UFE0042638/00001>
- Zheng, Y., & Chang, H. H. (2015). On-the-fly assembled multistage adaptive testing. *Applied Psychological Measurement*, 39(2), 104–118. <https://doi/10.1177/0146621614544519>

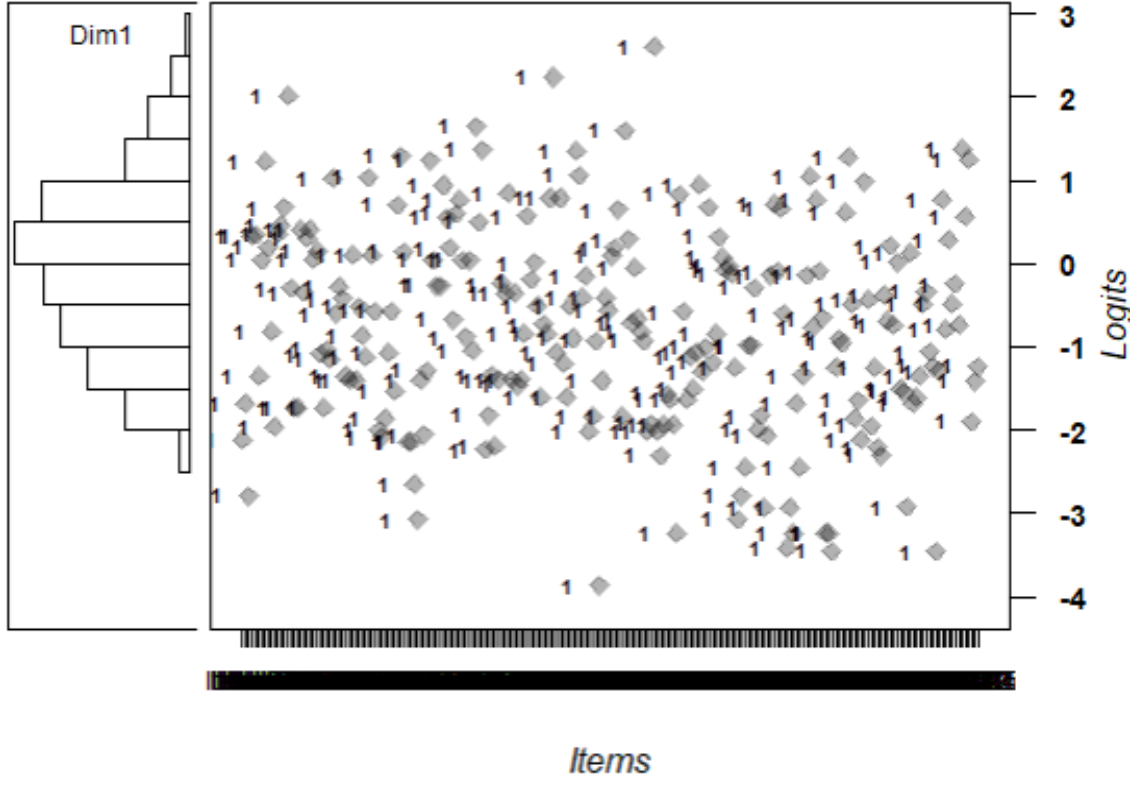
Ek-A: Her Bir Forma Göre Model Uyum İndeksleri

Form	Model	AIC	SABIC	HQ	BIC	logLik	χ^2	df
1	R-MTK	3286.157	3269.012	3318.842	3366.917	-1612.079	NA	NA
	R-MTTK	3278.848	3256.171	3322.076	3385.660	-1598.424	27.30*	10
2	R-MTK	3461.549	3445.302	3494.630	3543.225	-1699.774	NA	NA
	R-MTTK	3451.397	3429.909	3495.151	3559.421	-1684.699	30.15*	10
3	R-MTK	3768.483	3748.019	3807.494	3864.874	-1847.242	NA	NA
	R-MTTK	3772.335	3745.234	3823.999	3899.988	-1837.168	20.15	12
4	R-MTK	3854.205	3834.102	3893.376	3950.965	-1890.103	NA	NA
	R-MTTK	3834.129	3807.506	3886.004	3962.270	-1868.064	44.07*	12
5	R-MTK	3756.592	3735.763	3795.441	3852.611	-1841.296	NA	NA
	R-MTTK	3770.632	3743.048	3822.082	3897.793	-1836.316	9.96	12
6	R-MTK	3532.434	3514.289	3568.574	3621.683	-1732.217	NA	NA
	R-MTTK	3547.224	3523.210	3595.057	3665.348	-1728.612	7.21	11
7	R-MTK	3518.771	3499.379	3558.255	3616.256	-1722.385	NA	NA
	R-MTTK	3522.609	3496.928	3574.900	3651.711	-1712.305	20.16	12
8	R-MTK	3353.178	3335.034	3389.318	3442.428	-1642.589	NA	NA
	R-MTTK	3350.455	3326.440	3398.287	3468.579	-1630.228	24.72	11
9	R-MTK	3676.784	3659.607	3713.349	3767.019	-1804.392	NA	NA
	R-MTTK	3666.992	3644.257	3715.386	3786.420	-1788.496	31.79*	11

*R-MTK: Rasch Madde Tepki Kuramı; R-MTTK: Rasch Madde Takımı Tepki Kuramı

Ek-B: Birey-Madde Etkileşimi Haritası

Wright Map



EK-C: Etik Beyanı

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

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

beyan ederim.

28/06/2024

(İmza)

Sebahat GÖREN

EK-D: Etik Komisyon İzni

T.C.
HACETTEPE ÜNİVERSİTESİ REKTÖRLÜĞÜ
Eğitim Bilimleri Enstitüsü Müdürlüğü



Sayı : E-51944218-300-00002892884
Konu : Etik Komisyonu İzni (Sebahat GÖREN)

9.06.2023

EĞİTİM BİLİMLERİ ANA BİLİM DALI BAŞKANLIĞINA

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

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

Prof. Dr. Selahattin GELBAL
Enstitü Müdürü

Bu belge güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu: 41A9D9ED-3845-4BC3-AA77-BB9A6EF5C8D9

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

Adres: Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğü 06800

Beytepe-ANKARA

E-posta: ebe@hacettepe.edu.tr Elektronik Ağı: www.hacettepe.edu.tr

Telefon: (0 312) 297 85 70-71 Faks: (0 312) 299 85 66

Kep:

Bilgi için: Aysun ALTUN

Bilgisayar İşletmeni

Telefon: -





T.C.
KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ
REKTÖRLÜĞÜ
Sosyal ve Beşeri Bilimler Etik Kurulu

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

EĞİTİM BİLİMLERİ BÖLÜM BAŞKANLIĞINA

İlgi : 29.12.2020 tarihli ve 11930 sayılı yazı,

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

Prof. Dr. Nusret GÖKSU
Kurul Başkanı

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

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu :BEL95ZFLV Pin Kodu :78232

Belge Takip Adresi : https://ebys.ksu.edu.tr:443/enVision/Validate_Doc.aspx

Adres: Kahramanmaraş Sütçü İmam Üniversitesi Avcılar Kampüsü, 46100 -

Bilgi için: Mehmet TAKALAK

Onikişubat/Kahramanmaraş

Unvanı: Sekreter

Telefon:0344 300 27 01 Faks:0344 300 27 02

e-Posta: genelsekreterlik@ksu.edu.tr Elektronik Ağ: www.ksu.edu.tr

Kep Adresi: ksu.kahramanmaraş@hs01.kep.tr

Bu belge, güvenli Elektronik İmza ile imzalanmıştır.
Evrak sorgulaması (VALURL) adresinden yapılabilir.

T.C.

KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ
SOSYAL VE BEŞERİ BİLİMLER ETİK KURUL KARARLARI

Toplantı Tarihi: 04.01.2021 Toplantı Sayısı: 2020/1		
Üniversitemiz Sosyal ve Beşeri Bilimleri Etik Kurulu, Kurul Başkanı Prof. Dr. Nusret GÖKSU'nun Başkanlığında toplanarak aşağıdaki kararları almıştır, KARAR 1 : Dr. Öğr. Üyesi Levent YAKAR'ın 29.12.2020 Tarihli dilekçesi ve ekleri Prof. Dr. Nail YILDIRIM (raportör) 31.12.2020 tarih ve E. 49759 sayılı raporu görüşüldü. Üniversitemiz Öğretim Üyelerinden Dr. Öğr. Üyesi Levent YAKAR'ın "İngilizce Hazırlık Sınıflarında Bilgisayarda Bireyselleştirilmiş Test Kullanımı" başlıklı çalışması kapsamında kullanmak için uygulanacak açık uçlu görüşme formu (ankette yer alan soru, önerme ve/veya ölçeklerin alıntılanmasına yönelik akademik alıntılanma kurallarına uyma zorunluluğu araştırmayı yapan akademisyen ve varsa danışmanlığını yaptığı öğrenciye ait olmak koşuluyla) üniversite öğrencilerine uygulama talebinin uygun olduğuna oylama sonucunda oy birliği ile; Karar verildi		
BAŞKAN (İMZA) Prof. Dr. Nusret GÖKSU		
ÜYE (İMZA) Prof. Dr. Salih YEŞİL	ÜYE (İMZA) Prof. Dr. Ahmet NALÇACI	ÜYE (İMZA) Prof. Dr. Nail YILDIRIM
ÜYE (İMZA) Prof. Dr. Kemal TİMUR	ÜYE (İMZA) Prof. Dr. Faruk ÇİFTÇİ	

02.08.2022

İlgili Makama

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

Kahramanmaraş Sütçü İmam Üniversitesi

Doç. Dr. Levent YAKAR

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

28/06/2024

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

Tez Başlığı :Madde Takımlarından Oluşan Testlerde Bilgisayarda Bireyselleştirilmiş Test Uygulama

Desenlerinin Karşılaştırılması

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

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
27/06/2024	98	147413	06/06/2024	%5	2409349399

Uygulanan filtreler:

1. Kaynaklar hariç
2. Alıntılar dâhil
3. 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

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

Ad Soyadı: Sebahat GÖREN

Öğrenci No.: N18146172

Ana Bilim Dalı: Eğitim Bilimleri

İmza

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: ☐ Y.Lisans ☒ Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

Prof. Dr. Selahattin GELBAL

EK-F: Thesis/Dissertation Originality Report

28/06/2024

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

Thesis Title: A Comparison of Computerized Adaptive Testing Designs for Testlet Based 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
27/06/2024	98	147413	06/06/2024	5%	2409349399

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: Sebahat GÖREN

Student No.: N18146172

Department: Educational Sciences

Program: Educational Measurement and Evaluation

Status: ☐ Masters ☒ Ph.D. ☐ Integrated Ph.D.

Signature

ADVISOR APPROVAL

APPROVED
Prof. Selahattin GELBAL

EK-G: Yayımlama 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 ihlâl etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

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

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

28 /06 /2024

Sebahat GÖREN

"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 tez 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 tez 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.

