



**HACETTEPE ÜNİVERSİTESİ**  
**EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

Eğitim Bilimleri Ana Bilim Dalı  
Eğitimde Ölçme ve Değerlendirme Programı

AÇIKLAYICI MADDE TEPKİ MODELLERİNİN BİLGİSAYAR ORTAMINDA  
BİREYE UYARLANMIŞ TESTLERDE KULLANIMI



Osman TAT

Doktora Tezi

Ankara, 2020



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

AÇIKLAYICI MADDE TEPKİ MODELLERİNİN BİLGİSAYAR ORTAMINDA  
BİREYE UYARLANMIŞ TESTLERDE KULLANIMI

USE OF EXPLANATORY ITEM RESPONSE MODELS IN COMPUTER  
ADAPTIVE TESTS

Osman TAT

Doktora Tezi

Ankara, 2020

## Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,  
Osman TAT'ın hazırladıđı "Açıklayıcı Madde Tepki Modellerinin Bilgisayar Ortamında Bireye Uyarlanmış Testlerde Kullanımı" başlıklı bu çalıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölçme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiřtir.

J¼ri Bařkanı

Prof. Dr. řeref TAN



J¼ri Üyesi (Danıřman)

Prof. Dr. Nuri DOđAN



J¼ri Üyesi

Prof. Dr. Selahattin GELBAL



J¼ri Üyesi

Prof. Dr. H¼lya KELECİOđLU



J¼ri Üyesi

Doç. Dr. Celal Deha DOđAN



Bu tez Hacettepe Üniversitesi Lisans¼st¼ Eđitim, Öğretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından 03 / 12 / 2019 tarihinde uygun gör¼lm¼ř ve Enstit¼ Yönetim Kurulunca ..... / ..... / ..... tarihi itibarıyla kabul edilmiřtir.

Prof. Dr. Ali Ekber řAHİN  
Eđitim Bilimleri Enstitüsü M¼d¼r¼

## Öz

Bilgisayar ortamında bireye uyarlanmış testlerde madde havuzunun geliştirilmesi süreci oldukça zahmetli bir iştir. Bu geliştirme sürecinde madde parametrelerinin farklı alt gruplara, kullanım sırasına, farklı test formatlarına göre değişim göstermemesi beklenir. Bunun için yapılan analizlerde böylesi bir durumun tespit edildiği maddeler elenir. Bu çalışmanın amacı açıklayıcı madde tepki modellerinin bilgisayar ortamında bireye uyarlanmış testlerde kullanımının, madde havuzu geliştirme sürecine, uygulanan madde sayısına, yetenek parametrelerinin kestiriminin kesinliğine etkisini incelemektir. Araştırmada her birinde 100 maddenin olduğu 10 madde havuzu ve 1440 katılımcıya ait cevap örüntüleri simülasyonla üretilmiştir. Geliştirilen tüm madde havuzları Rasch model, örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modeli ile kalibre edilmiştir. Ardından her bir modelden elde edilen cevap örüntüleri ve önsel yetenek puanları ile 10 döngüden oluşan post-hoc simülasyonları gerçekleştirilmiş ve tüm havuzlardan elde edilen sonuçların ortalamasına ulaşılmıştır. Post-hoc simülasyonlarında yetenek puan kestirimi olarak BSD (EAP), sonlandırma kuralı olarak kesinlik ve minInfo kuralı, madde kullanım sıklığı kontrol kuralı olarak randomesque kullanılmıştır. Açıklayıcı madde tepki modeli ile yapılan post-hoc simülasyonu “catR” paketine araştırmacı tarafından yazılan yeni fonksiyonlar ile gerçekleştirilmiştir. Araştırmada tüm madde havuzunda olduğu gibi post-hoc simülasyonlarında da alt gruplar göz ardı edildiğinde tüm modellerin çok yakın yetenek puan ortalaması kestirdiği belirlenmiştir. Alt gruplar dikkate alındığında ise modeller arasında manidar farklar tespit edilmiştir. Ayrıca parametre kestirim modelinin maddenin kullanım sıklığı, test uzunluğu ve test örtüşme oranı üzerinde etkisinin olmadığı tespit edilmiştir. Çalışmada madde havuzunda birey ve madde eşdeğişkenlerinin etkin olduğu maddelerin bulunduğu durumlarda, bilgisayar ortamında bireye uyarlanmış testlerin açıklayıcı madde tepki modellerine dayalı parametrelerle yapılması önerilmiştir.

**Anahtar sözcükler:** bireye uyarlanmış test, açıklayıcı madde tepki modelleri, Rasch modeli, minInfo, kesinlik, randomesque, madde kullanım sıklığı

## Abstract

The item bank development stage in computer adaptive testing is extremely challenging. It is assumed that the item difficulties are constant among different sub-groups, different positions and various test forms. The items violate these assumptions are eliminated. This may result in more time-consuming item pool development stage. The main purpose of this research is to investigate how average test length, item exposure, test overlap and precision of ability parameters change when explanatory item response models are utilized in computer adaptive testing. The study analysis conducted with simulated 10 item pools with 100 items and 1440 candidates in each. Each item bank calibrated using Rasch model, latent regression, linear logistic test model and latent regression linear logistic test model. Next, response patterns and prior ability estimates used for post-hoc simulations conducted in 10 replications for each item bank and models. The simulations are based on EAP estimation, two stop rules (precision and minInfo) and the item exposure control rule randomesque. The computer adaptive testing simulations based on explanatory item response models conducted using a modified version of "catR". It is reported that if the sub-groups in population are ignored in post-hoc simulations all models estimate very similar ability score mean. It also found that explanatory item response models have no effect on average test length, test overlap and item exposure rate. It is an important finding that latent regression and linear logistic test model achieved to reduce item exposure rate for the first 20 items.

**Keywords:** computer adaptive test, explanatory item response models Rasch model, minInfo, precision, randomesque, item exposure

## Teşekkür

Doktora tez sürecimin her aşamasında desteğini yanımda hissettiğim, katkıları ile gelişimim üzerinde büyük pay sahibi olan değerli danışmanım Prof. Dr. Nuri DOĞAN'a, lisansüstü eğitim hayatımın neredeyse tamamında kendilerinden çok şey öğrendiğim ve örnek aldığım değerli hocalarım Prof. Dr. Selahattin GELBAL ve Prof. Dr. Hülya KELECİOĞLU'na, tezin planlanması ve yazılması sürecinde önerileri ve eleştirileri ile katkılarını esirgemeyen değerli hocam Prof. Dr. Şeref TAN'a, tez savunma jürimde önemli geri bildirimlerinden faydalandığım sayın hocam Doç. Dr. Celal Deha DOĞAN'a, dört yıllık zaman diliminde birçok konuda fikir alış verişi yaptığım sevgili oda arkadaşlarım Arş. Gör. Dr. Başak ERDEM KARA ve Arş. Gör. Dr. Abdullah Faruk KILIÇ'a ve yoğun çalışma şartlarında beni her zaman destekleyen ve bana anlayış gösteren sevgili eşim Özlem KÖSELE TAT'a teşekkür ederim.

## İçindekiler

Kabul ve Onay.....	i
Öz.....	ii
Abstract.....	iii
Teşekkür.....	iv
Tablolar Dizini.....	vii
Şekiller Dizini.....	ix
Simgeler ve Kısaltmalar Dizini.....	xi
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	6
Araştırma Problemi.....	7
Sınırlılıklar.....	8
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	9
Madde Tepki Kuramı.....	9
Açıklayıcı Madde Tepki Modelleri.....	16
Bilgisayar Ortamında Bireye Uyarlanmış Test.....	22
İlgili Araştırmalar.....	28
Bölüm 3 Yöntem.....	36
Araştırmanın Türü.....	36
Çalışma Grubu.....	36
Verilerin Analizi.....	40
Değerlendirme Kriterleri.....	48
Bölüm 4 Bulgular ve Yorumlar.....	51
Birinci Alt Probleme İlişkin Bulgular ve Yorumlar.....	51
İkinci Alt Probleme İlişkin Bulgular ve Yorumlar.....	56
Üçüncü Alt Probleme İlişkin Bulgular ve Yorumlar.....	62

Dördüncü Alt Probleme İlişkin Bulgular ve Yorumlar.....	64
Beşinci Alt Probleme İlişkin Bulgular ve Yorumlar.....	90
Altıncı Alt Probleme İlişkin Bulgular ve Yorumlar .....	93
Yedinci Alt Probleme İlişkin Bulgular ve Yorumlar.....	96
Sekizinci Alt Probleme İlişkin Bulgular ve Yorumlar .....	97
Dokuzuncu Alt Probleme İlişkin Bulgular ve Yorumlar .....	97
Onuncu Alt Probleme İlişkin Bulgular ve Yorumlar.....	98
Bölüm 5 Sonuç, Tartışma ve Öneriler .....	100
Sonuç.....	100
Tartışma.....	101
Öneriler .....	104
Kaynaklar .....	106
Ek A. Madde Havuzlarına İlişkin Faktör Yükleri.....	116
EK B. Birinci Birey Değişkenine Göre DMF Analizi Sonuçları .....	118
Ek C. İkinci Birey Değişkenine Göre DMF Analizi Sonuçları .....	120
Ek Ç. Birinci Madde Değişkenine Göre DMF Analizi Sonuçları .....	123
Ek D. İkinci Madde Değişkenine Göre DMF Analizi Sonuçları.....	125
EK-E: Etik Komisyonu Onay Bildirimi .....	127
EK-F: Etik Beyanı .....	128
EK-G: Doktora Tez Çalışması Orijinallik Raporu .....	129
EK-Ğ: Dissertation Originality Report .....	130
EK-H: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı.....	131

## Tablolar Dizini

Tablo 1 Açıklayıcı Madde Tepki Modelleri.....	18
Tablo 2 Madde Havuzu Üretim Koşulları .....	38
Tablo 3 Alt Gruplar Bazında Katılımcı Sayısı .....	38
Tablo 4 Madde Havuzuna İlişkin Faktör Yükleri .....	42
Tablo 5 Madde Havuzlarındaki İlk 20 Madde İçin DMF Analizi Sonuçları .....	44
Tablo 6 Madde Güçlük Parametrelerine (b) İlişkin Betimsel İstatistikler .....	51
Tablo 7 Madde Güçlük Parametrelerine İlişkin Tekrarlı ANOVA Sonuçları .....	53
Tablo 8 Madde Güçlük Parametrelerine İlişkin İkili Karşılaştırma Sonuçları.....	54
Tablo 9 Madde Havuzundan Kestirilen Yetenek Puanlarına İlişkin Betimsel İstatistikler .....	56
Tablo 10 Madde Havuzlarında Alt Gruplar İçin Kestirilen Yetenek Puanlarına İlişkin t-Testi Sonuçları .....	58
Tablo 11 Madde Havuzlarındaki Kestirimlere İlişkin Model-veri Uyum İndeksleri.	62
Tablo 12 Rasch Model Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Betimsel İstatistikler .....	64
Tablo 13 Rasch Model Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları .....	66
Tablo 14 ÖR Modeli Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Betimsel İstatistikler .....	68
Tablo 15 ÖR Modeli Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları .....	69
Tablo 16 ÖR Modeli Parametrelerine Dayalı BBT Simülasyonlarında Alt Gruplar İçin Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları.....	71
Tablo 17 DLTM Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Betimsel İstatistikler .....	75
Tablo 18 DLTM Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları .....	76
Tablo 19 DLTM Parametrelerine Dayalı BBT Simülasyonlarında Alt Gruplar İçin Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları.....	78
Tablo 20 ÖRDLTM Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Betimsel İstatistikler .....	83

Tablo 21 ÖRDLTM Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları .....	84
Tablo 22 ÖRDLTM Parametrelerine Dayalı BBT Simülasyonlarında Alt Gruplar İçin Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları.....	85
Tablo 23 BBT Simülasyonlarında Model Bazında Kullanılan Ortalama Madde Sayısı .....	90
Tablo 24 BBT Simülasyonlarında Model Bazında Madde Kullanım Sıklığı .....	94
Tablo 25 BBT Simülasyonlarında Model Bazında İlk 20 Maddeye Ait Madde Kullanım Sıklığı .....	94
Tablo 26 BBT Simülasyonlarında Test Örtüşme Oranı .....	97



## Şekiller Dizini

Şekil 1. Rasch modeline ilişkin madde karakteristik eğrisi.....	15
Şekil 2. BBT işleyiş şeması .....	23
Şekil 3. Verilerin geniş ve uzun biçimleri .....	45
Şekil 4. AMTM için BBT post-hoc simülasyon şeması.....	47
Şekil 5. Alt gruplarda ortalama madde güçlüğü .....	53
Şekil 6. Madde güçlüklerine ilişkin ortalama RMSE değerleri.....	55
Şekil 7. Madde havuzunda alt gruplar için kestirilen ortalama yetenek puanları ..	57
Şekil 8. Madde havuzlarından kestirilen yetenek puanlarının dağılımı .....	58
Şekil 9. Madde havuzlarında kestirilen yetenek puanlarına ilişkin ortalama RMSE değerleri .....	59
Şekil 10. Madde havuzlarında kestirilen yetenek puanlarına ilişkin korelasyon katsayıları.....	61
Şekil 11. Madde havuzu bazında model-veri uyum indeksleri.....	63
Şekil 12. Rasch model parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarının dağılımı .....	65
Şekil 13. Rasch model parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarına ilişkin korelasyon katsayıları.....	67
Şekil 14. Örtük regresyon parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarının dağılımı .....	70
Şekil 15. ÖR modeli parametrelerine dayalı BBT simülasyonlarında alt gruplar için kestirilen yetenek puanlarına ilişkin ikili karşılaştırma analizi.....	72
Şekil 16. ÖR modeli parametrelerine dayalı BBT simülasyonlarında alt gruplar için kestirilen yetenek puanlarına ilişkin korelasyon katsayıları .....	74
Şekil 17. DLTM parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarının dağılımı.....	77
Şekil 18. DLTM parametrelerine dayalı BBT simülasyonlarında alt gruplar için kestirilen yetenek puanlarına ilişkin ikili karşılaştırma analizi.....	80
Şekil 19. DLTM parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarına ilişkin korelasyon katsayıları .....	82
Şekil 20. ÖRDLTM parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarının dağılımı.....	86

Şekil 21. ÖRDLTM parametrelerine dayalı BBT simülasyonlarında alt gruplar için kestirilen yetenek puanlarına ilişkin ikili karşılaştırma analizi.....	88
Şekil 22. ÖRDLTM parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarına ilişkin korelasyon katsayıları .....	89
Şekil 23. Rasch model parametrelerine dayalı BBT simülasyonlarında en sık karşılaşılan test uzunluğu frekansları .....	91
Şekil 24. ÖR modeli parametrelerine dayalı BBT simülasyonlarında en sık karşılaşılan test uzunluğu frekansları .....	92
Şekil 25. DLTM parametrelerine dayalı BBT simülasyonlarında en sık karşılaşılan test uzunluğu frekansları .....	92
Şekil 26. ÖRDLTM parametrelerine dayalı BBT simülasyonlarında en sık karşılaşılan test uzunluğu frekansları .....	93
Şekil 27. BBT simülasyonlarında model bazında ilk 20 maddeye ait madde kullanım sıklığı .....	95
Şekil 28. BBT simülasyonlarında model bazında başarı ile sonlandırılan test oranı .....	96
Şekil 29. BBT simülasyonlarında model bazında ortalama RMSE değerleri .....	98
Şekil 30. BBT simülasyonlarında model bazında ortalama yanlılık değerleri .....	99

## Simgeler ve Kısaltmalar Dizini

**AMO:** Ağırlıklandırılmış Maksimum Olabilirlik (WLE)

**AMTM:** Açıklayıcı Madde Tepki Modelleri

**BBT:** Bilgisayar Ortamına Bireye Uyarlanmış Test

**BMO:** Birlikte Maksimum Olabilirlik (JML)

**BSD:** Beklenen Sonsal Dağılım (EAP)

**DLTM:** Doğrusal Lojistik Test Modeli

**DMF:** Değişen Madde Fonksiyonu

**DOKM:** Doğrusal Olmayan Karma Modeller

**DYF:** Değişen Yüzey Fonksiyonu

**GDKM:** Genelleştirilmiş Doğrusal Karma Modeller

**GDM:** Genelleştirilmiş Doğrusal Modeller

**KMO:** Koşullu Maksimum Olabilirlik (CML)

**KTK:** Klasik Test Kuramı

**MKE:** Madde Karakteristik Eğrisi

**MMO:** Marjinal Maksimum Olabilirlik (MML)

**MO:** Maksimum Olabilirlik (ML)

**MPK:** Madde Parametre Kayması

**MSD:** Modal Sonsan Dağılım (MAP)

**MTK:** Madde Tepki Kuramı

**ÖR:** Örtük Regresyon

**ÖRDLTM:** Örtük Regresyon Doğrusal Lojistik Test Modeli

**RM:** Rasch Model

**RMSE:** Hata Kare Ortalamasının Karekökü

## Bölüm 1

### Giriş

Bu bölümde problem durumuna, araştırmanın amacı ve önemine, problem cümlesine ve sınırlılıklara yer verilmiştir.

#### Problem Durumu

Davranış bilimlerinde kuramsal veya uygulamaya dönük birçok yapı doğrudan gözlemlenemez. Bu yapıları ölçmede temel işleyiş, ilgili yapıya dolaylı kanıtlar sağladığı düşünülen gözlemlenebilir değişkenlerin elde edilmesi şeklindedir. Bu gözlemlenebilir kanıtların elde edilmesi işlemi, ilgili yapının varlığını gösteren maddelerden oluşan testlerin geliştirilmesi anlamına gelir (Braeken ve Tuerlinckx, 2009). Test geliştirmenin ve uygulamanın geçmişi çok eskilere dayansa da standart doğrusal testlerin geliştirilmesi ve uygulanmasında en büyük gelişim 20. yüzyılda görülmüştür. Amerika Birleşik Devletleri'nde ilk defa 1901'de uygulanan üniversiteye giriş sınavı ve 1926 yılında uygulanan ilk Bilimsel Değerlendirme Sınavı'ndan (Scholastic Assessment Test, SAT) bu yana standart doğrusal testlerin üniversiteye yerleştirme, profesyonellik lisansı alma veya gelişimin izlenmesi gibi birçok amaçla kullanımı, dünyanın hemen hemen her yerinde yaygınlaşmıştır (van der Linden ve Glas, 2010). Test puanlarının incelenmesinde geleneksel olarak klasik test kuramından (KTK) faydalanılmaktadır. Bu kurama göre test puanları gerçek puan ve bir miktar hatanın toplamından oluşur. Başka bir ifadeyle gözlemlenen test puanının bir bölümü örtük değişken olan bireyin yeteneğinden, kalan bölümü ise hata teriminden meydana gelir (Boeck ve Wilson, 2004; Lord, 1980). Klasik test kuramına dayalı testlerden günümüzde hala yaygın bir şekilde yararlanılmasına karşın bu kuramın birtakım sorunlarının olduğu bilinmektedir. KTK'da kullanılan maddelere ait indeksler, maddelerin uygulandığı katılımcı grubundan elde edilir ve katılımcıların yetenek puanları teste dahil edilen maddelerin özelliklerine bağlıdır. Bu durum katılımcı ve test özelliklerinin birbirinden tam olarak ayıramaması sonucunu doğurmakta ve KTK'nın en zayıf yönü olarak görülmektedir (Hambleton, Swaminathan ve Rogers, 1991; Paek ve Cole, 2019). Günümüzde test geliştirme, tasarım ve puanlanmasının temel bileşeni olarak testin bütünü yerine bireysel maddelerin görülme eğilimi vardır. Buradan yola çıkarak testin nasıl yapılandırıldığına ve test puanına odaklanan KTK'nın yanında madde

karakteristiğine ve maddelerin nasıl bir test meydana getirdiklerine yoğunlaşan madde tepki kuramının (MTK) kullanımı yaygınlaşmaya başlamıştır (Ayala, 2009; Reckase, 2009).

Ölçme problemlerinin çözümünde geniş bir potansiyele sahip olduğundan ve kuramsal olarak doğrulanabilen ölçme ilkelerine dayandığından standart testlerin gelişiminde daha fazla kullanılmaya başlanan (Embretson ve Reise, 2000) madde tepki kuramı, maddelere verilen cevaplar ile bireylerin yetenek veya özellikleri arasındaki ilişkiyi tanımlayan matematiksel modeller bütünüdür (DeMars, 2010). Uygun şartlarda ve model veri uyumu sağlandığında KTK'ya alternatif olan bu test kuramı, madde karakteristiklerinin gruptan bağımsız olması, birey yeteneğini tanımlayan ölçümlerin testten bağımsız olması, modellerin test düzeyinden ziyade madde düzeyinde kurulması, güvenilirliğin kestiriminde birçok paralel test yapma zorunluğunu ortadan kaldırması ve her bir yetenek kestirimi için ayrı ayrı hata terimi sunması açısından oldukça güçlü bir yaklaşımdır (Ayala, 2009; Hambleton ve diğerleri, 1991). Her ne kadar ilk MTK modelleri Rasch veya üç parametrelilik lojistik model (3PL) gibi ikili puanlanan (dichotomous) verilere dayalı modeller olsa da günümüzde geniş bir psikometrik alanda uygulanabilen modeller mevcuttur (Embretson ve Reise, 2000). Boyutluluk bakımından tek ve çok boyutlu modeller, maddelerin puanlanması bakımından ikili, nominal, ordinal ve çoklu puanlanan (polytomous) modellerin yanında, parametrik olmayan modeller, cevaplama sürelerini dikkate alan modeller, monoton olmayan modeller, hiyerarşik cevap modelleri ve genelleştirilmiş doğrusal karma model (GDKM) ve doğrusal olmayan karma model (DOKM) yaklaşımından da bahsedilebilir (van der Linden, 2016).

Madde tepki kuramı modelleri madde cevaplarına uygulanan geleneksel modeller kümesinden çok daha geniş bir ailenin üyesidir. Bu bakış açısına geniş istatistiksel yaklaşım denir. Bu yaklaşıma göre birçok madde tepki modeli genelleştirilmiş doğrusal veya doğrusal olmayan karmaşık modellerin özel halidir (McCulloch ve Searle, 2001). Bir testte birden fazla maddeye verilen yanıtlar cevaplayıcıda yuvalanmış (nested) tekrarlı ölçümler olarak düşünülebilir. GDKM ve DOKM bu şekildeki normal dağılım veya dağılmayan tekrarlı ölçümlerin analiz edilmesini sağlayan modeller kümesidir (Jeon ve Rijmen, 2016). MTK modellerinin GDKM veya DOKM yaklaşımı ile ele alınması, madde ve kişiye ilişkin değişkenlerin, bu değişkenlerin etkileşiminin ve bunlar dışındaki birçok dışsal değişkenin sabit ve

rastgele etki olarak modellenmesinin yolunu açmıştır (Weirich, Hecht ve Böhme, 2014). Bu yaklaşımdan, bireylere ve maddelere ilişkin değişkenlerin etkilerinin tek aşamada modellenebildiği, açıklama ve betimleme özelliklerinin bir arada tanımlanabildiği açıklayıcı madde tepki modelleri (AMTM) doğmuştur (Boeck ve Wilson, 2004).

Madde tepki kuramındaki gelişmelere paralel olarak test geliştirme ve uygulamalarında da yeni yaklaşımlar gündeme gelmiştir. Bu yaklaşımlardan biri son yıllarda tanınırlığı artan bilgisayar ortamında bireye uyarlanmış test (BBT) yöntemidir. BBT etkileşimli bilgisayar teknolojisi yardımı ile bir testin geliştirilmesi, bireye uygulanması ve puanlanması yöntemi olarak tanımlanabilir (Reckase, 2009). BBT'nin geleneksel kâğıt-kalem testlerine göre birçok üstünlüğünün olduğu bilinmektedir. Daha yüksek ölçme kesinliği, daha düşük test süresi, standartlaştırılmış yönergeler ve esnek uygulama zamanı bu yöntemin en sık vurgulanan yönlerindedir. Çoğu BBT uygulamasında madde seçimi ve yetenek kestiriminde yararlanılan algoritmalar MTK'ya dayanmaktadır. Dolayısı ile, madde kalibrasyonunda faydalanılan model tek boyutlu modellerden biri ise tüm maddelerin oluşturduğu soru havuzunun tek boyutlu olması, maddelerin yerel bağımsız olması ve testin hız testi olmaması varsayımlarını karşılanması gerekmektedir (Segall, 1996).

Bilgisayar ortamında bireye uyarlanmış test uygulamalarında katılımcılara, potansiyellerini en verimli biçimde sergileyebilecekleri maddelerin sorulması amaçlanır. Farklı katılımcıların cevapladıkları ortak maddeler olmayabilir. MTK ilkeleri, bu durumda hem bireylere en uygun maddelerin seçiminde, hem farklı madde setlerinden elde edilen puanların eşitlenmesinde kullanılır (Embretson ve Reise, 2000). Bir testte katılımcıların soruları önceden elde etmesi onlara avantaj sağlayan ve sınavın güvenliğini zedeleyen bir durumdur. BBT uygulamalarında sınav güvenliğini sağlamak ve her yetenek düzeyine uygun madde uygulama olasılığını arttırmak için geniş bir madde havuzuna ihtiyaç duyulur (Glas ve van der Linden, 2003; McLeod ve Schnipke, 2006). Birçok test geliştirici, ortalama test uzunluğunun en az altı-sekiz katı arasında maddenin bulunduğu bir soru havuzunun yüksek kesinlik elde etmek için gerekli olduğunu belirtmiştir. Aynı zamanda teorik olarak ideal bir soru bankasındaki maddelerin güçlük parametrelerinin tek biçimli (uniform) dağılım göstermesi gerektiğini, ayırıcılık parametrelerinin yüksek ve şans

parametrelerinin düşük olması gerektiğini belirtmişlerdir. Bu temel varsayımların uygulamalarda karşılanması her zaman mümkün olmayabilir (Özbaşı ve Demirtaşlı, 2015; Segall, 2005).

Kâğıt-kalem testlerinde olduğu gibi BBT uygulamalarında da madde parametrelerinin, katılımcıların sahip olduğu yetenek düzeyi sabit tutulduğunda, demografik gruplar arasında farklılaşmaması beklenir. Başka bir ifade ile soru havuzundaki herhangi bir maddenin aynı yetenek düzeyine sahip farklı gruplardaki katılımcılara avantaj sağlamaması yani değişen madde fonksiyonu (DMF) göstermemesi gerekir. Maddelerin DMF göstermesi BBT uygulamalarında kâğıt-kalem testlerine göre daha önemli bir sorun olarak görülebilir. Çünkü BBT’de genellikle daha az madde uygulanır. Bu durumda her bir DMF’li maddenin yetenek puanı üzerindeki etkisinin oransal olarak daha yüksek olma ihtimali vardır (Zwick, 2000). Madde havuzu geliştirme aşamasında genellikle maddelerin DMF açısından incelenmesi ve önemli düzeyde DMF gösteren maddelerin düzeltilmesi veya elenmesi gerekir. Ancak bu durum madde havuzu geliştirme aşamasını daha da zorlaştırmakta veya test açısından önemli bazı maddelerden yararlanılamaması sonucunu doğurabilmektedir.

Uygulamalar arasında farklı işlemeyen madde parametrelerinin elde edilmesi için bireyin algısını etkileyebilecek dışsal faktörlerin (context effect) ortadan kaldırılması veya kontrol edilmesi hayatidir. Genellikle birden fazla maddeden oluşan testlerde maddelerin uygulanma sırası, başka bir ifade ile madde pozisyonu uygulamalar veya kitapçıklar arasında farklılık göstermektedir. Maddenin uygulanma sırası önemli bir dışsal faktördür. Çünkü bir maddenin testin sonuna yakın bir yerde uygulanması katılımcıların yorgunluğu veya artan dikkatsizliği nedeniyle testin başında uygulanmasına göre genellikle daha zorlayıcı olmaktadır (Weirich, Hecht ve Böhme, 2014). Bu durum, madde pozisyonlarının tamamen bireye bağlı olduğu BBT uygulamalarının önemli sorunlarındanıdır.

Madde tepki kuramı madde parametrelerinin alt gruplar arasında farklılaşmadığını varsayar. Ancak alt gruplar öğretim programındaki değişimler, alıştırımların tekrarlanması ve maddelere maruz kalma gibi birçok nedenden dolayı zaman içerisinde değişirler. Bu değişme zamanla madde parametrelerinin de farklılaşmasına neden olur ve bu durum madde parametre kayması (MPK, item parameter drift) olarak adlandırılır. PISA ve TIMSS gibi büyük çaplı uluslararası

sınavlarda madde parametrelerinin deęişmezlięi önce DMF ardından MPK analizleri yapılarak saęlanmaya alıřılır. Ancak, pilot ve asıl uygulamalarda gizil alt grupların varlıęı tek boyutlu MTK modelleri ile tam olarak belirlenemedięi iin DMF gibi MPK'nın da tam olarak nüne geilebildięi sylenemez (Park, Lee ve Xing, 2016). Kâğıt-kalem testlerinde olduęu gibi BBT uygulamalarında da madde parametrelerinin zaman ierisinde kayma gstermesi olasıdır. Bu durumda, farklı dnemlerde uygulanan maddelerin MPK gsterip gstermediklerinin belirlenmesi, gsteren maddelerin soru havuzundan ıkarılması veya bu soruna farklı bir czm getirilmesi gerekmektedir (van der Linden ve Glas, 2010).

Deęişen madde fonksiyonu, madde parametre kayması gibi durumların yanında bazı durumlarda madde zellięi veya yzeyi ile belirli bir grup zellięinin etkileşiminden bahsedilebilir. Bařka bir ifade ile madde parametreleri zerinde etkili olan madde zellięi, bazı durumlarda katılımcıların sahip olduęu grup yelięine baęlıdır. rneęin bir maddenin testin sonunda uygulanması kadın katılımcılara dezavantaja sebep olurken, erkekler iin bu durum gerekleşmeyebilir. Bylesine madde zellikleri veya yzeyi ile katılımcı grubu etkileşiminin madde parametreleri zerindeki etkisine deęişen yzey fonksiyonu (differential facet functioning) denir (Kim, Cohen ve Park, 1995). Deęişen yzey fonksiyonu (DYF) BBT uygulamalarında karřılařılabilecek istenmeyen durumlardan biridir.

Sonuç olarak, BBT'nin gerek kendi işleyişinden gerekse de dayandıęı madde tepki kuramından kaynaklanan birok sorunu olduęu grlmektedir. DMF, MPK veya DYF gsteren maddelerin elenmesi olduka zor olan madde havuzu geliřtirme srecini daha gleřtirebilmektedir. Madde kalibrasyonu srecinde maddelerin uygulanma pozisyonu sabitken, BBT uygulamalarında maddelerin uygulanma pozisyonu sabit deęildir. MTK modellerinin kullanıldıęı BBT uygulamalarında ayrıca MTK'nın doęası gereęi katılımcıların kiřisel farklılıkları gz ardı edilerek, madde formlarının eřitlilięi sınırlandırılarak veya uygulama srecinde ortaya ıkabilecek sorunlar (madde uygulanma sırası vb.) dikkate alınmayarak uyarlama sreci sadece bireyin yetenek puanına yoęunlaşmaktadır. Oysa birey ve madde deęişkenlerinin ve bunlarının etkileşiminin uyarlama srecinde dikkate alınmasının, BBT uygulamalarında yetenek parametresinin daha az hata ile kestirilmesi, ortalama test uzunluęu, madde kullanım sıklıęı zerinde olumlu etkisi olabileceęi dřnlmektedir (Park, Joo, Cornillie, van der Maas ve van den Noortgate, 2018). Gnmzde bahsi

geçen sorunların çözümünde etkili olabilecek farklı yöntemlere rastlamak mümkündür. Örneğin birden fazla grubun bulunduğu verilerde grup değişkeni, madde veya birey parametrelerinin kestiriminde çok gruplu madde tepki modelleri (multiplegroup IRT) ile modellenebileceği gibi birden fazla yüzeyin olduğu durumlarda çok yüzeyli MTK modeli (multifaceted IRT) kullanılabilir. Bu modellerin kullanıldığı durumların tümünü birlikte ele alabilen açıklayıcı madde tepki modellerinden yararlanılabilir. Maddeler üzerinde etkili olan birey değişkenleri birey açıklayıcı madde tepki modelleri yardımı ile kontrol altında tutularak yetenek kestiriminde DMF sorunu ile baş edilebilir. Madde açıklayıcı madde tepki modelleri kullanılarak maddelerin uygulanma pozisyonuna, madde parametre kaymasına ve maddelerin farklı formatlarının kullanıma ilişkin sorunlar çözümlenebilir. Ayrıca simülasyonlarda birey ve madde değişkenlerinin ortak etkisinin parametre kestirimleri üzerindeki etkisi kontrol altında tutulabilir. Ancak, kategorik birey ve madde değişkenlerinin kullanıldığı durumlarda çok gruplu MTK veya puanlayıcı güvenilirliğinin test edilmesinde kullanılan çok yüzeyli MTK modeli ile oldukça benzer sabit etkiler (madde güçlük parametresi) kestiren açıklayıcı madde tepki modellerinin BBT uygulamalarında kullanımına rastlanmamıştır.

### **Araştırmanın Amacı ve Önemi**

Madde ve bireye ilişkin eşdeğişkenlerin (kovaryantların) yanında gizil değişkenlerin de modellenebildiği, karmaşık ölçme durumlarına kolaylıkla uygulanabilen (Wilson, Zheng ve McGuire, 2012) açıklayıcı madde tepki modellerinin nispeten yeni olduğu ve uygulamalarına az rastlandığı bilinmektedir (Atar ve Aktan, 2013). Çoğu geleneksel madde tepki modelinde, bazı katılımcı gruplarının diğerlerinden daha yüksek performans göstermelerine etki eden faktörler veya bazı maddelerin belirli bir grup için daha zor olmasına sebep olan değişkenlere ilişkin açıklamalar yapılmaz. Bunun yerine sadece betimlemelerle yetinilir. Oysa faktörler üzerinden yapılabilecek açıklamalar madde ve bireye ait değişkenlerin de analize dahil edildiği açıklayıcı madde tepki modelleri ile mümkündür (Jeon ve Rijmen, 2016). Açıklayıcı madde tepki modellerinin doğrusal testlerde kullanımı zamanla artmaya başlamış olsa da BBT uygulamalarında kullanımına henüz rastlanmamıştır.

Bu arařtırmada, madde kalibrasyonunda birey aıklayıcı madde tepki modeli olan rtk regresyon kullanılarak BBT simlasyonlarında DMF'nin kontrol altında tutulmasının madde ve yetenek parametrelerine etkisi incelenmiřtir. Ayrıca alıřmada, maddelere iliřkin deęiřkenlerin modellenebileđi AMTM'nin bir tr olan doęrusal lojistik test model+hata (DLTM+hata) kullanılmıřtır. BBT uygulamalarında maddelerin farklı formunun kullanılması, madde parametre kayması veya madde uygulanma sırası gibi birok etkenin kontrol altında tutulmasının, parametreler zerindeki etkisi arařtırılmıřtır. Son olarak ifte aıklayıcı madde tepki modeli olan rtk regresyon doęrusal lojistik test modeli (RDLTM) kullanılarak, madde ve birey deęiřkenlerinin etkileřiminin BBT uygulamalarında kontrol edilmesi saęlanmıřtır. Bu kontrol durumunun parametre kestirimlerine etkisi gzlemlenmeye alıřılmıřtır. Bu alıřmanın, BBT uygulamalarında bahsedilen bu problemlerin kontrol edilmesi durumunda, parametrelerin nasıl etkilendiđini ortaya koyması aısından nemli olduđu dřnlmektedir.

### **Arařtırma Problemi**

Madde havuzlarında kestirilen madde ve yetenek parametreleri ve bilgisayar ortamında bireye uyarlanmış testlerde elde edilen yetenek parametreleri, ortalama test uzunluđu, madde kullanım sıklıđı, test rtřme oranı, RMSE ve yanlılık deęerleri a) deęiřen madde fonksiyonu, madde parametre kayması, maddelerin farklı formlarının kullanılması, maddenin uygulanma sırası, deęiřen yzey fonksiyonu ve benzeri durumlar aıklayıcı madde tepki kuramı modelleri ile kontrol altında tutulduđunda, b) Rasch modeli kullanılarak gz ardı edildiđinde nasıldır?

**Alt problemler.** Bu alıřma kapsamında ařađıdaki 10 alt probleme yanıt aranacaktır:

1. rtk regresyon, doęrusal lojistik test modeli ve rtk regresyon doęrusal lojistik test modellerinden elde edilen madde glk parametreleri Rasch modeli ile elde edilen madde glk parametrelerinden farklılařmakta mıdır?
2. rtk regresyon, doęrusal lojistik test modeli ve rtk regresyon doęrusal lojistik test modellerinden elde edilen yetenek parametreleri Rasch modeli ile elde edilen yetenek parametrelerinden farklılařmakta mıdır?

3. Rasch, örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinin model-veri uyumları nasıldır?

Rasch, örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinden elde edilen parametreler ile iki farklı sonlandırma kuralına (kesinlik ve minInfo) göre gerçekleştirilen BBT uygulamalarında;

4. kestirilen yetenek parametreleri nasıldır?
5. kullanılan ortalama madde sayısı nasıldır?
6. madde kullanım sıklığı nasıldır?
7. testi başarı ile sonlandırılan birey oranları nasıldır?
8. ortalama test örtüşme (overlap) oranı nasıldır?
9. ortalama RMSE değerleri nasıldır?
10. ortalama yanlılık (bias) değerleri nasıldır?

### **Sınırlılıklar**

1. Araştırma, gerçek veri özelliklerini yansıttığı düşünülen ve normal dağılım gösteren simülasyon verileri ile sınırlıdır. Birey ve madde değişkenlerinin ve bunların etkileşiminin parametreleri üzerinde etkili olduğu gerçek verilere ulaşamadığı için simülasyon verileri kullanılmıştır.
2. R platformunda yapılan açıklayıcı madde tepki modellerine ilişkin analizlerde teknik yetersizliklerle karşılaşılması (RAM sorunu) ve bu analizlerin ve BBT simülasyonlarının oldukça fazla zaman almasından dolayı madde havuzları 100 madde ile sınırlandırılmıştır.

## Bölüm 2

### Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu çalışmada madde tepki kuramında 1PL modelin özel hali olan Rasch modelinden, açıklayıcı madde tepki modellerinden (AMTM) örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinden yararlanılmıştır. Bunların dışında bu modellerden elde edilen parametrelerle post-hoc simülasyonlarına dayalı BBT uygulamalarına yer verilmiştir. Bundan dolayı bu başlık altında araştırmanın kuramsal temelini oluşturan MTK, AMTM ve BBT konularına sırası ile yer verilmiştir.

#### Madde Tepki Kuramı

Psikometrinin iki temel test kuramı olan klasik test kuramı ve madde tepki kuramı varlıklarını bir yıl arayla yayımlanan iki çalışmaya borçludurlar. Charles Spearman 1904 yılında “İki Şeyin Kanıtı ve Ölçülmesi” isimli makalesinde ilk kez daha sonra klasik test kuramı olarak adlandırılacak olan kuramın temel varsayımı olan belirli sayılılar altında gözlenen puanların gerçek ve rastgele hata puanlarının bütünü olduğu fikrini ortaya atmıştır. Alfred Binet'nin madde tepki kuramının gelişimindeki rolü ise Spearman'inki kadar belirgin olamamıştır. Binet'nin 1905 yılında yayımladığı makalesinin Fransızca yazılmış olmasının, makalenin geniş kitlelere ulaşamamasında etkili olduğu belirtilmektedir (Ostini ve Nering, 2006; van der Linden, 2016). Binet'nin fikrinin önemini farkına varan ilk kişi 1925'te yayınladığı makalesinde bu fikirden faydalanan Louis Thurstone olmuştur. Lord ve Rasch ise dağılım fonksiyonlarını maddelere verilen cevapların bir fonksiyonu olarak kullanmayı başaran ve bu konudaki karmaşayı çözen ilk araştırmacılar olmuşlardır (Baker ve Kim, 2017; van der Linden, 2016)

Lord ve Rasch'ın matematiksel temellerini oluşturduğu MTK, temel olarak belirli bir gizil değişken (yetenek, özellik vb.) düzeyinde, bireyin bir maddeye doğru cevap verme olasılığının bir dağılım fonksiyonu ile tanımlanmasını sağlayan modeller kümesidir (Hambleton ve diğerleri, 1991; Weirich ve diğerleri, 2014). Bu fonksiyona madde-cevap fonksiyonu denir. Örneğin, Eşitlik 1'de lojistik modellerin en genel hali olan dört parametrelili lojistik modele (4PL) ait madde-cevap fonksiyonu görülmektedir.

$$P_j(\theta) = (P(X_j = 1 | \theta, a_j, b_j, c_j, d_j) = c_j + (d_j - c_j) \frac{\exp [a_j(\theta - b_j)]}{1 + \exp [a_j(\theta - b_j)]} \quad (1)$$

Eşitlik 1’de görülen  $P_j(\theta)$ , j maddesinin doğru cevaplanma olasılığıdır ve  $\theta$  ne kadar yüksekse bu olasılık da o kadar yüksektir.  $(a_j, b_j, c_j, d_j)$  madde parametrelerine ilişkin vektördür. Bunlardan  $a_j$  ayırıcılık parametresi olup, madde güçlüğüne denk gelen noktada madde karakteristik eğrisine teğet olan doğrunun eğimine eşittir. Bu parametre maddeye doğru cevap verme olasılığının yetenek düzeyleri arasındaki değişiminin ölçüsüdür.  $b_j$  madde güçlük parametresidir ve y ekseninde c noktası ile 1 arasındaki mesafenin yarısına denk gelen yetenek düzeyini gösterir. Şans parametresi olarak adlandırılan  $c_j$ ,  $-\infty$  yetenek düzeyindeki (hiç yeteneğin olmadığı durum) bir bireyin j maddesine doğru cevap verme olasılığını temsil eder.  $d_j$  parametresi ise dikkatsizlik parametresi olarak adlandırılır ve bireyin stres, yorgunluk gibi durumlardan kaynaklı olarak bir maddeye yanlış cevap verme durumunun kontrol altında tutulmasını sağlamak için kullanılır (Lord, 1980; Magis, 2013). 4PL modelden dikkatsizlik parametresinin çıkarılması ile 3PL model elde edilir. 3PL modelden şans parametresi çıkarıldığında ise 2PL modele ulaşılır. 2PL modelde tüm madde ayırıcılıkları 0’a eşitlendiğinde 1PL model elde edilirken 1’e eşitlendiğinde Rasch modele ulaşılır. Madde tepkileri kategorik veya sürekli, iki kategorili veya çok kategorili, kategoriler sıralı veya sırasız olabileceği gibi test performansının kaynağı tek bir yetenek (tek boyutlu) veya çok sayıda yetenek (çok boyutlu) olabilir. Günümüzde tüm bu durumlara uygun MTK modeli mevcuttur (Hambleton ve Jones, 1993). Ancak, bu çalışmada tek boyutlu, ikili puanlanan maddelere uygulanan Rasch modeli kullanılacağından, bu modelin teknik özelliklerine ve varsayımlarına detaylı ver verilmiştir.

**Tek boyutlu madde tepki modelleri.** Günümüzde madde tepki uygulamalarında kullanılmak üzere çok sayıda model geliştirilmiştir. Bu modeller cevap örüntülerinin boyutluluğu (tek ve çok boyutlu), cevapların puanlanma düzeyi (dichotomous, polytomous), madde parametre sayısı (1PL, 2PL, 3PL, 4PL) ve madde karakteristik eğrisinin şekline ilişkin istatistiksel varsayımlar bakımından sınıflandırılabilir (Bejar, 1983; McDonald, 1982). Cevapların boyutluluğu bakımından birçok çok boyutlu madde tepki kuramı modeli geliştirilmiş (Spray, Davey, Reckase, Ackerman ve Carlson, 1990) olmasına karşın, günümüzde hala

tek boyutlu modeller daha yaygın kullanılmaktadır (Hambleton, Swaminathan ve Rogers, 1991).

Tek boyutlu MTK modelleri, test maddeleri ile birey karakteristiği (yetenek) arasındaki ilişkiyi tek bir parametrenin olduğu matematiksel ifade yardımı ile tanımlayan modeller bütünüdür (McDonald, 1982; Reckase, 2009). Bu bakımdan bütün tek boyutlu MTK modellerinin ortak varsayımı, bir soru setinin tamamının aynı niteliği ölçtüğüdür. Başka bir ifade ile tek boyutluluk varsayımdır (Finch ve French, 2015; Hambleton ve diğerleri, 1991). Bu modellerdeki tek boyutluluk, gözlenen değişkenlerin (maddelerin) gözlenemeyen örtük değişkenin fonksiyonu olduğu anlamına gelmektedir. Eğer birey bu tek boyutlu gizil uzaya sahipse bu uzayda konumu belirlenebilir ve başka bireylerle karşılaştırılabilir (Ayala, 2009).

Bir testte sonuç olarak tek bir yetenek puanının raporlanabilmesi için tüm maddelerin ortak, baskın bir yapıyı paylaştıklarının gözlemlenebilir kanıtlarının olması gerekir. Tek boyutluluğun test edilmesinin en basit yöntemlerinden biri maddeler arası korelasyon matrisine dayalı özdeğerlerdir (eigenvalues). Ancak, bu yöntem sadece sürekli verilerde kullanılabilir. Örtük yetenek puanının sürekli olduğu ve normal dağıldığı, gözlenen değişkenlerin (maddeler) ikili puanlandığı durumlarda tetrakorik korelasyona dayalı açıklayıcı faktör analizi ile tek boyutluluk test edilir (DeMars, 2010). Varsayımın ihlali durumunda, özellikle madde güçlüklerinin yüksek kestirildiğini belirten çalışmalar bulunmaktadır (Ansley ve Forsyth, 2016). Temel varsayımı olan tek boyutluluk ile diğer modellerden ayrılma da tüm MTK modellerinde olduğu gibi tek boyutlu modellerin üç varsayımı daha vardır. Birinci varsayıma göre bireyin sahip olduğu gizil değişkendeki bir değişim belirli bir maddenin doğru cevaplanma olasılığında değişime sebep olur ve bu durum madde karakteristik eğrisi (MKE) veya madde karakteristik fonksiyonu (MKF) denen grafikler yardımı ile gösterilir (Embretson ve Reise, 2000; Fox, 2010).

İkinci varsayıma göre tek boyutlu gizil bir değişkeni ölçen bir grup maddeye verilen yanıtlar, ilgili değişken (yetenek) sabit tutulduğunda birbirinden bağımsızdır. Yani yetenek sabit tutulduğunda, bireyin iki maddeye verdiği yanıtlar birbirinden bağımsızdır, birbirini etkilemez. Bu varsayımın tanımladığı durum yerel bağımsızlık olarak adlandırılır ve maddeye tepki verme şekli üzerinde tek bir gizil değişkenin etkili olması durumu olan tek boyutluluk şartına bağlıdır. Yani tek boyutluluk sağlanmadığında yerel bağımsızlıktan da bahsedilemez. Ancak tek boyutluluğun

varlığında yerel bağımsızlık aranabilir. Örneğin bir maddede doğru cevaba yönlendiren bir ipucunun varlığı veya başka bir maddenin çözümü kolaylaştıran bir bilgi sunması durumunda yerel bağımsızlık zarar görebilir. Çünkü maddeye doğru cevap verme olasılığı, bireyin ölçülmek istenen yeteneğinin yanında ipucunu keşfetme yeteneğinden de etkilenmiş olur. Ancak, yerel bağımsızlık çok boyutlu gizil değişken sabit tutulduğunda çok boyutlu MTK modellerine genellenebilir. Koşullu kovaryans kuramının perspektifinden, maddeler arasındaki ilişkinin yerel bağımlılık düzeyinde olduğu durumlarda bu ilişki modellenmediğinde çok boyutluluk gündeme geldiği vurgulanmaktadır (Ayala, 2009; Fox, 2010; Hambleton ve diğerleri, 1991; Levy, Mislevy ve Sinharay, 2009). MTK modellerinde tüm madde setinden yerel bağımsızlık ihlalinin test edilmesini sağlayan Q2 (van den Wollenberg, 1982) ve R2 (Glas, 1988) gibi yöntemlerin yanında, eşleştirilmiş maddelerden yerel bağımsızlık ihlalinin test edildiği Q3 (Yen, 1984),  $X^2$  ve  $G^2$  (Chen ve Thissen, 1997) gibi yöntemler de geliştirilmiştir (Debelak ve Koller, 2019). Yen'in Q3 yöntemi en sık kullanılan ve görece basit olan yöntemlerden biridir ve madde artık değerleri arasındaki korelasyona dayanır. Bu yöntemde madde çiftlerin artık değerleri arasındaki korelasyonun 0.2'den yüksek olmaması, yerel bağımsızlık şartının sağlandığı anlamına gelir (Finch ve French, 2015).

Son varsayıma göre MTK modellerinin kullanılacağı testin hız testi olmaması gerekmektedir. Çünkü testin hız testi olması durumunda maddelere cevap verme durumunu etkileyen yeni bir örtük değişkenin varlığı gündeme gelir. Örneğin tüm maddelere cevap vermek için yeterli zamanın olmadığı durumlarda maddelere cevap vermede öğrencinin konuya ilişkin yeteneğinin yanında, öğrencinin bireysel çabukluğu da etkili olur. Hızın da kontrol altında tutulmadığı yani tüm katılımcılar için sabitlenmediği durumlarda yerel bağımsızlık ihlal edilmiş olur (Ayala, 2009).

***Madde parametrelerinin kestirimi.*** Kalibre edilmiş madde havuzundan yetenek puanı kestirimi, madde parametre kestiriminden görece daha kolaydır. Madde parametrelerinin elde edilmesi daha karmaşık işlemlere dayanmaktadır. Hem yetenek hem de madde parametrelerinin bilinmediği durumda kestirim yapmak bağımsız değişkenlerin bilinmediği durumlarda lojistik regresyon analizi yapmaya benzetilebilir (DeMars, 2010; Embretson ve Reise, 2000). Günümüzde en sık kullanılan madde parametre kestirimi yöntemi Bock ve Aitkin (1981) tarafından geliştirilen marjinal maksimum olabilirlik yöntemidir. Bunun dışında birlikte

maksimum olabilirlik (BMO, joint maximum likelihood), koşullu maksimum olabilirlik (KMO, conditional maximum likelihood), en küçük kareler kestirimi ve eşleştirme kestirimi (pairwise estimation) de literatürde karşılaşılan yöntemlerden bazılarıdır.

Marjinal maksimum olabilirlik yönteminde bilinmeyen yetenek düzeyleri sorunu, cevap örüntüsüne ait olasılıkların, popülasyon dağılımından elde edilen beklentiler şeklinde ifade edilmesi ile çözülür (expectations). Bu yöntemde cevap örüntüleri popülasyondan rastgele elde edilen örneklemeler olarak kabul edilir. Kestirim yöntemi Bock ve Aitkin (1981) tarafından geliştirilen iterasyonlara dayanan beklenti-maksimizasyon (BM, expectaiton-maximization) algoritması ile yürütülür. Bu algoritmanın beklenti (B) ve maksimizasyon (M) olmak üzere iki aşaması olduğu belirtilmiştir. Beklenti aşamasında her bir yetenek düzeyinde (quarature point) olması beklenen kişi sayısının yanında her bir maddeyi doğru cevaplama beklenen kişi sayısı hesaplanır. Daha sonra maksimizasyon aşamasında kestirim eşitlikleri kullanılarak en yüksek olabilirliği sağlayan madde parametreleri elde edilir. Bu parametreler kullanılarak beklentiler tekrar güncellenir. Bu süreç belirlenen kesme noktasında yakınsama (convergence) sağlanıncaya kadar devam eder. Yakınsama sağlandığında, nihai parametrelerin ve standart hataların kestirilmesi için elde edilen değerlerle Newton-Gauss prosedürü işletilir (Embretson ve Reise, 2000; Sheu, Chen, Su ve Wang, 2005).

Birlikte maksimum olabilirlik kestiriminde amaç tüm veri olabilirliklerinin yetenek parametresi  $\theta$  ve madde parametresi  $p$  ile birlikte maksimize edilmesidir. Önce  $p$  madde parametresi sabit kabul edilip olabilirlik fonksiyonu  $L(\theta, p)$   $\theta$ 'ya göre maksimize edilir. Ardından  $\theta$  sabit tutulup  $L(\theta, p)$   $p$ 'ye göre maksimize edilir. Bu süreç iterasyonlar yardımı ile model parametrelerine yakınsama sağlanıncaya kadar tekrar eder ve her aşamada parametreler güncellenir. Koşullu maksimum olabilirlik kestirimi Rasch modelin bazı temel özelliklerine dayanmaktadır. Bazı durumlarda örtük yetenek düzeyinin kestirilmesi yeterli istatistiklerin elde edilmesi ile mümkündür. Örneğin Rasch modelde toplam doğru cevap sayısı yetenek puanının kestirilmesi için yeterli bir istatistiktir. Çünkü madde ayırıcılıkları eşit olduğundan hangi maddelerin doğru cevaplandığı, hangilerinin yanlış cevaplandığına bakılmaksızın aynı yetenek puanı kestirilir. Bu temel ilkeye dayalı olarak bilinmeyen yetenek düzeyi yerine bilinen yeterli istatistikler kullanılarak tüm olabilirlikler yeniden elde edilebilir. Bu şekilde birey parametrelerine ilişkin kestirimlere ihtiyaç

duyulmadan yetenek kestirilebilir. Bu yaklaşım KMO kestiriminin en büyük avantajıdır. KMO sadece Rasch modelde kullanılan bir yöntemdir.

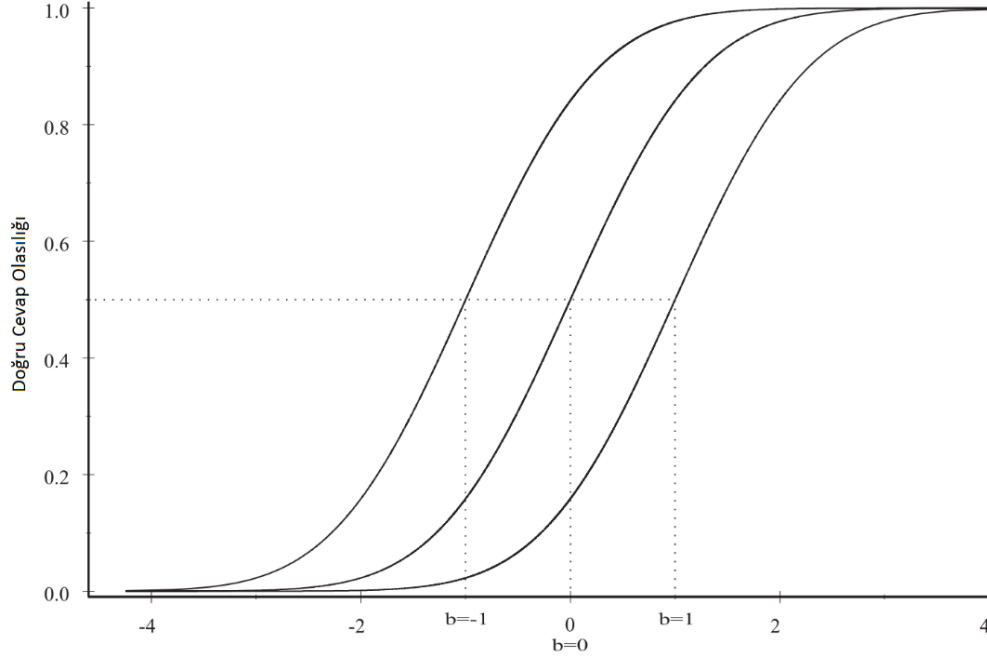
**Yetenek parametresinin kestirilmesi.** Standart testler gibi madde kalibrasyonlarının yapılarak madde parametrelerinin elde edildiği durumlar yetenek parametresinin hesaplanması için en kolay durumlardır. MTK'da maksimum olabilirlik (MO), ağırlıklandırılmış maksimum olabilirlik (AMO), modal sonsal dağılım (MSD) ve beklenen sonsal dağılım (BSD) en sık kullanılan yetenek kestirim yöntemlerinden bazılarıdır. Madde parametrelerinin mevcut olduğu durumlarda  $\theta$  yetenek düzeyine sahip bir bireyin, bir maddeye doğru cevap verme olasılığı  $p(\theta)$  ve yanlış cevap verme olasılığı  $1-p(\theta)$ 'dir. MO kestirimi yapılırken cevap örüntüsü göz önüne alınarak her bir madde için doğru veya yanlış cevap verme olasılıklarının olabilirlikleri, olabilirlik fonksiyonu ile hesaplanır. Ardından yerel bağımsızlığın sonucu olarak tüm olabilirlikler çarpılır ve olabilirlik dağılımı elde edilir. Dağılımın tepe noktası birey için MO kestirimi olarak kabul edilir (Boeck ve Wilson, 2004; DeMars, 2010).

Yetenek parametresinin, bireyin içinde bulunduğu evrendeki dağılımına önsel (prior) dağılım denir. Bayes istatistiğinde gözlenen veriye ait olabilirlik fonksiyonu ile önsel dağılım çarpıldığında ortaya çıkan sonuç sonsal olabilirlik (posterior likelihood) olarak adlandırılır. Oluşan dağılımın tepe noktası modal sonsal dağılım veya maksimum sonsal dağılım (MSD) yetenek puanı olarak adlandırılırken, çok sayıda replikasyonun ortalamasına dayalı kestirim puanına, beklenen sonsal dağılım (BSD) kestirimi denir (DeMars, 2010; Hambleton ve diğerleri, 1991; Reckase, 2009).

**Rasch modeli.** Tek boyutlu 1PL modelin özel bir hali olan Rasch modeli en basit ve en yaygın kullanılan MTK modellerinden biridir. Modelin genel matematiksel gösterimi Eşitlik 2'de görüldüğü gibidir. Eşitliğe göre  $\theta$  yetenek düzeyindeki i bireyin b güçlük düzeyindeki k maddesine doğru cevap verme olasılığı bireyin yeteneği ile maddenin güçlüğü arasındaki farka logit link yardımı ile bağlanmıştır. Bu genel olarak, bireyin yeteneği maddenin güçlüğünden ne kadar yüksekse maddeye doğru cevap verme olasılığı o kadar yüksektir şeklinde yorumlanır (Baker, 2001; Fox, 2010). Bir maddenin tüm yetenek düzeyleri ile doğru cevaplanma olasılığı arasındaki ilişki Eşitlik 2 yardımı ile kurulduğunda maddenin madde

karakteristik eğrisi elde edilir (Boeck ve Wilson, 2004). Şekil 1’de üç maddenin Rasch modeli ile elde edilen karakteristik eğrisi görülmektedir.

$$P(Y_{ik} = 1|\theta_i, b_k) = \frac{\exp(\theta_i - b_k)}{1 + \exp(\theta_i - b_k)} = (1 + \exp(\theta_i - b_k))^{-1} \quad (2)$$



Şekil 1. Rasch modeline ilişkin madde karakteristik eğrisi

Şekil 1 incelendiğine Rasch modelinin temel özellikleri göze çarpmaktadır. Örneğin eğri üzerinde maddenin doğru cevaplanma olasılığının %50 olduğu noktaya karşılık gelen yetenek düzeyi, maddenin b parametresi yani madde güçlük parametresi olarak bilinir. Yine her üç maddenin b yetenek düzeyinde eğriye çizilen dik çizginin kestiği noktada eğrilere teğet çizilen doğrunun eğimi birbirine eşittir. Bu noktadaki doğrunun eğimi maddenin ayırıcılık indeksi olarak bilinir. Bir parametrelili lojistik modelin özel bir hali olan Rasch modelinde ayırıcılık 1’e sabitlenmiştir. Bundan dolayı farklı madde güçlüğüne sahip maddelerin karakteristik eğrileri birbirine paralel konumlanmaktadır. Maddelerin sadece madde güçlükleri bakımından farklılaşması anlamına gelen bu durum Rasch modelin zayıf yönüdür. Şekil 1’de ayrıca maddelerin doğru cevaplanma olasılığının soldan sağa doğru gittikçe arttığı ve yetenek sonsuza doğru gittikçe olasılığın bire yaklaştığı görülmektedir (Fox, 2010; Lord, 1980).

Rasch 1960 yılında bağımlı değişken olarak kabul ettiği bir maddeyi doğru cevaplamanın log odds veya logitini üssel dağılım fonksiyonu yardımı ile yetenek ve

madde güçlüğü arasındaki farka eşitleyerek üssel dağılım fonksiyonları ailesinin önemli özelliklerinden faydalanmıştır. Bu sayede Rasch modelinde yetenek ve güçlük parametrelerinin cebirsel ayrımı mümkün olmaktadır. Ayrıca karşılanması zor olmayan az sayıda varsayımı olması, savunucuları tarafından modelin matematiksel olarak üstün olduğu şeklinde savunulmaktadır. Buna karşın modelin tüm madde ayırıcılıklarını sabitlemesi model karşıtları tarafından fazla sınırlayıcı bulunmasına neden olmaktadır (Fox, 2010; van der Linden, 2016).

### **Açıklayıcı Madde Tepki Modelleri**

Geleneksel madde tepki kuramı modelleri bireyin bilişsel becerilerine, başarısına veya yeteneğine ilişkin sunduğu bilginin yanında maddeye ilişkin güçlük düzeyine, ayırıcılık düzeyine ve model veri uyumuna dair bilgi verir. Ancak araştırmacılar ve uygulayıcılar bu betimsel ölçümleri kullanarak katılımcılar veya maddeler ile ilgili kararlar alırken madde tepki modelleri ölçme araçlarının tasarımından kaynaklı sistematik etkileri tanımlayamazlar. Başka bir ifade ile bu modeller, bu araçların tasarımından kaynaklı maddeler veya katılımcılar arasındaki ortak değişkenliği açıklayamazlar. Cevapların ortak noktasının ölçülmesi test geliştirme sürecinin önemli bir parçasıdır. Çünkü bu şekilde test geliştiriciler fiziksel, kültürel, bilişsel, iletişimsel, dilsel ve benzeri özelliklerin de aralarında bulunduğu özelliklerin yapıyla ilişkisini ölçebilmektedir. Örneğin, geleneksel MTK modellerinde bir testi alan katılımcıların testin dilini anadili olarak kullanması veya ikincil dil olarak kullanması ölçülen hedef yapı ile doğrudan ilgili değildir. Bu modellerle katılımcının testin dilini anadili ve ikinci dili olarak kullanmasından kaynaklanan performans farkını doğrudan ölçmek mümkün değildir. Benzer şekilde eşdeğer iki test formundan birinde soruların yarısının görsel olması, diğer formda bu maddelerin yazılı olması durumunda madde formatının, maddenin güçlüğü üzerindeki etkisini açıklayamaz (Stanke ve Bulut, 2019).

Yaygın kullanılan çoğu MTK modeli her bir madde için sadece bireysel güçlük parametresi ve tüm katılımcılar için yetenek parametresi kestirir. Bu ölçme modellerinde, madde ve birey parametrelerindeki farklılığın varlığına ve bu olası farklılığın nedenlerine ilişkin açıklayıcı değişkenlere yer verilmez (Hartig, Frey, Nold ve Klieme, 2012; Ravand, 2015). Ölçme işlemi madde ve birey parametrelerini betimlemekten çok, açıklama amacıyla yapıyorsa bu, birinci basamakta ölçme

sonuçlarının (madde veya yetenek parametrelerinin) elde edildiği, ikinci basamakta ise elde edilen ölçümlerle dışsal faktörlerin ilişkilendirildiği iki aşamalı bir süreç ile sağlanabilir. Örneğin geleneksel MTK modelleri ile alt gruplar ayrı ayrı ölçeklenir ve elde edilen yetenek puanları daha sonra dışsal değişkenlerin bağımsız değişken olarak kullanıldığı regresyon analizinde bağımlı değişken olarak kullanılabilir. İki aşamalı yaklaşım denen bu ölçme yaklaşımı ile değişkenlerin madde güçlükleri veya yetenek puanlarına etkisi incelenebilir (Briggs, 2008). Alternatif olarak madde parametrelerini açıklamak için dışsal değişkenlerin doğrudan modellendiği tek aşamalı bir yöntem kullanılabilir (Atar ve Aktan, 2013). Geleneksel MTK modellerinin yetersiz kaldığı durumlarda kullanımı yaygınlaşmaya başlayan açıklayıcı madde tepki modelleri (AMTM), madde ve bireye ait içsel ve dışsal değişkenlerin bir tek analizde modellenmesini sağlayan, açıklama ve betimleme işlemini aynı anda gerçekleştiren yeni bir psikometrik yaklaşımdır (Boeck ve Wilson, 2004; Kahraman, 2014). Bu yaygınlık özellikle son on yılda psikoloji, istatistik ve eğitim bilimleri araştırmalarında gözlenmektedir (Santi, Kulesz, Khalaf ve Francis, 2015).

AMTM aslında birçok geleneksel MTK modelinin geliştirilmiş doğrusal karma modeller (GDKM, generalized linear mixed models [GLMM]) ve doğrusal olmayan karma modeller (DOKM, non-linear mixed models [NLMM]) formunda tekrar modellenmesi yoluyla içsel ve dışsal değişkenlerin de rol aldığı esnek psikometrik yaklaşımlardır. Geliştirilmiş doğrusal modeller (GDM) sürekli değişkenler için kullanılan doğrusal regresyon yönteminin kategorik cevaplar için olan daha geniş halidir (Briggs, 2008; Wilson ve diğerleri, 2012). Gözlemlerle değişkenler arasındaki ilişki doğrusal bağlantı yardımı ile kurulmadan önce bağlantı fonksiyonunda dönüşüm gerektiren modeller GDM olarak adlandırılır. Geliştirilmiş doğrusal karma modeller ve doğrusal olmayan karma modeller ise GDM'nin daha geniş kümesidir ve tekrarlı ölçümler (madde tepkileri) gibi normal dağılım gösteren veya normal dağılmayan kümelenmiş verilerin analizinde kullanılan modellerin genel adıdır (Jeon ve Rijmen, 2016). Bu bakımdan hem GDKM'nin hem de DOKM'nin iki önemli istatistiksel modelin bir ürünü olduğu söylenebilir. Bunlar rastgele etkilerin modellenebildiği doğrusal karma modeller ve normal dağılmayan verilerle analizi mümkün kılan geliştirilmiş doğrusal modellerdir (Bolker ve diğerleri, 2009). Geleneksel MTK modellerinden Rasch ve 1PL model bir GDKM iken ayıricılık parametrelerinin bulunduğu 2PL veya 3PL gibi

modeller DOKM grubuna dahildir. Çünkü MTK modelleri ikili değişkenlerin (madde cevapları) beklenen değerlerinin logit veya probit dönüşümlerinin doğrusal modelleridir ve bu doğrusal modellerdeki bir veya daha fazla katsayı rastgele değişken olarak işlev görür (Sheu, Chen, Su ve Wang, 2005). Örneğin Rasch modelinde madde güçlüğü maddeye özgü sabit etki iken yetenek parametresi katılımcılar arasında rastgele değişen bir rastgele etkidir. Sabit etkiler gözlemlerin bir fonksiyonu olarak farklılık göstermeyen, bilinmeyen sabitlerken, rastgele etkiler genellikle normal dağılımdan çekilen ve gözlem birimlerine (katılımcılar) göre farklılık gösteren değerlerdir (Boeck ve diğerleri, 2011). AMTM genel olarak Tablo 1’de görüldüğü şekliyle sınıflandırılır. Madde ve birey değişkenlerinin her ikisinin de bulunmadığı modeller çifte tanımlayıcı model olarak adlandırılırken, her iki değişken türünün bulunduğu modeller çifte açıklayıcı modeller olarak adlandırılır. Birey değişkenlerinin modellendiği, madde değişkenlerinin modellenmediği durumlar birey açıklayıcı modeller, tersi durumlardaki modeller madde açıklayıcı modeller şeklinde adlandırılır (Boeck ve Wilson, 2004).

Tablo 1

*Açıklayıcı Madde Tepki Modelleri*

Madde Yordayıcıları	Birey Yordayıcıları	
	Yok	Var
Yok	Çifte Tanımlayıcı	Birey Açıklayıcı
Var	Madde Açıklayıcı	Çifte Açıklayıcı

**Çifte tanımlayıcı model (Rasch model).** Hiçbir açıklayıcı eşdeğişkenin bulunmamasından dolayı çifte tanımlayıcı model olarak adlandırılan ve AMTM’nin temelini oluşturan Rasch modeli GDKM kapsamında Eşitlik 3’teki gibi gösterilir.

$$\eta_{pi} = \theta_p X_{i0} + \sum_{k=1}^K \beta_i X_{ik} \quad (3)$$

Eşitlik 3’te  $\eta_{pi}$   $p$  bireyinin  $i$  maddesine doğru cevap verme olasılığı,  $\theta_p$   $p$  bireyine ait yetenek parametresi,  $\beta_i$   $i$  maddesine ait kolaylık parametresini temsil eder. Rasch modelinde  $X_{i0}$  tüm maddeler için bire eşittir. Eğer  $i=k$  ( $k=1, \dots, K$ ) ise  $X_{ik}$  bire eşittir, diğer durumlarda 0’a eşittir. Bu modelde birey yetenek puanlarının evrenden rastgele çekildikleri ve yetenek puanlarının normal dağıldığı, ortalamalarının 0’a eşit olduğu ( $\theta_p \sim N(0, \theta_p^2)$ ) varsayılır. Eşitlik 3,  $\eta_{pi} = \theta_p + \beta_i$  şeklinde basit bir forma dönüştürülebilir. Dikkatli incelendiğinde  $\beta_i$ ’nin bir soruyu

doğru cevaplama olasılığında pozitif katkı sağladığı yani maddeyi kolaylaştırdığı görülebilir. Bu parametre bu şekli ile madde güçlük parametresinin tersidir. Bu durum alışlagelen regresyon gösteriminden kaynaklıdır ve kestirim sonrası elde edilen katsayının negatifi alınarak madde güçlük parametresi elde edilir (Boeck, Bakker ve diğerleri, 2011). Eşitlik 3 sadece Rasch modelinin GDKM formunda gösterilmiş halidir ve açıklayıcı madde tepki modellerinin temel biçimi olduğu söylenebilir. Bu temel üzerine eklenecek değişkenler ile AMTM asıl anlamına kavuşmaktadır (Briggs, 2008).

#### **Birey açıklayıcı madde tepki modeli (örtük regresyon modeli).**

Katılımcılara ait açıklayıcı değişkenlerin eklendiği ama maddelere ait herhangi bir değişkeninin bulunmadığı modeller birey açıklayıcı madde tepki modelleri olarak bilinir. Birey açıklayıcı madde tepki modellerinden biri olan örtük regresyon modeli 1PL modelden elde edilen yetenek puanı için birey eşdeğişkenlerinin sabit etkilerinin olduğu örtük regresyon analizi anlamına gelir. Eşitlik 4 ve 5'te genel hatları ile örtük regresyon modeli görülmektedir.

$$\theta_p = \sum_{j=1}^J v_j Z_{(p)j} + \varepsilon_p \quad (4)$$

$$\eta_{pi} = \sum_{j=1}^J v_j Z_{(p)j} + \varepsilon_p + \beta_i X_{pi} \quad (5)$$

Eşitlik 4'te birey değişkeni olan  $Z_{(p)j}$   $j$  ( $j=1, \dots, J$ ) özelliğine sahip  $p$  bireyine ait değer,  $v_j$   $j$  kişi özelliği için sabit katsayıdır.  $\varepsilon_p \sim N(0, \theta_\varepsilon^2)$  birey değişkenleri kontrol altında tutulduğuna kalan birey etkisidir. Örtük regresyon modelinde yetenek puanlarının normal dağılması varsayımı vardır. Bundan dolayı bu model, örneklem evren ortalamasından farklılaşmadığı sürece faydalı olabilmektedir (Boeck, Bakker ve diğerleri, 2011).

#### **Madde açıklayıcı madde tepki modeli (doğrusal lojistik test modeli).**

Madde göstergelerinin yanında maddelere özgü değişkenlerin de dahil edildiği modellere madde açıklayıcı madde tepki modelleri denir. Madde yordayıcılarının modellenmesi doğrusal lojistik test modeli (DLTM) veya çok düzeyli MTK uygulamaları ile yapılabilir. Fischer tarafından geliştirilen doğrusal lojistik test modeli madde güçlük parametresinin doğrudan maddeye ilişkin faktörlerin bir fonksiyonu

şeklinde tanımlanmasını sağlayan bir yöntemdir (Hartig ve diğerleri, 2012). DLTM'de maddelere ait değişkenlerin nasıl modellendiği Eşitlik 6'da görülmektedir.

$$\eta_{pi} = \theta_p + \sum_{k=1}^K \beta_k X_{(p,i)k} \quad (6)$$

Eşitlik 6'da Maddelere ait eşdeğişkenin K tane kategorisi olduğunu varsayalım. Eğer i maddesi k özelliğine ilişkin değeri taşıyorsa bu durumda  $X_{(p,i)k}$  terimi bire, taşımıyorsa 0'a eşitlenir. Örneğin i maddesinin uygulandığı dönemler 2017 ve 2018 yılları olsun. Madde güçlüklerinin dönemlere göre nasıl olduğu incelenirken k 2017 olduğu durumda 2017 yılına ait güçlük parametresi kestirilir. DLTM yaklaşımı ile Eşitlik 6'da olduğu gibi sadece bir yordayıcının modellendiği yapılardan çok daha karmaşık modelleri üretmek mümkündür (Boeck ve diğerleri, 2011; Wilson ve diğerleri, 2012). Ancak eşitlikte sunulan modelde aynı kategorideki tüm maddeler için eşit ve hatasız madde kolaylığı (madde güçlüğü'nün negatifi) kestirilir. Bu temel haliyle tüm aynı kategorideki maddelerin eşdeğer olduğu varsayılır ki bu durum çoğu zaman gerçekçi olmayan bir varsayımdır. Modele her bir madde için madde düzeyindeki artık değer gibi işlev gören rastgele hata terimi eklenerek DLTM+hata modeli elde edilir (DiTrapani, Rockwood ve Jeon, 2018). Eşitlik 7'de görülen DLTM+hata modelinin yardımı ile her bir maddenin kendine özgü madde kolaylık parametresi hesaplanabilir. Yüksek hata varyansı madde değişkenlerinin güçlüklerini açıklamada daha az başarılı olduğu anlamına gelir (Boeck, Bakker ve diğerleri, 2011; Boeck, Cho ve Wilson, 2011; Cho, Boeck ve Lee, 2017).

$$\eta_{pi} = \theta_p + \sum_{k=1}^K \beta_k X_{(p,i)k} + \varepsilon_i \quad (7)$$

**Çifte açıklayıcı madde tepki modelleri (örtük regresyon doğrusal lojistik test modeli).** Hem madde hem de birey değişkenlerinin birlikte modellendiği ve en az sınırlamanın olduğu modelin genel hali Eşitlik 8'de görüldüğü gibidir.

$$\eta_{pi} = \sum_{j=1}^J v_j Z_{pj} + \varepsilon_p + \sum_{k=1}^K \beta_k X_{ik} \quad (8)$$

Eşitlik 8'de birey değişkenlerinden  $Z_{pj}$  eğer  $p=j$  şartı sağlanıyorsa bire eşittir. Aynı durum madde değişkeni olan  $X_{ik}$  için de geçerlidir. Yani bir i maddesi k

özelliğini sağlıyorsa bire eşit olur. Diğer durumlarda bu değişkenler 0'a eşittir (Boeck, Bakker ve diğerleri, 2011). Birey ve madde değişkenlerinin etkileşim terimlerini de modelleyerek açıklayıcı MTK modellerini daha kompleks hale getirmek mümkündür. Daha önce de bahsedildiği gibi sunduğu esneklikten dolayı GDKM yaklaşımının, MTK modellerinin sınırlayıcı birçok yönünden sıyrılmanın yeni ve önemli bir yolu olduğu söylenebilir.

**Kestirimde kullanılan yazılımlar ve model-veri uyumu.** R ve SAS gibi büyük platformlardaki paketlerde kullanılan en yaygın iki kestirim yaklaşımı log-likelihood yakınsama (approximating the log-likelihood) ve model yakınsamadır (approximating the model). İkinci kategorideki algoritma Laplace yakınsamadır ve daha fazla ilerleme gösterilemeyen noktaya kadar doğrusal yakınsamanın tekrarlanmasına dayalıdır. Bu süreçte tipik olarak bir önceki basamakta kestirilen parametrelerle GDKM katsayıları güncellenir yeniden kestirim yapılır. İterasyon belirlenen kesme puanında değişimin olmadığı durumda sonlandırılır. SAS'ta en önemli iki paketten GLIMMIX Laplace yakınsamaya dayalı doğrusallaştırma yaklaşımını kullanırken NLMIXED log-likelihood'a dayalı integral yakınsama algoritması kullanılmaktadır (Zhang ve diğerleri, 2011). R platformunda GDKM kestirimi yapabilen çok sayıda paket bulunmaktadır. Bunlardan en önemlilerinden "lme4" (Bates, Mächler, Bolker ve Walker, 2015) birçok yakınsama yöntemini gerçekleştirebilmektedir. Örneğin paket log-likelihood yakınsama için Gauss-Hermite algoritması kullanabilirken, Laplace yakınsama yöntemini de sunmaktadır (Zhang ve diğerleri, 2011).

Açıklayıcı madde tepki modellerinde  $-2\text{LogLikelihood}$  değerlerinin karşılaştırılmasına dayalı hipotez testi, sadece iç içe geçmiş (nested) modellerin model-veri uyumunu karşılaştırmada kullanılmaktadır. Bu yöntem ayırık modellerin karşılaştırılmasında kullanılmamaktadır. Bunun dışında Akaiki bilgi kriteri (AIC), Schwartz'ın bayes bilgi kriteri (BIC) model-veri uyumunun karşılaştırılmasında kullanılmaktadır. Bu kriterin değerlerin küçülmesi model-veri uyumunun iyileşmesi anlamına gelir (Boeck, Bakker ve diğerleri, 2011).

**AMTM'nin uygulama alanları.** Bir sınavda testin birden çok formu olduğu durumlarda katılımcıların maddelere cevap verme davranışlarının test formundan bağımsız olduğu varsayılır. Bu varsayımının ihlal edildiği durumda madde konum etkisi veya test konum etkisinden bahsedilebilir. Maddeye cevap verme davranışının

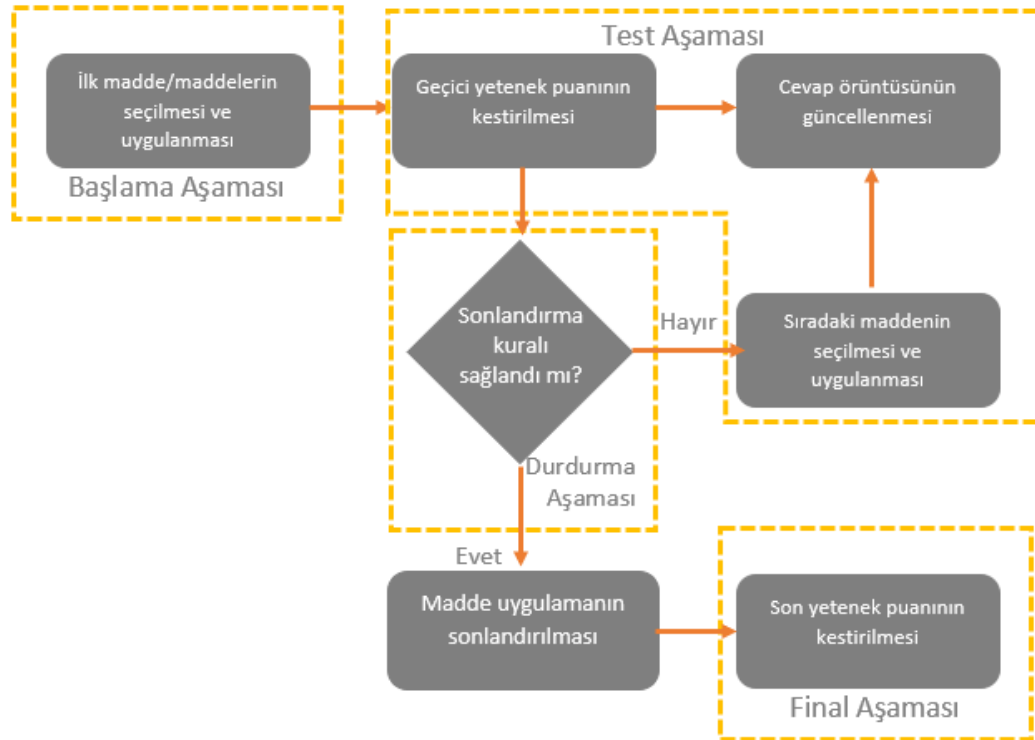
altında yatan yapıyla ilişkili olmayan dışsal bir etki katılımcıların yeteneği, madde parametreleri ve test karakteristiği (geçerlik, güvenilirlik vb.) hakkında yanlış çıkarımlara sebep olabilmektedir. Açıklayıcı madde tepki modelleri, madde konum etkisini ve alt testin konum etkisinin (çok sayıdaki alt testin uygulanma sırası) yanında bunlarının etkileşiminin de kontrol altında tutulmasını sağlayacak özelliklere sahiptir. Model seçimi test tasarımına bağlıdır. Örneğin farklı test formu sayısının az olduğu durumlarda çok gruplu MTK (multiple group IRT) veya çok yüzeyli MTK (multi-faceted IRT) modelleri kullanılabilir. Buna karşın GDKM'ye dayalı AMTM, çok sayıda madde ve test konum etkisini bir arada modellemeyi sağlayabilecek kadar esnekler. Bunun dışında katılımcı alt gruplarına bağlı değişen madde fonksiyonu (DMF), madde parametre kayması, değişen yüzey fonksiyonu (DYF) gibi birçok etkinin tespit edilmesinde ve istatistiksel olarak kontrol edilmesinde AMTM'den yararlanılabilir (Rose, Nagy, Nagengast, Frey ve Becker, 2019).

### **Bilgisayar Ortamında Bireye Uyarlanmış Test**

Bireye uyarlanmış test (BBT) fikri, ilk olarak Binet ve Simon tarafından 1908 yılında zekâ testi geliştirme sürecinde ortaya atılmış oldukça eski bir test yaklaşımıdır. Binet ve Simon uyarlama işlemini, geliştirdikleri maddelerin zekâ yaşlarına göre sınıflandırılması ve her deneğe zekâ yaşına uygun soruların sunulması şeklinde temellendirmişlerdir. Uygulamada deneklerin doğru cevap verdiği soruların ölçtüğü zeka düzeyinin, katılımcıların zeka düzeyi olduğu varsayılmıştır (Hambleton ve diğerleri, 1991; van der Linden ve Glas, 2010; Weiss, 1983). Bilgisayar teknolojisinin bulunmadığı dönemlerde kâğıt-kalem testi olarak uygulanan bu yöntemde maddelerin güçlük düzeylerinin belirlenmesi, uygun madde seçimi ve testin sonlandırılması aşaması test uygulayıcıları tarafından gerçekleştirilmiştir. Bu yönüyle ilkel olan uyarlamalı test yöntemi, esas fikri aynı kalmakla birlikte, bilgisayar teknolojisinin gelişimi ile şimdiki tanımına ulaşmıştır. Günümüzde artık BBT, bilgisayar teknolojisinin yardımı ile bir testin geliştirilmesi, katılımcılara uygulanması ve puanlanmasını sağlayan modern bir psikometrik yöntem olarak tanımlanmaktadır (Reckase, 2009). Son yıllarda ölçme kesinliği, kısa test süresi ve uygulama zamanının esnekliği gibi birtakım avantajlarından dolayı bilgisayar uyarlamalı testler popülerliğini arttırmakta ve kullanım alanını genişletmektedir. Amerika Birleşik Devletleri'nde lisansüstü eğitim kurumlarına

başvurmak için girilmesi gereken GRE (Graduate Record Examinations) 1993 yılında bilgisayar uyarlamalı test olarak yapılmaya başlanmış ve sonraki beş yıl içerisinde 500 binden fazla katılımcı GRE'nin BBT formu ile sınava tabi tutulmuştur (Zwick, 2000).

BBT uygulamasına geçilmeden önce olabildiğince tüm yetenek düzeylerini ölçebilecek güçlük düzeyine sahip çok sayıda maddeden oluşan bir madde havuzu geliştirilir. Tek boyutlu bir MTK modeli ile kalibrasyon yapılacaksa faktör analizleri yardımı ile madde havuzunun faktör yapısı ortaya konur. Tek boyutluluk ve yerel bağımsızlık şartını sağlamayan maddeler elendikten sonra kalan maddelerin parametreleri (madde gücüğü, ayırıcılık indeksi vb.) MTK modellerinden bir veya birkaçı yardımı ile kestirilir. Maddelerin ve testin genelinin model veri uyumu incelenir ve en iyi uyum sağlayan veya testin amacına en uygun olan modelin kestirdiği parametreler madde havuzunun parametreleri olarak kabul edilir. Nihai madde havuzunda ortalama bir BBT uygulamasında kullanılacak madde sayısının en az altı-sekiz katı kadar maddenin olması beklenir (Ayala, 2009; Segall, 2005). Madde havuzu geliştirme aşaması aslında BBT uygulamalarının en önemli aşaması olarak düşünülebilir. Çünkü, madde havuzu ne kadar iyi ise, BBT algoritmaları o kadar kaliteli işleyebilir (Wainer ve Dorans, 2000).



Şekil 2. BBT işleyiş şeması (Magis, Yan ve Davier, 2017)

BBT uygulamasının işleyişi genel olarak Şekil 2’de görüldüğü gibidir. BBT uygulamalarının işleyişi basit bir döngü içermektedir. BBT uygulamasına başlamadan önce başlatma kuralının belirlenmesi gerekmektedir. Başlatma kuralı olarak genellikle orta güçlükte bir sorunun seçilmesi yöntemi kullanılsa da katılımcılara ait önsel yetenek dağılımları veya son yetenek parametresinin hesaplanmasına katılmayan birkaç sorunun ardından kestirilen geçici yetenek parametresine göre en fazla bilgi veren maddenin seçilmesi de başlatma kuralı olarak kullanılabilir. Kural belirlendikten sonra başlatma kuralına uygun ilk madde uygulandıktan sonra maddeye verilen tepkiye göre geçici bir yetenek parametresi kestirilir. Bu noktada madde tepki kuramının lojistik modelleri, aşamalı tepki veya kısmi puan modeline dayalı yetenek kestirim yöntemlerinden MO, BSD veya MSD gibi kestirim yöntemleri devreye girer. Yetenek kestiriminin ardından sonlandırma kuralının sağlanıp sağlanmadığı sistem tarafından kontrol edilir. Sonlandırma kuralı BBT uygulamasının ne zaman sonlandırılacağına belirtildiği prosedür olarak tanımlanmaktadır. Uygulanan madde sayısının önceden belirlenmiş bir sayıya ulaşması, yetenek kestiriminin kesinliğinin veya standart hatasının belirli bir noktaya yakınsaması veya belirli bir güven aralığı içinde kalması sonlandırma kuralı olarak kullanılabilir. Geçici yetenek parametresinin elde edilmesinin ardından sonlandırma kuralı karşılanmışsa uygulama son bulur. Aksi halde cevap örüntüsü güncellenerek bir sonraki madde tekrar aynı yöntemlerle belirlenir ve süreç tekrar başlar. Bu durum sonlandırma kuralının karşılandığı zamana veya soru havuzunda uygulanacak madde kalmayınca kadar devam eder (Reckase, 2009).

BBT kritik üç temel sorunun cevabı aranarak tasarlanır. Bunlar:

1. Teste başlamak için ilk madde nasıl seçilmelidir?
2. İlk soru uygulandıktan sonraki maddeler nasıl seçilmelidir?
3. Testi sonlandırmaya nasıl karar verilir (Wainer ve Dorans, 2000)?

Bu üç soru bağlamında ilk ve sonraki maddelerin seçim algoritması, bu algoritmaların dayandığı geçici ve son yetenek puanı kestirimi ve testin sonlandırma kuralının BBT’nin temel bileşenleri olduğu söylenebilir (Segall, 1996). Yine bunlarla ilgili olarak bilgi fonksiyonlarının hesaplanması, madde kullanım sıklığının (exposure) kontrolü ve içerik dengeleme (content balancing) BBT’nin önemli alt başlıklarıdır.

**Başlama kuralı.** BBT'de maddelerin uygulanabilmesi genellikle bilgi fonksiyonlarına dayanmaktadır. Bilgi fonksiyonlarının hesaplanmasında madde parametrelerinin yanında yetenek puanının da bulunması gerekmektedir. BBT'de ilk maddenin uygulanmasında bilgi fonksiyonundan yararlanabilmek için bireyin varsa önsel yetenek puanı kullanılabilir. Bu şekilde yapılan işlemlerde mevcut yetenek düzeyinde en fazla bilgi sunan tek bir madde seçilebileceği gibi en fazla bilgi sunan birkaç maddeden rastgele bir madde de seçilebilir. Önsel bilginin olmadığı durumlarda bireyin geldiği popülasyonun ortalama yetenek puanı bireye atanabilir. Bu durumda test genellikle popülasyon ortalamasına göre orta güçlükteki bir madde ile başlatılır. Bunların dışında, son yetenek puanının hesaplanmasına dahil edilmeyen rastgele seçilen birkaç madde ile BBT başlatılabilir (Babcock ve Weiss, 2013; Segall, 2005).

**Madde seçimi ve yetenek kestirimi.** BBT uygulamalarında yetenek kestirimi ve madde seçimi eşzamanlı olarak sınav esnasında gerçekleşmektedir. Bundan dolayı bilgisayarların, test uzmanları ve psikometristler gibi karmaşık birçok ölçme probleminin üstesinden gelmesi gerekmektedir. MTK yaklaşımı ile gerçekleştirilen BBT uygulamalarında geçici ve son yetenek puanları daha önce bahsedilen kestirim yöntemleri olan MO, AMO, MSD veya BSD gibi yöntemlerle yapılır. Yine MTK'ya dayalı bir BBT uygulamasında ilk madde uygulandıktan sonraki maddelerin seçilmesinde bu geçici yetenek puanlarının veya dağılımlarının kullanıldığı maksimum bilgi kriteri (maximum information criterion), Owen'ın Bayes yaklaşımı (Owen's approximate bayes procedure), maksimum-global bilgi kriteri (maximum-global information criterion) ve Likelihood-ağırlıklandırılmış bilgi kriterinden (likelihood-weighted information criterion) yararlanır (van der Linden ve Glas, 2000).

Fisher tarafından geliştirilen test bilgi fonksiyonu beklenen bilgi değerini yetenek parametresinin bir fonksiyonu şeklinde tanımlamaktadır. Birnbaum (1968) doğrusal testlerin oluşturulmasında test bilgi fonksiyonunu temel kriter olarak kullanan ilk kişidir. İlk başlarda BBT için madde seçim kriteri olarak kullanılmasa da kısa zamanda bu yönü ile kullanılmaya başlanmıştır. Bu kriter gereğince k. madde için maksimum bilgi k-1. maddeye verilen cevap ve o maddede hesaplanan geçici yetenek puanı ile Eşitlik 9'daki gibi hesaplanır. Her aşamada hesaplanan geçici yetenek puanı ile madde havuzundaki en yüksek bilgiyi sunan madde belirlenir ve

uygulanır. Bu süreç sonlandırma kriterinin sağlanmasına kadar devam eder (Thompson, 2009; van der Linden ve Glas, 2000).

$$i_k = \operatorname{argmax}_j \{U_1, \dots, U_{k-1}, U_j(\hat{\theta}_{i1}, \dots, u_{1ik-1})\} : j \in R_k \quad (9)$$

BBT literatüründe en sık rastlanan bir diğer madde seçim yöntemi Kullback-Leibler (KL) bilgi fonksiyonuna dayalı madde seçim algoritmasıdır. KL bilgi fonksiyonu iki olabilirlik arasındaki mesafe ölçümüne dayanır. Bu mesafe ne kadar fazla ise yani KL ne kadar büyük ise iki olabilirlik o kadar kolay ayrılabilir. KL bilgi kriteri Eşitlik 10 yardımı ile hesaplanır.

$$K_k(\hat{\theta}_{k-1}, \theta_0) \equiv E \left[ \log \frac{L(\theta_0 | U_1, \dots, U_k)}{L(\hat{\theta}_{k-1} | U_1, \dots, U_k)} \right] = \sum_{k=1}^k K_{ik}(\hat{\theta}_{k-1}, \theta_0) \quad (10)$$

KL bilgisi cevap değişkenlerinin mevcut yetenek düzeyi olan  $\hat{\theta}_{k-1}$  ve gerçek yetenek puanı olan  $\theta_0$  arasındaki farkın (discriminate) derecesini verir (van der Linden ve Glas, 2000). Ancak, bireyin sahip olduğu gerçek yetenek düzeyi bilinmediğinden BBT'de mevcut yetenek düzeyi ( $\hat{\theta}$ ) kullanılarak olabilirlik fonksiyonu yardımı ile  $\theta_0$  kestirilmeye çalışılır (Magis, Yan ve Davier, 2017). Eğer başlangıçtaki geçici yetenek parametresi değeri nihai  $\theta$  değerinden çok farklı ise Fisher bilgi fonksiyonunun, CAT uygulamalarında başlangıçta kullanılması önerilmemektedir. Bunun yerine KL tavsiye edilmektedir (Thompson, 2009; van der Linden, 2016).

**Sonlandırma kuralı.** Simülasyon döngüsünün ne zaman sonlandırılacağına karar veren yöntemle sonlandırma kuralı denir. Belirli bir madde sayısına ulaşmak, katılımcının son yetenek puanının istenen kesinlik düzeyinde kestirilmesi veya madde havuzunda uygulanmayan maddelerin en yüksek bilgi düzeyinin belirlenen kesme puanının altında kalması en yaygın karşılaşılan sonlandırma kurallarıdır (Han, 2018; Reckase, 2009). BBT uygulamalarında en önemli hedef katılımcıların yetenek puanlarının kâğıt-kalem testlerine oranla daha az madde ile ve daha az hata ile kestirilmesidir. Bu bakımdan en yaygın kullanılan kesinlik (precision) sonlandırma kuralında her bir birey için yapılan simülasyon döngüleri belirlenen standart hata değerinin altına ulaşıldığı anda sonlandırılır. Diğer bir kural olan minInfo kuralı ise uygulama esnasında geçici yetenek düzeyinde kalan maddelerden elde edilen en yüksek bilgi eğer kesme değerinden daha küçük ise uygulama sonlandırılır (Magis ve Barrada, 2017).

**BBT'nin avantajları ve dezavantajları.** Alternatifi olarak kullanımı devam eden doğrusal testlerle kıyaslandığında BBT'nin, çok daha az madde ile yeteneği etkin bir biçimde kestirebildiği belirtilmiştir (Ali ve Chang, 2014). Ayrıca doğrusal kağıt-kalem testlerinde katılımcılara ait test sonuçlarından yüksek düzeyde bilgi elde etmek daha zordur (van der Linden, 2005). Bu avantajlarına karşın BBT uygulamasına ilişkin birtakım zorluklar da bulunmaktadır. BBT uygulaması sırasında madde seçim algoritmalarının kriterlerine uyan bazı maddeler birden fazla kullanılabilirken, bazı maddeler hiç kullanılmayabilir. Bu durumda maddelere birden fazla maruz kalan bireylerin bu maddeleri ezberleme ihtimalleri artar. Ayrıca tekrar tekrar aynı sorunun uygulanması katılımcıların motivasyonunu da etkileyebilir. Bu durumda, özellikle zor maddelere ait ön bilgisi olan katılımcılar test sonuçlarını ve test güvenliğini etkileyebilir. Bazı test uygulayıcıları soru havuzunu sık sık güncelleyerek veya madde klonlama yöntemlerinden yararlanarak bu durumun önüne geçmeye çalışsa da bu oldukça zor bir süreçtir (Glas ve van der Linden, 2003; McLeod ve Schnipke, 2006). Bu sorunu gidermenin başka bir yolu ise maddeye maruz kalma (exposure) durumunun kontrol edilmesinde kullanılan yöntemlerden yararlanmaktır. Örneğin maddeye maruz kalma yöntemlerinden Sympton-Hetter yöntemiyle maddeye maruz kalma durumu kontrol edilebilir. Ancak, bu yöntemde uygun kontrol parametrelerinin belirlenmesi için önceden tüm olasılıkların hesaba katıldığı uzun simülasyonlar yapmak gerekir (van der Linden, 2006).

Alanyazında BBT uygulamasının beklenen tutarlılık düzeyine erişilmeden sonlandırıldığı, çok fazla sayıda maddenin uygulandığı veya istenen ölçme kesinliğini karşılamada başarısız olduğu da rapor edilmektedir. Sonlandırma kriterinin testin ve katılımcıların psikometrik özellikleri veya katılımcıların tepki verme şekilleri göz önünde tutulmadan belirlenmesi, sözü edilen problemlerle karşılaşılmasına sebep olabilmektedir (Babcock ve Weiss, 2013). Ayrıca BBT uygulamalarında birden fazla konu alanında test yapılıyorsa her bir katılımcıya benzer içeriğe sahip test uygulanması veya her konu alanından maddelerin sorulmasını sağlamak her zaman mümkün olmamaktadır. Maksimum bilgiyi elde edecek ve her katılımcıya benzer konu içeriğinin sorulmasını sağlayacak soru setlerinin, uygun madde seçim algoritmaları ile belirlenmesi, belirtilen sorunların çözülmesinde etkili olabilir (van der Linden, 2005).

## İlgili Arařtırmalar

**Açıklayıcı madde tepki modellerine ilişkin arařtırmalar.** Randall, Cheong ve Engelhard (2010) yürüttükleri çalışmada geniş ölçekli sınavlarda geleneksel madde tepki modellerinin özel hali olan çok yüzeyli MTK modelinin madde düzeyindeki ölçme değişmezliğini ölçmek için kullanımını ve açıklayıcı madde tepki modellerinin ölçme değişmezliği ölçümlerinin nasıl geliştirebileceğini göstermeyi amaçlamışlardır. Çalışmanın verileri Gürcistan'da ülke genelinde yerleştirme amacıyla kullanılan problem çözme testinden elde edilmiştir. 74 okulda kümelenmiş, aralarında engellilerin de olduğu 868 öğrencinin verilerinin kullanıldığı arařtırmada arařtırmacılar genel anlamda ölçme değişmezliği, özel anlamda ise öğrencilerin engellilik durumlarına göre değişen madde fonksiyonuna ilişkin kavramsal bir çerçeve geliştirmişlerdir. Arařtırmada her iki yöntemin de katılımcıların engellilik durumlarına göre madde düzeyindeki ölçme değişmezliğini belirleme noktasında benzer sonuçlar ürettiği belirlenmiştir.

Bulut, Palma, Rodriguez ve Stanke (2015) gençlerde algılanan destek ve pozitif kişiliğin 6. ve 12. sınıflar olmak üzere sınıflar düzeyinde ve Latin öğrencilerde İngilizce öğrenen alt gruplar arasında ölçme değişmezliğini açıklayıcı madde tepki modelleri ile incelemiřlerdir. Arařtırmada Latin öğrencilerin destek ve pozitif kişilik algılarının gelişimsel düzeylerine ve dil öğrenme durumlarına göre farklılaştığı belirlenmiştir. Pozitif kişilik maddelerine cevap verme durumunun orta okuldan liseye geçişte farklılaştığı vurgulanmıştır. Ayrıca açıklayıcı madde tepki modelleri ile aynı anda hem değişen madde fonksiyonunun hem madde parametre kaymasının modellenebileceği belirtilmiştir.

Chen, Yang, Bulut, Cui ve Xin (2019) uyuşturucu madde kullanımını etkileyen kişilik faktörlerini 573 kişiden elde edilen verilerle, açıklayıcı madde tepki modelleri yardımıyla incelemiřlerdir. Arařtırmada kaygı düzeyi, dürtüsellik, duygu arayışı ve umutsuzluk değişkenleri birey değişkeni olarak kullanılmıştır. Çalışmada cinsiyet, alkol kullanımı ve bu iki değişkenin etkileşiminin yasa dışı madde kullanımını manidar bir şekilde yordadığı rapor edilmiştir. Kişilik özelliği olarak da sadece duygu arayışının madde kullanımını manidar bir şekilde yordadığı belirlenmiştir. Ayrıca cinsiyet ve umutsuzluğun etkileşiminin de madde kullanımında manidar bir yordayıcı olduğu sonucuna ulaşılmıştır.

Kahraman (2014) çalışmasında madde sayısının oldukça az olduğu bir performans testinin parametrelerinin kestirimine açıklayıcı madde tepki modellerinin katkı sağlayıp sağlamadığını incelemiştir. 767 kişiye uygulanan altı maddelik bir testin verilerine dayanan araştırmada herhangi bir açıklayıcı değişkenin olmadığı, sadece cinsiyet değişkeninin olduğu, sadece maddeyi cevaplama süresinin olduğu, sadece maddenin uygulanma sırasının olduğu ve sadece başka bir performans testinin puanlarının olduğu dört farklı model, açıklayıcı kısmi puan modeli ile modellenmiştir. Çalışmada az sayıda sorunun uygulandığı otantik performans testlerinde, uygun açıklayıcı değişkenlerle birlikte modellendiğinde, açıklayıcı madde tepki modellerinin parametre kestiriminde etkili olduğu sonucuna ulaşılmıştır.

Weirich, Hecht ve Böhme (2014) çalışmalarında açıklayıcı madde tepki modellerinden doğrusal lojistik test model ( $DLTM+\epsilon$ ) ile maddelerin test içerisindeki konumlarının maddeye verilen tepkilere etkisini incelemişlerdir. Araştırmada sonuç olarak  $DLTM+\epsilon$  modelinin madde konumlarının etkisinin ve bu etkinin farklı alt gruplarda aynı olup olmadığının belirlenmesinde kullanılabileceği sonucuna ulaşılmıştır. Özellikle alt gruplarda madde konum etkisinin eşit olup olmamasının ortak maddeler üzerinden yapılan test eşitleme çalışmalarda önemli bir faktör olduğu vurgulanmıştır.

Atar (2011) TIMSS 2007 Türkiye matematik testi verilerini kullandığı çalışmasında Rasch model, örtük regresyon modeli, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modelinin sonuçlarını karşılaştırmıştır. Örtük regresyon modelinde birey değişkenlerinden sadece özgüven değişkeninin birey yetenekleri üzerinde manidar bir etkisi olduğu belirlenmiştir. Ayrıca doğrusal lojistik test modelinde maddelerin bilişsel süreçlerinin madde güçlüklerini nasıl etkiledikleri incelenmiştir. Buna göre aynı yetenek düzeyindeki bir bireyin bilgi gerektiren bir maddeyi doğru cevaplama olasılığı 0.5 iken bu oran, akıl yürütme gerektiren maddeler için 0.27'ye düşmektedir. Başka bir ifade ile akıl yürütme gerektiren maddeler daha zordur. Ayrıca hem birey hem madde değişkenlerinin aynı anda modellendiği örtük regresyon doğrusal lojistik test modelin örtük regresyon ve  $DLTM$  yerine kullanılabileceği belirtilmiştir.

Atar ve Aktan (2013) çalışmalarında TIMSS 2007 Türkiye fen bilgisi verilerini kullanarak 2PL örtük regresyon modeli ile tek boyutlu 2PL modelinin sonuçlarını karşılaştırmışlardır. Örtük regresyon analizi yardımı ile fen başarısı üzerinde

cinsiyetin, fen bilgisine yönelik özgüveninin ve ebeveyn eğitim düzeylerinin etkisi incelenmiştir. 2PL örtük regresyon modelinin geleneksel iki parametrelili lojistik modelden daha iyi model-veri uyumu gösterdiği belirlenmiştir. Ayrıca SAS prosedürü olan NLMIXED ile MULTILOG yazılım sonuçları karşılaştırılmış ve örtük regresyon analizinde her iki yazılımın da madde güçlükleri ve ayırıcılıklarını oldukça yakın kestirdiği sonucuna ulaşılmıştır.

Briggs (2008) çalışmasında 433 onuncu sınıf öğrencisinin bir fen bilgisi testine (Partnership for the Assessment of Standards-based Science) verdiği cevaplar ve demografik değişkenler ile açıklayıcı madde tepki modellerinin grup farklılıklarını karşılaştırmadaki potansiyelini göstermeyi amaçlamıştır. Çalışmada yetenek parametrelerinin katılımcıların etnik kökeni tarafından manidar bir biçimde yordandığı sonucuna ulaşılmıştır. Beyaz grup üyelerine oranla siyahi öğrencilerin 0.61 logit, hispanik öğrencilerin ise 0.62 logit daha düşük yetenek ortalamasına sahip olduğu belirlenmiştir.

Goldhammer, Martens ve Lüdtke (2017) yaptıkları araştırmada Programme for the International Assessment of Adult Competencies (PIAAC) çalışmasının Kanada verilerini kullanarak test sorumluluğunu (test taking engagement) açıklayıcı madde tepki modelleri ile incelemiştir. Yerleştirme amacıyla kullanılmadığı için motivasyonun düşük olduğu (low-stakes) PIAAC'ta test sorumsuzluğunun düşük eğitim düzeyi, düşük bilişsel yetenek ve testin dilinin katılımcının ana dili olmaması ile açıklanabileceği vurgulanmıştır. Cinsiyetin sorumsuzluk üzerinde etkisi olmamasına karşın sadece problem çözmede yaşın pozitif etkiye sahip olduğu belirlenmiştir.

Min, Zickar ve Yankov (2018) çalışmalarında madde özelliklerinin maddenin psikometrik özelliklerini ve kişilik ölçeği geliştirmeyi nasıl etkilediğini incelemiştir. Kişilik ölçeğinin dört kategoride (negatif kelime kullanma, ince düşünme, sosyal kabul ve yanlış anlaşılma) gruplandığı madde özellikleri madde ayırıcılığı ve madde güçlüğü ile açıklayıcı madde tepki modelleri yardımı ile etkileşime sokulmuştur. Araştırmada negatif kelime bulundurma maddelerinin ayırıcılıklarının daha yüksek olduğu tespit edilmiştir.

Fagginger Auer, Hickendorff ve van Putten (2018) çalışmalarında altıncı sınıftan küçük çocukların çok rakamlı bölme işlemlerinde zihinsel veya yazılı

stratejileri seçmede etkili olan faktörleri açıklayıcı madde tepki modelleri ile incelemişlerdir. Ayrıca eğitimin, zihinsel stratejilerden çok yazılı stratejileri seçme noktasında etkisini araştırmışlardır. Çalışmada, strateji seçimindeki kişisel farklılığın öğrencilerin matematiğe karşı değil, strateji seçimine karşı tutum ve inançlarından kaynaklandığı tespit edilmiştir. Ayrıca yazılı stratejilerin kullanımının aynı zamanda öğretmenin yazılı strateji kullanmanın önemini vurgulaması ile de ilişkili olduğu sonucuna ulaşılmıştır.

Hecht, Weirich, Siegle ve Frey (2015a) 19107 dokuzuncu sınıf öğrencisinin geniş ölçekli bir sınavdaki fen bilgisi testine ilişkin verileri ile madde konum etkisini ve madde çiftlerinin aynı kitapçıkta bulunma etkisini incelemişlerdir. Açıklayıcı madde tepki modelleri ile yapılan analizler sonucunda konum etkisi kontrol edilmesinin madde parametre kestirimlerine ait yanlışlık ve RMSE değerlerinde küçük bir pozitif etkinin gözlemlendiği belirtilmiştir. Öte yandan kitapçık etkisinin göz ardı edilebilecek düzeyde olduğu sonucuna ulaşılmıştır.

Hecht, Weirich, Siegle ve Frey (2015b) doğrusal lojistik test modeli yardımı ile geniş kapsamlı bir teste katılan 9044 dokuzuncu ve onuncu sınıf öğrencisinin verilerini kullanarak kitapçık etkisinin madde parametrelerine etkisini incelemişlerdir. 1021 maddenin 74 kitapçık formunda uygulandığı sıvanda göz ardı edilemeyecek düzeyde kitapçık etkisinin olduğu belirlenmiştir.

**Bilgisayar ortamında bireye uyarlanmış testlere ilişkin araştırmalar.** Yao (2012) çalışmasında simülasyon verisi ile farklı test uzunluğu ve sonlandırma kurallarında beş çok boyutlu bilgisayar uyarlamalı test madde seçim prosedürünü karşılaştırmıştır. Standart hata ve tahmin edilen standart hata azaltma sonlandırma kuralları ile beş madde seçim algoritması olan en küçük açı (minimum angle,  $A_g$ ), hacim ( $V_m$ ), doğrusal kombinasyonun hata varyansını küçültme ( $V_1$ ), birleşik skorun hata varyansını küçültme ( $V_2$ ) ve Kullback-Leibler (KL) bilgisinin performansları karşılaştırılmıştır. Araştırmada KL bilgisine dayalı madde seçim yönteminde en az sayıda madde ile istenen kesinlik düzeyine ulaşıldığı belirlenmiştir. KL'den sonra en başarılı yöntem  $V_m$  iken en fazla madde  $A_g$  yönteminde kullanılmıştır. Sonlandırma kurallarında ise standart hata yönteminin birkaç madde ile tahmin edilen standart hata azaltma yönteminden daha başarılı olduğu gözlemlenmiştir.

Van der Linden ve Glas (2000) çalışmalarında Hukuk Fakültesi Kabul Testi (LSAT) verilerini kullanarak ağırlıklandırılmış-sapma yöntemi (WDM) ve gölge test yaklaşımının bilgisayar uyarlamalı testlerdeki performanslarını karşılaştırmışlardır. Araştırmada her iki madde seçim yönteminin de madde kullanım sıklığı açısından eşit derecede iyi olduğu tespit edilmiştir. Buna karşın ağırlıklandırılmış-sapma yönteminin içerik dengelemede birtakım ihlallerinin olduğu, daha yüksek yanlılık değerleri ve daha az kesinlikte yetenek parametresi kestirdiği belirtilmiştir.

Eroğlu ve Kelecioğlu (2015) 250 ve 500 maddeden oluşan simülasyon verisi ile farklı yetenek kestirim yöntemleri kullanarak ölçme kesinliği ve ortalama test uzunluğunu incelemişlerdir. Maksimum olabilirlik ve beklenen sonsal dağılımın yetenek kestiriminde kullanıldığı çalışmada sabit test uzunluğu, standart hata, standart hata-minimum madde, teta yakınsama ve teta yakınsama-minimum madde sonlandırma kurallarından sabit uzunluğun 20 madde, standart hata için kesme noktasının 0.22 olduğu ve teta yakınsama kesme noktasının 0.02 olduğu durumlarda RMSE ve yanlılık değerlerinin diğer kriterlere göre küçük kestirildiği ve uyum katsayılarında farklılığa sebep olmadığı gözlenmiştir. Ayrıca ortalama test uzunluğu ile RMSE değerleri arasında negatif bir ilişki olduğu belirlenmiştir.

Chen, Hou, Fitzpatrick ve Dodd (2016) simülasyon verisi kullandıkları çalışmalarında bilgisayar uyarlamalı testlerde popülasyon dağılımının maksimum olabilirlik (MO) ve beklenen sonsal dağılım (BSD) kestirimine etkisini incelemişlerdir. MO ve BSD için normal dağılım gösteren önsel yeteneklerin kullanıldığı ve BSD için tekbiçimli (uniform) çarpık dağılım gösteren iki veri seti kullanılmıştır. Her ne kadar önsel yetenek puanlarına yakın kestirim yapmasa da BSD yöntemi hem normal dağılım hem de uniform dağılım gösteren önsel yetenek puanları ile MO yöntemine benzer sonuçlar kestirmiştir. Ayrıca BSD için optimal quadratür noktasının belirlenebilmesi için çok sayıda deneme yapmak gerektiği vurgulanmıştır.

Choe, Kern ve Chang (2017) çalışmalarında sabit test uzunluğunun kullanıldığı bilgisayar uyarlamalı testlerde madde seçim algoritması olarak beklenen cevaplama süresine dayalı maksimum Fisher bilgisi yönteminin, testin bitirilmesi için gerekli süre ve yetenek kestiriminin kesinliğine etkisini incelemişlerdir. Bu yöntemle yaptıkları simülasyonlarda madde kullanım sıklığında dengesizlik, test süresinde azalma ve yetenek puanının kestiriminde daha yüksek hata gözlenmiştir. Ayrıca beklene cevaplama süresinin optimal bir şekilde merkezlenmesi (centering) ve

ağırlıklandırılması ile aynı yetenek kestirim kesinliğinde daha düşük cevaplama süresi ortalaması ve varyansının yanında daha düşük test örtüşme (overlap) oranı saplanmıştır.

Chen, Hou ve Dodd (2016) çalışmalarında bilgisayar uyarlamalı testlerin ilk aşamalarında beş madde seçim kuralı olan Fisher bilgisi (FI), Fisher aralık bilgisi (Fisher's interval information, FII), Fisher'in sonsal dağılımlı bilgisi (FIP), Kullback-Leibler bilgisi (KL) ve Kullback-Leibler sonsal dağılımlı bilgisinin (KLP) yetenek kestiriminin kesinliğindeki etkililiğini incelemiştir. Simülasyonların ilk aşamalarında önsel yetenek puanının -3 ve -2 olduğu durumlarda FII, FIP, KL ve KLP'nin FI'dan bariz bir şekilde daha iyi olduğu ve kullanılan madde sayısı 10 ve daha fazla olduğu durumlarda ise herhangi bir yöntemin diğerine üstünlük sağlamadığı belirlenmiştir.

Revuelta, Ximénez ve Olea (2016) çalışmalarında 557 birinci sınıf psikoloji bölümü öğrencisine bir İspanyolca konuşanlar için İngilizce testini üç farklı uyarlama yöntemi (adaptive, easy adaptive, sabit uzunluk) ile uygulamışlardır. Aynı zamanda araştırmacılar maddelere tekrar dönme durumunu dört farkı (maddeye dönme yok, test sonunda dönme, her beş maddede bir dönme, her maddeden sonra dönme) şeklinde tasarlayarak maddelere tekrar cevap verme durumunun yetenek kestiriminin kesinliğine etkisini incelemiştir. Araştırmada sonuç olarak yetenek kesinliği ile uyarlama türü arasında bir ilişki olduğu gözlenmiştir. Aynı zamanda maddeye tekrar dönme durumunun sonuçlar üzerinde önemli olduğu sonucuna ulaşılmıştır.

Leroux, Lopez, Hembry ve Dodd (2013) çalışmalarında madde kullanım sıklığı kontrol prosedürü olan süreçsel-sınırlanmış standart hata (PR-SE) ile daha yaygın kullanılan randomesque, Sympson\_Hetter (HS) ve madde kullanım sıklığı kontrolünün yapılmadığı durumları karşılaştırmışlardır. Küçük ve büyük iki madde havuzu ve sabit uzunluk ve değişken uzunluk sonlandırma kurallarına göre 3PL ile BBT simülasyonları gerçekleştirilmiştir. Çalışmada PR-SE'nin madde kullanım sıklığını kontrol noktasında SH ile benzer sonuçlar verdiği belirtilmiştir. Ayrıca her iki madde havuzunda da PR-SE yöntemi neredeyse tüm maddelere uygulanmışken, diğer yöntemler küçük madde havuzunun yaklaşık %52'sine, büyük madde havuzunun ise yaklaşık %80'ine uygulanmıştır. PR-SE'nin SH ile karşılaştırıldığında daha küçük test örtüşme oranına sahip olduğu vurgulanmıştır.

Leung, Chang ve Hau (2003) çalışmalarında içerik dengeleme (content balancing) yöntemlerini karşılaştırmışlardır. Araştırmacılar bu yöntemlerden sınırlandırılmış BBT (constrained CAT, CCAT), düzenlenmiş sınırlandırılmış BBT (modified constrained CAT, MCCAT) ve düzenlenmiş multinomial modelin (MMM) performanslarını farklı test uzunlukları ve farklı hedeflenen maksimum madde kullanım sıklığı oranında karşılaştırmışlardır. Araştırmada içerik dengelemenin ölçme kesinliğinde ve madde havuzundan yararlanma üzerinde sistematik bir etkisinin olmadığı belirlenmiştir. Bu üç içerik dengeleme yönteminden MMM'nin diğer yöntemlere kıyasla madde kullanım sıklığını düşürmede daha başarılı olduğu raporlanmıştır.

Şenel ve Kutlu (2018) çalışmalarında görme engelli öğrencilerin kullanabilecekleri işitmeye dayalı BBT uygulamaları tasarlayarak bu alanda yazılım geliştirme ve test tasarlama yöntemleri ile yeni uygulamalarının önünü açmayı amaçlamışlardır. Yazılım geliştirme ve ön deneme uygulamalarına 196, asıl uygulamalarda 608 görme engelli orta okul öğrencisi ile yürütülen çalışma kapsamında SesliTest olarak adlandırılan ve geçerliği-güvenirliği olduğu vurgulanan test geliştirilmiştir. Kestirimlerinde üç parametrelili lojistik modeli kullanan yazılım ile BBT'nin görme engelli bireyler için de kullanılabileceği belirtilmiştir.

Brown ve Iwashita (1996) araştırmalarında farklı dilsel geçmişi olan katılımcıların yerleştirme amacı ile kullanılan Japonca gramer testinin BBT uygulamalarındaki performansını incelemiştirler. Avustralya, Çin ve Japonya'dan 14000 öğrencinin her biri 225 çoktan seçmeli maddeden 50'sine cevap vermiştir. Madde güçlüklerinin farklı ülkelerden gelen öğrenci gruplarına göre oldukça farklı olduğu belirlenmiştir. Araştırmada farklı dilsel geçmişe sahip katılımcıların bu özelliği göz ardı edildiğinde testin geçerliğinin zayıflayabileceği vurgulanmıştır. Bu bakımdan testin bu şekliyle yetenek puanlarının kestiriminde yanlış sonuçlar doğurabileceği belirtilmiştir.

**İlgili araştırmaların özeti.** Açıklayıcı madde tepki modellerine ilişkin çalışmalar çok sayıda olmamakla birlikte son yıllarda artış göstermektedir. Bu alana ilişkin literatürde bazı çalışmalarda, geleneksel MTK yöntemleri ile yapılan kestirimler ile AMTM kestirimlerinin karşılaştırıldığı görülmektedir. Yine katılımcılara ilişkin değişkenlerin örtük regresyon analizi ile incelendiği ve yetenek kestiriminde bu değişkenlerin etkisinin araştırıldığı çalışmalara da rastlanmaktadır. Birey

değişkenlerinin modellendiği ve DMF'nin AMTM ile incelendiği çalışmalar da mevcuttur. Ayrıca doğrusal lojistik test modelinin madde parametre kayması veya maddelerinin sahip olduğu bilişsel süreçlerin incelenmesinde kullanıldığı çalışmalara da rastlanmıştır. AMTM'nin henüz sadece doğrusal testlerde uygulandığı gözlenmiş ve herhangi bir BBT uygulamasına rastlanmamıştır.

BBT'ye yönelik çalışmaların ise genellikle madde seçim algoritmalarının, madde kullanım sıklığı kontrol yöntemlerinin, sonlandırma kurallarının, madde havuzu genişliğinin kestirim kesinliği üzerindeki etkisini inceledikleri görülmektedir. Ayrıca az sayıda uyarılma türünün karşılaştırıldığı çalışmalara ve görsel testlerin yanında işitsel test olarak da uygulanabileceğini belirten çalışmalara rastlanmıştır.



## Bölüm 3

### Yöntem

Bu bölümde araştırmanın türüne, çalışma grubuna, verilerin analizine ve değerlendirme kriterlerine yer verilmiştir.

#### Araştırmanın Türü

Bu araştırmada, açıklayıcı madde tepki modellerinin bilgisayar ortamında bireye uyarlanmış testlerde kullanımının bazı parametreler üzerindeki etkisi geleneksel madde tepki modelleri ile karşılaştırılarak ortaya konmaya çalışılmıştır. Madde ve birey değişkenlerinin parametre kestirimleri üzerindeki etkisinin belirlenebilmesi açısından geliştirilen madde havuzu kullanılarak farklı koşullar altında BBT simülasyonları (post-hoc) yapılmış ve bulguların mevcut literatüre katkıları tartışılmıştır. Mevcut uygulamaların dışında yeni bir tekniğin kullanımının önünü açacağı düşünülen çalışmada, yöntemlerden elde edilecek olası farklı sonuçların karşılaştırılması ve betimlenmesi amaçlandığından araştırmanın betimsel araştırma kapsamında olduğu düşünülmektedir. Betimsel araştırmalar belirli bir durumu veya varlığı mümkün olan en detaylı ve tüm yönleri ile betimlemeye, açıklamaya çalışan araştırmalardır (Fraenkel, Wallen ve Hyun, 2012). Ayrıca, simülasyonlarla üretilen verilerle yapılan analizlere ve post-hoc simülasyona dayalı bilgisayar uyarlamalı test sonuçlarının tartışıldığı bu araştırmanın simülasyon araştırması olduğu söylenebilir. Belirli değişkenlerin incelenmesinde zaman zaman kullanılan simülasyonlar gerçek verilerde de sıklıkla karşılaşılan birtakım durumların araştırmacılar tarafından gözlemlenmesine imkân tanır. Ayrıca bu simülasyonlarda elde edilen sonuçlar, gerçek verilerle yürütülebilecek çalışmalara hipotetik alt yapı sunarlar (Fraenkel ve diğerleri, 2012).

#### Çalışma Grubu

**Simülasyona dayalı veri üretim süreci.** Bu araştırmada farklı koşulları bir arada sağlayan simülasyon verileri kullanılmıştır. Çalışmada simülasyon verisi kullanılmasının temel nedeni simülasyon çalışmalarında araştırmanın konusu olan değişkenlerin dışındaki faktörlerin kontrol altında tutulabilmesidir. Ayrıca katılımcılara ve maddeye ilişkin eşdeğişkenlerin madde ve yetenek parametreleri üzerinde tek başına veya birlikte etkili olduğu, çok sayıda madde içeren gerçek

uygulamalara dayalı verilere ulaşılamamasıdır. Birden fazla uygulamaya ait gerçek maddelerden madde havuzu oluşturmak makul bulunmamıştır. Çünkü farklı uygulamalardan elde edilen puanların eşitlenmesi gerekmektedir. Ayrıca bir uygulamada DMF gösteren maddeler, tüm uygulamalar birleştirildiğinde DMF göstermeyebilir.

Madde havuzu genişliği başka bir ifade ile madde havuzunda kaç maddenin olacağı madde karakteristiğine, testin güvenliğine, sınavın doğasına (sonuçlarının katılımcıların geleceğinde önemli rolü olan sınavlar vb.) ve kapsamı beklenen içeriğin genişliği gibi birçok faktöre bağlı olarak belirlenir. Genel olarak madde havuzunun, bir BBT uygulamasındaki ortalama test uzunluğunun 8-12 katı kadar maddeden oluşması gerektiği belirtilmektedir. Örneğin ortalama test uzunluğunun 25 olduğu bir BBT uygulamasında madde havuzunun 200-300 madde içermesi beklenmektedir (Ayala, 2009). Bu araştırma kapsamında her biri 100 maddeden oluşturulan 10 farklı madde havuzu üretilmiştir. Üretilen madde havuzlarının her birinin kalibrasyonları dört farklı model ile yapılmış ve elde edilen parametreler ile 10 döngü olacak şekilde post-hoc simülasyonları gerçekleştirilmiştir. R platformunda yapılan açıklayıcı madde tepki modellerine ilişkin analizlerde teknik yetersizliklerle karşılaşılmasından (RAM sorunu), bu analizlerin ve BBT simülasyonlarının oldukça fazla zaman almasından dolayı madde havuzları 100 madde ile sınırlandırılmıştır.

Bazı önemli avantajlarının ortaya çıkabilmesi ve altta yatan algoritmaların etkili çalışabilmesi açısından BBT madde havuzlarının farklı yetenek düzeylerinin her biri için yeterli sayıda nitelikli madde barındırması gerekmektedir (Wainer ve Dorans, 2000). Tüm farklı yetenek düzeylerini kapsamasının yanında bu araştırmanın önemli noktalarından bir olan Rasch model ve açıklayıcı madde tepki modellerinin analiz sonuçlarının karşılaştırılabilmesi açısından madde ve birey değişkenlerinin etkili olduğu madde parametrelerinin üretilmesine ihtiyaç duyulmuştur. Bu doğrultuda, her bir madde havuzunda Rasch modele göre ikili puanlanan 100 maddeye ait geçici güçlük parametreleri üretilmiştir. Üretilen bu madde havuzlarının tümünde ilk 20 maddenin birey ve madde eşdeğişkenlerinin kategorilerine göre farklı değerler alması sağlanmıştır. Bu aşamada izlenen yol Tablo 2'de sunulmuştur. Örneğin tüm soru bankalarında E alt grubunda olan katılımcılar için ilk beş maddenin güçlük parametresi ortalaması -1 iken, K alt grubundakiler için bu maddelerde güçlük ortalaması 1'dir. Bu maddelerde E

grubunda olanların daha az zorlanması, K alt grubundakilerin ise daha dezavantajlı olması amaçlanmıştır. Yine 16-20 maddelerde, bu maddelerin x formunu alan katılımcılar için madde güçlük ortalaması -1, y formunu alanlar için 0 ve z formunu alan katılımcılar için birdir. Bu maddelerin x formunu alan katılımcılar için avantaj sağlaması planlanmıştır. İlk 20 madde dışında kalan maddelerin güçlük parametreleri ise normal dağılım gösterecek ve ortalaması 0 olacak şekilde üretilmiştir. Tüm madde güçlükleri üretilen yetenek puanlarını kapsayacak şekilde (-2, 2) aralığında üretilmiştir.

Tablo 2

*Madde Havuzu Üretim Koşulları*

Eşdeğişken	Alt Grup	Madde Numarası	Dağılım	a	Ortalama b	En Küçük b	En Büyük b	c
1. Birey	E	1-5	Normal	1.00	-1.00	-2.00	.00	.00
	K	1-5	Normal	1.00	1.00	.00	2.00	.00
2. Birey	A	6-10	Normal	1.00	-1.00	-2.00	.00	.00
	B	6-10	Normal	1.00	.00	-1.00	1.00	.00
1. Madde	C	6-10	Normal	1.00	1.00	1.00	2.00	.00
	i	11-15	Normal	1.00	-1.00	-2.00	.00	.00
2. Madde	j	11-15	Normal	1.00	1.00	.00	2.00	.00
	x	16-20	Normal	1.00	-1.00	-2.00	.00	.00
	y	16-20	Normal	1.00	.00	-1.00	1.00	.00
Eşdeğişken Etkisiz	z	16-20	Normal	1.00	1.00	.00	2.00	.00
		21-100	Normal	1.00	.00	-2.00	2.00	.00

Tablo 3

*Alt Gruplar Bazında Katılımcı Sayısı*

Birey Eşdeğişkenleri		E			K			Satır Toplam
Madde Eşdeğişkenleri		A	B	C	A	B	C	
i	x	40	40	40	40	40	40	240
	y	40	40	40	40	40	40	240
	z	40	40	40	40	40	40	240
i Toplam		120	120	120	120	120	120	
j	x	40	40	40	40	40	40	240
	y	40	40	40	40	40	40	240
	z	40	40	40	40	40	40	240
j Toplam		120	120	120	120	120	120	
Sütun Toplam		240	240	240	240	240	240	1440

Geçici madde parametrelerinin üretilmesinin ardından tüm madde havuzlarında katılımcı alt gruplarında dengeli bir dağılım sağlanabilmesi için (-2, 2) aralığında normal dağılım gösteren  $N(0,1)$  1440 geçici yetenek parametresi üretilmiştir. Her madde havuzunda 1440 katılımcı olmasının sebebi, katılımcıların 36 alt gruba (2X3X2X3) dengeli dağılımını sağlamak ve faktör analizi için madde sayısına oranla uygun örneklem büyüklüğüne ulaşılmasıdır. Tablo 3'te hangi alt grupta kaç katılımcı olduğu görülmektedir. Alt grupların analizlerde uygun bir biçimde temsil edilebilmesi için her bir madde havuzu için üretilen yetenek parametrelerin yarısına rastgele olacak biçimde E kategorisi (720), diğer yarısına K kategorisi (720) atanmıştır. Ardından E ve K kategorilerindeki parametrelerin her biri üçe bölünmüş ve her bir parçaya A (240, 240), B (240, 240) ve C (240, 240) kategorileri atanmıştır. Daha sonra EA, EB, EC, KA, KB ve KC alt gruplarını oluşturan yetenek parametreleri i ve j alt gruplarını oluşturacak şekilde yeniden rastgele ikiye ayrılmıştır. Son olarak her biri 120 puandan oluşan 12 alt grubun (EA<sub>i</sub>, EA<sub>j</sub>, EB<sub>i</sub>, EB<sub>j</sub>, EC<sub>i</sub>, EC<sub>j</sub>, KA<sub>i</sub>, KA<sub>j</sub>, KB<sub>i</sub>, KB<sub>j</sub>, KC<sub>i</sub>, KC<sub>j</sub>) her biri rastgele üç eşit bölüme ayrılmıştır. Üç parçaya x, y ve z kategorileri tanımlanmış ve bu şekilde her birinde 40 katılımcının olduğu 36 (EA<sub>ix</sub>, KA<sub>jz</sub> vb.) alt grup meydana getirilmiştir. Bu şekilde alt grupların eşit oranda temsil edilmesi sağlanmıştır.

Geçici parametreler üretildikten sonra katılımcıların hangi alt gruba ait olduğu göz önünde bulundurularak 10 madde havuzu için cevap örüntüleri üretilmiştir. Örneğin, EA<sub>ix</sub> alt grubundaki katılımcıların cevap örüntüleri, madde havuzlarındaki ilk beş maddenin güçlük düzeyi E, sonraki beş madde A, 11-15 arasındaki maddeler i ve 16-20 arasındaki maddeler x madde güçlüğü kriterlerine uygun üretilmiş madde parametreleri ile bu katılımcılara ait geçici yetenek puanlarından elde edilmiştir. Geçici parametrelerin üretilmesinde belirli aralıklarda normal dağılıma uygun veri üretmeyi sağlayan R platformundaki "msm" (Jackson, 2011) paketi; cevap örüntülerinin üretilmesinde ise "catR" (Magis ve Barrada, 2017) paketi kullanılmıştır.

Bu araştırmada simülasyona dayalı veri üretildiği için birey ve madde değişkenleri görgül (ampirik) değişkenler olarak atanmamıştır. Bu değişkenler gerçek veri olmadığından gerçek görgül değişken isimleri ile adlandırılmamıştır. Gerçek veriler kullanılarak birinci birey değişkeninin cinsiyet, ikinci birey değişkeninin sosyo-ekonomik düzey, birinci madde değişkeninin madde formatı ve ikinci madde değişkeninin maddenin uygulanma sırası olduğu çalışmalar

yürütülebilir. Böyle bir çalışmada örneğin EA grubu, alt sosyo-ekonomik düzeydeki erkek katılımcıları temsil edebilir. Benzer bir biçimde kcjz grubu bir grup maddeyi testin sonlarında açık uçlu biçimde alan yüksek sosyo-ekonomik düzeye sahip kadın katılımcıları tanımlamak için kullanılabilir. Bu durum yapılacak çalışmanın amacı ve değişkenlerin kategori sayısına bağlı olarak çeşitlendirilebilir.

## **Verilerin Analizi**

Verilerin analizi aşamasında, BBT simülasyonları gerçekleştirilmeden önce Rasch modelin varsayımları test edilmiştir. Ardından madde havuzlarında istenen maddelerde DMF varlığını tespit etmek için DMF analizleri gerçekleştirilmiştir. Daha sonra Rasch modeli ve açıklayıcı madde tepki modelleri ile madde kalibrasyonu gerçekleştirilmiş madde ve önsel yetenek parametreleri elde edilmiştir. Son olarak BBT simülasyonları gerçekleştirilmiştir.

Bir parametrelili lojistik modelin özel bir hali olan Rasch modelinin doğru kestirimler yapabilmesi için verilerin tek boyutluluk, yerel bağımsızlık, model-veri uyumu ve testin hız testi olmaması varsayımlarını karşılaması gerekmektedir. Madde tepki kuramında tek boyutluluk tüm maddelerin aynı gizil özelliği (latent trait) ölçmesi anlamına gelmektedir. Başka bir ifade ile verilen cevaplardaki tesadüfi olmayan tüm varyans madde güçlüğü ve bireyin yeteneğinin tek bir boyutu ile açıklanabiliyorsa verilerin tek boyutlu olduğu söylenebilir. Sonuçlarının tek bir puan olarak raporlandığı tek boyutlu testlerde tüm maddelerin ortak bir yapıyı paylaştıkları varsayımı vardır (DeMars, 2010). Genel olarak tek boyutluluk maddeler arası korelasyon matrisine ait özdeğerler yardımı ile test edilir. Ancak özdeğerler sadece normal dağılılan sürekli verilerle gerçekleşen faktör analizinde faydalanılan bir yöntemdir. Cevapların ikili puanlandığı (dichotous) ve ölçülen gizil özelliğin (yetenek) sürekli olduğu durumlarda faktör yapısı, tetrakorik korelasyona dayalı faktör çıkarma yöntemleri yardımı ile belirlenir. Bu araştırmada, her biri 10 maddeden oluşan 10 soru havuzunun tek boyutluluğu tetrakorik korelasyona dayalı temel bileşenler analizi ile incelenmiştir. Açıklayıcı faktör analizi sonuçlarına ilişkin detaylar Tablo 23'te görülmektedir. Temel bileşenler analizi yapılmadan önce Bartlett küresellik testi yapılmış ve örneklem uygunluğunu için de Kaiser-Meyer-Olkin (KMO) istatistiği hesaplanmıştır. Bartlett testi değişkenlere ait korelasyon matrisinin birim matrise denk olup olmadığını sınavan bir hipotez testidir. Korelasyon

matrisinin birim matrisininden manidar bir biçimde farklılaşması değişkenler arasında ilişkinin 0'a eşit olmadığı anlamına gelir. Bu durumda testin manidar olması değişkenlerin faktörleşebileceği şeklinde yorumlanır. KMO istatistiği ise değişkenler arası korelasyon karelerinin, kısmi korelasyon karelerine oranının ölçüsüdür. Bu istatistik 0 ve 1 arasında değerler alır. KMO istatistiğinin 1'e yakın bir değerde olması faktör analizinde ayrı ve güvenilir faktörlerin elde edilebileceğinin göstergesidir. İstatistiğin kabul edilebilir en küçük değerinin 0.5 olduğu belirtilmiştir. Bu değerden daha küçük bir istatistik elde edildiğinde veri sayısının artırılması veya analiz edilecek değişkenlerin tekrar belirlenmesi gerekebilir. 0.9'dan büyük değerler ise örneklem uygunluğunun mükemmel düzeyde olduğunu göstermektedir (Field, 2009). Bartlett küresellik testinin tüm madde havuzlarında manidar çıkması ve KMO değerinin 0.96-0.98 arasında değerler alması ile örneklemelerin faktör analitik yöntemlerle analize uygun olduğu sonucuna ulaşılmıştır.

Temel bileşenler analizleri sonucunda madde havuzlarındaki varyansın %25-%28'inin tek bir temel bileşen tarafından açıklanabildiği belirlenmiştir. Bu durumda, kalibrasyonların kabul edilebilir olması için test varyansının en az %20'sinin tek faktör tarafından açıklanmasına ilişkin varsayımın (Reckase, 1979) karşılandığı söylenebilir. Hata kareleri ortalamalarının kökü (RMS) bir madde havuzu (3. Madde havuzu için 0.05) dışında tüm soru havuzlarında 0.04 bulunmuştur. Ayrıca tetrakorik korelasyon matrisinde çapraz değerlere dayalı model uyum indekslerinin 0.96-0.98 arasında değerler aldığı belirlenmiştir. Bu durum madde havuzlarında model uyumunun sağlandığı şeklinde yorumlanmıştır. Ortalama faktör yükünün 0.50 ile 0.52 arasında değerler aldığı gözlenmiştir. Analizlerde altı madde havuzunda birkaç maddede (1. Havuz=1, 3. Havuz=2, 6. Havuz=3, 7. Havuz=1, 8. Havuz=2, 10. Havuz=1) 0.4'ten küçük faktör yüküne rastlanmıştır. Ancak havuz genişliğine oranla bu madde sayılarının az olmasından dolayı madde havuzlarından herhangi bir madde çıkarmaya ihtiyaç duyulmamıştır. Analiz sonucunda verilerin tek boyutluluk şartını sağladığı belirlenmiştir. Temel bileşenler analizinden elde edilen tüm faktör yükleri Ek A'da sunulmuştur. Temel bileşenler analizinde birinci faktörün diğer faktörlere oranla baskın olmasından dolayı Tablo 4 ve EK A'da sadece bu faktöre ilişkin yükler sunulmuştur. Analizler bir R paketi olan "psych" (Revelle, 2018) ile gerçekleştirilmiştir.

Tablo 4

*Madde Havuzuna İlişkin Faktör Yükleri*

Madde Havuzu	Madde Sayısı	RMS	Uyum İndeksi	Açıklanan Varyans (%)	KMO	Bartlet Test ( $\chi^2_{4950}$ )	Ortalama Yük	En Küçük Yük	En Büyük Yük
Havuz 1	100	.04	.98	27	.92	2898.32*	.52	.36	.60
Havuz 2	100	.04	.98	27	.92	2881.89*	.52	.40	.61
Havuz 3	100	.05	.96	25	.79	3329.16*	.50	.35	.58
Havuz 4	100	.04	.98	28	.94	2853.97*	.52	.44	.59
Havuz 5	100	.04	.98	26	.90	2677.65*	.51	.40	.58
Havuz 6	100	.04	.97	26	.87	2924.12*	.51	.37	.58
Havuz 7	100	.04	.97	26	.90	2776.47*	.50	.36	.56
Havuz 8	100	.04	.97	26	.91	2792.22*	.51	.38	.60
Havuz 9	100	.04	.98	27	.92	287.50*	.52	.43	.59
Havuz 10	100	.04	.98	27	.92	2855.90*	.52	.37	.59

\*p&lt;.05

Diğer tek boyutlu madde tepki kuramı modellerinde olduğu gibi Rasch modelde de yerel bağımsızlık varsayımının karşılanması gerekmektedir. Yerel bağımsızlık, ölçmeye konu olan özellik (yetenek) sabitlendiğinde (kontrol altında tutulduğunda) iki farklı maddeye cevap vermenin bağımsız olduğu anlamına gelir. Başka bir ifade ile cevaplar arasındaki ilişki sadece bireyde var olan örtük özelliğe (yeteneğe) atfedilebilir. Yerel bağımsızlığın test edilmesinde en yaygın kullanılan yöntemlerden biri Yen'in (1984) geliştirdiği Q3 yöntemidir. Bu yöntem madde artık değerleri arasındaki korelasyona dayalı basit ve etkili bir yöntemdir. Madde artık değerleri arasındaki korelasyonun 0.2'den büyük olmaması yerel bağımsızlık şartının sağlandığı şeklinde yorumlanır (Finch ve French, 2015). Yerel bağımsızlık R paketi olan "sirt" (Robitzsch, 2019) kullanılarak Yen'in Q3 yöntemi ile test edilmiştir. Hesaplamalar sonucunda tüm madde havuzlarında madde artık değerler çiftleri arasındaki korelasyonun -0.14 ile 0.12 arasında değiştiği gözlenmiştir. Hiçbir madde havuzunda 0.2'den büyük korelasyon katsayısına rastlanmaması yerel bağımsızlık varsayımının karşılandığı şeklinde yorumlanmıştır.

Bazı katılımcılar, test süresinin yetmemesinden dolayı testin sonlarındaki maddeleri yanıtlayamayabilir. Bu durumda, mevcut madde tepki kuramı modelleri katılımcının bu maddelere verdiği tepkiyi tanımlayamaz. Çünkü tek boyutlu MTK modellerinde sadece var olan cevaplar üzerinden kestirimler yapılır. Cevap verilip verilmeyeceğine odaklanılmaz. Bundan dolayı MTK modellerine ilişkin analizlerde

katılımcıların tüm test maddelerine erişebilecekleri kadar sürelerinin olması gerekmektedir. Başka bir ifade ile testin hız testi olmaması gerektiği vurgulanmaktadır (Lord, 1980). Testin hız testi olmadığı genellikle testin tamamını bitiren katılımcı oranı, testin %75'ini bitiren katılımcı oranı ve katılımcıların %80'nin cevapladığı madde sayılarının karşılaştırılması ile belirlenir. Yaklaşık tüm katılımcıların maddelerin tamamına yakınına cevap verdikleri durumda, hızın test performansı üzerinde etkili bir faktör olmadığı sonucuna varılır (Hambleton ve diğerleri, 1991). Bu araştırmada tüm madde havuzlarında katılımcıların tamamı için cevap örüntüleri eksiksiz bir şekilde üretildiğinden testlerin hız testi olmadığı söylenebilir.

**DMF analizleri.** Veri üretim sürecinde tüm madde havuzlarının ilk 20 maddesinin güçlük parametreleri üzerinde etkili olan birey ve madde eşdeğişkenleri geliştirilmiştir. Bu eşdeğişkenlerin etkisinin ilk 20 maddeye ait cevap örüntülerine yansıyor yansımadağı DMF analizleri yardımı ile incelenmiştir. İki kategorili değişkenler için DMF analiz yöntemi olarak (birinci birey ve madde eşdeğişkenleri) madde tepki kuramına dayalı olan Lord'un Ki-kare yöntemi (Lord, 1980) kullanılmıştır. Üç kategorili değişkenler için (ikinci katılımcı ve madde eşdeğişkenleri) Lord'un genelleştirilmiş Ki-kare yönteminden (Kim ve diğerleri, 1995) yararlanılmıştır. İlk 20 maddenin DMF analizlerine ilişkin bulgular Tablo 5'te sunulmuştur. Tablo 5 incelendiğinde, tüm madde havuzlarında ilk beş maddenin birinci birey, 6-10 arasındaki maddelerinin ikinci birey, 11-15 arasındaki maddelerin birinci madde ve 16-20 arasındaki maddelerin ikinci madde eşdeğişkenine göre yüksek düzeyde (C düzeyinde) DMF gösterdiği görülmektedir. Bu durumda veri üretim sürecinde hedeflenen maddelerde istenen etkinin oluşturulduğu söylenebilir. Tüm madde havuzları için yapılan DMF analizlerine ilişkin bulgular Ek B, Ek C, Ek D ve Ek E'de sunulmuştur. Hem Lord'un ki-kare yöntemi hem de Lord'un genelleştirilmiş ki-kare yöntemi ile DMF analizleri bir R paketi olan "difR" (Magis, Béland, Tuerlinckx ve Boeck, 2010) ile gerçekleştirilmiştir.

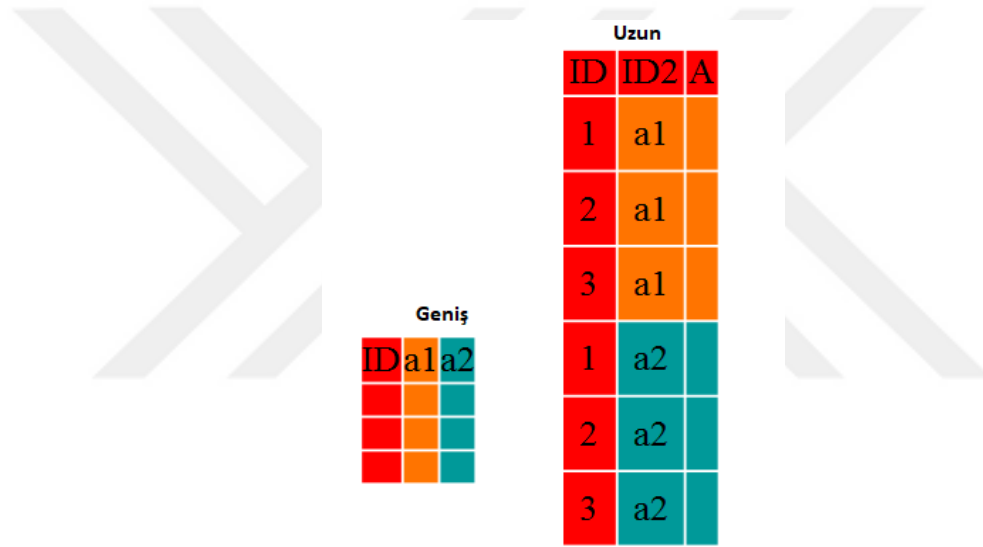
Tablo 5

## Madde Havuzlarındaki İlk 20 Madde İçin DMF Analizi Sonuçları

Değişken	Madde	Lord'un $\chi^2$ İstatistiği									
		1. Madde Havuzu	2. Madde Havuzu	3. Madde Havuzu	4. Madde Havuzu	5. Madde Havuzu	6. Madde Havuzu	7. Madde Havuzu	8. Madde Havuzu	9. Madde Havuzu	10. Madde Havuzu
Birinci Birey Değişkeni (E, K)	M1	438.74*	204.05*	110.69*	311.03	55.31*	132.67*	220.90*	229.55*	360.97*	98.02*
	M2	137.50*	44.95*	39.61*	169.08	15.05*	239.07*	223.53*	47.61*	97.45*	153.80*
	M3	194.96*	197.22*	307.67*	131.47*	18.98*	261.93*	211.66*	289.12*	254.33*	117.69*
	M4	76.77*	149.48*	566.63*	261.26*	165.76*	141.47*	135.16*	53.38*	259.81*	204.45*
	M5	177.04*	111.95*	358.81*	249.67*	227.18*	388.72*	322.58*	126.97**	185.81*	334.17*
	M6	156.37*	233.23*	310.01*	219.72*	65.50*	305.33*	88.50*	329.40*	90.29*	122.73*
İkinci Birey Değişkeni (A, B, C)	M7	49.81*	177.49*	227.26*	213.44*	259.78*	296.38*	113.30*	229.14*	167.10*	193.43*
	M8	193.30*	133.77*	234.56*	60.97*	44.81*	136.08*	217.56*	383.88*	148.75*	257.31*
	M9	215.56*	399.94*	100.10*	137.18*	32.89*	127.80*	333.11*	118.59*	110.06*	144.65*
	M10	195.93*	208.12*	265.68*	156.02*	42.27*	100.85*	134.08*	106.04*	118.29*	222.39*
Birinci Madde Değişkeni (i, j)	M11	313.86*	87.06*	41.18*	263.64*	174.98*	246.24*	84.45*	98.34*	388.98*	114.60*
	M12	6.90*	223.66*	209.58*	174.10*	192.66*	358.91*	164.53*	97.15*	122.83*	236.66*
	M13	408.28*	282.72*	438.49*	179.02*	389.54*	17.60*	195.04*	158.35*	109.39*	24.86*
	M14	219.12*	305.39*	64.94*	152.20*	261.97*	151.42*	415.50*	344.36*	43.80*	42.05*
	M15	140.82*	359.38*	251.17*	131.94*	87.54*	376.11*	248.37*	343.09*	69.77*	260.12*
İkinci Madde Değişkeni (x, y, z)	M16	141.92*	297.94*	43.95*	171.51*	168.85*	183.42*	315.52*	217.64*	59.70*	298.42*
	M17	176.84*	80.24*	130.16*	142.13*	150.30*	259.11*	111.87*	115.28*	130.86*	214.58*
	M18	132.54*	69.53*	37.29*	64.25*	38.83*	432.33*	112.68*	229.66*	281.07*	147.85*
	M19	99.63*	76.58*	131.28*	109.39*	70.83*	70.73*	287.03*	123.30*	208.25*	75.66*
	M20	257.90*	198.41*	81.57*	243.73*	124.59*	153.87*	79.28*	281.33*	74.09*	361.28*

\*C Düzeyinde DMF

**Madde havuzlarının kalibrasyonuna ilişkin veri analizi.** Verilerin üretilmesi, Rasch model varsayımlarının test edilmesi ve DMF analizlerinin ardından soru bankalarında post-hoc simülasyonlarda kullanılacak olan madde parametrelerinin ve önsel yetenek puanlarının elde edilmesi süreci gerçekleşmiştir. Bu aşamada, Rasch modele, eşvaryanslı (homoscedastic) örtük regresyon modeline, eşvaryanslı doğrusal lojistik test modeline ve eşvaryanslı örtük regresyon doğrusal lojistik test modeline ilişkin analizler R paketi olan “lme4” (Bates ve diğerleri, 2015) ile gerçekleştirilmiştir. Cevap örüntüleri, birey ve madde eşdeğişkenlerinin bulunduğu veri setleri, “lme4” paketinde analizlerin gerçekleştirilebilmesi için öncelikle Şekil 1’de olduğu gibi geniş veri formatından uzun veri formatına dönüştürülmüştür.



Şekil 3. Verilerin geniş ve uzun biçimleri

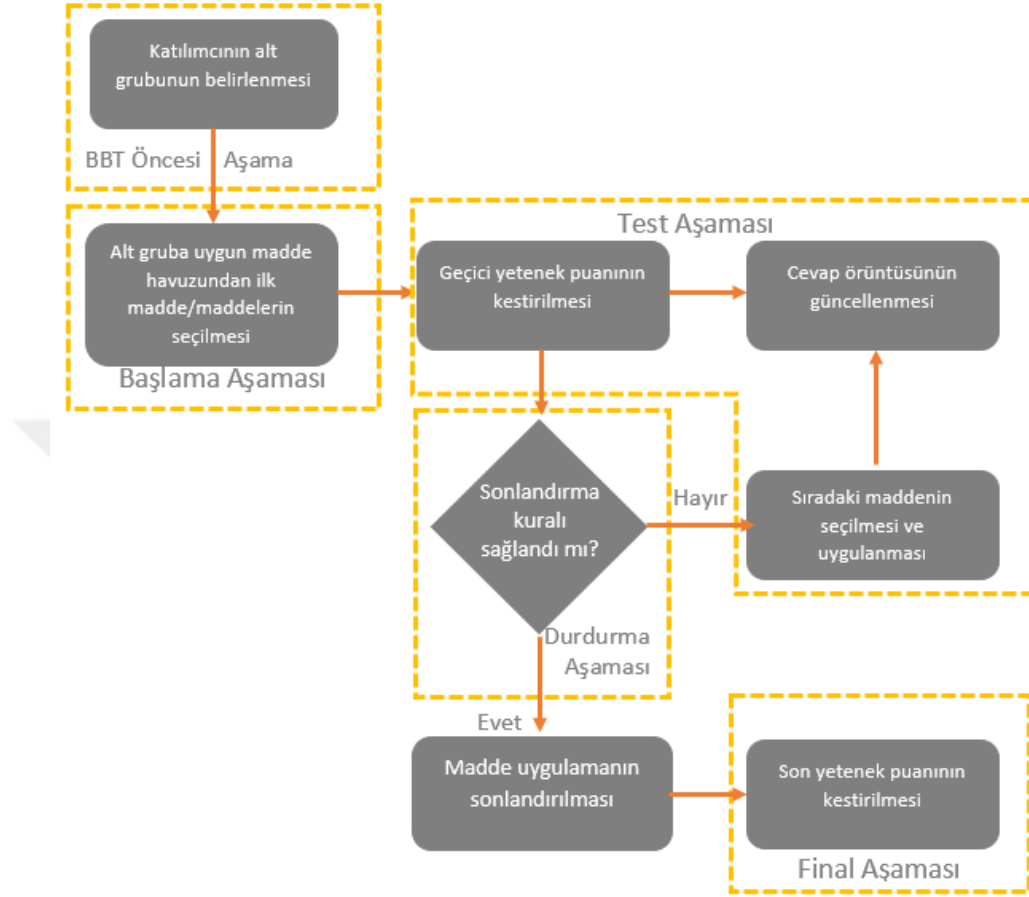
Veriler uygun biçime getirildikten sonra her bir madde havuzu için ayrı ayrı dört model çalıştırılmıştır. Her bir model için Laplace yakınsama algoritmasına dayalı bobyqa optimizasyon kullanılmıştır. Maksimum 50.000 iterasyon ile sınırlandırılan kestirimlerde tüm modellerin başarılı bir biçimde yakınsama (converge) gösterdiği tespit edilmiştir. Daha sonra tüm madde havuzları ve modeller için sabit etkiler (madde güçlükleri) ve rastgele etkiler (yetenek parametreleri) elde edilmiştir. Rasch modelden, 10 madde havuzu için 10 farklı madde güçlük ve yetenek parametresi seti elde edilmiştir. Örtük regresyon analizlerinde her bir madde havuzu için altı farklı (EA, EB, EC, KA, KB, KC) madde güçlük parametresi seti olacak şekilde 60 farklı güçlük parametresi seti ve 10 yetenek parametre seti kestirilmiştir. Doğrusal lojistik test modelinde yine her bir madde havuzu için altı farklı (ix, iy, iz, jx, jy, jz) madde

güçlük parametresi seti ve 10 yetenek parametresi seti elde edilmiştir. Örtük regresyon doğrusal lojistik test modelde her bir madde havuzu için 36 madde güçlüğü seti kestirilmesine karşın bunlardan sadece üç alt grup için kestirilen parametreler post-hoc simülasyonlarında kullanılmıştır. Analizler için gerekli süre ve bulguların raporlanmasındaki güçlükler göz önünde bulundurularak ilk 20 maddenin tamamında en avantajlı olan eaix (en düşük madde güçlük ortalaması, tüm eşdeğişkenler bakımından en avantajlı grup), ilk beş ve 11-15 arasındaki maddelerde avantajlı, 6-10 ve 16-20 arasındaki maddelerde dezavantajlı olan eciz (orta madde güçlük ortalaması) ve ilk 20 maddenin tamamında dezavantajlı olan kcjz (en yüksek madde güçlük ortalaması, tüm eşdeğişkenler bakımından en dezavantajlı grup) alt gruplarına ait madde parametreleri ile BBT simülasyonları gerçekleştirilmiştir.

**BBT simülasyonlarına ilişkin veri analizi süreci.** Madde havuzu kalibrasyon sürecinde madde ve önsel yetenek parametrelerinin elde edilmesinin ardından post-hoc simülasyonu sürecine geçilmiştir. Rasch modelden elde edilen parametrelerin kullanıldığı BBT simülasyonları için “catR” (Magis & Barrada, 2017) paketi kullanılmıştır. AMTM'den elde edilen parametrelerle gerçekleştirilen BBT simülasyonlarında katılımcıların ait oldukları alt gruplara ait madde parametrelerinin kullanılması gerekmektedir. Bu yöntemle post-hoc simülasyonu yapan yazılım bulunamadığından yazar “catR” paketi için yeni fonksiyonlar geliştirmiştir. Geliştirilen yeni fonksiyonların ardından “catR” paketine ait post-hoc simülasyon algoritması, Şekil 4'teki işleyiş şemasında özetlenmiştir.

Şekli 4'te açıklayıcı madde tepki modelleri ile kestirilen önsel yetenek puanları ve madde parametreleri ile gerçekleştirilen post-hoc simülasyonlarının işleyişi görülmektedir. Şemaya göre simülasyon, katılımcının hangi alt gruba ait olduğunun belirlenmesi ile başlamaktadır. Katılımcının ait olduğu alt grup belirlendikten sonra algoritma, alt gruba ait madde parametrelerini ve katılımcının önsel yetenek puanını kullanarak tüm maddeler için maksimum Fisher bilgi fonksiyonu hesaplamaktadır. En fazla bilgi veren üç maddeden birini rastgele seçerek katılımcıya uygulamaktadır. Uygulanan maddenin ardından geçici yetenek puanı hesaplanmakta ve sonlandırma kuralının sağlanıp sağlanmadığı değerlendirilmektedir. Değerlendirmeye bağlı olarak uygulamanın sonlandırılmasına veya cevap örüntüsünün güncellenerek yeni bir maddenin

uygulanmasına karar verilmektedir. Bu süreç tüm katılımcılar için sonlandırma kuralı sağlanana veya madde havuzundaki tüm maddeler kullanılına kadar devam etmektedir.



Şekil 4. AMTM için BBT post-hoc simülasyon şeması

Rasch model için catR, diğer üç model için “catR” paketinin düzenlenmiş bir biçiminin kullanıldığı simülasyonlar her bir madde havuzu için 10 tekrar olacak şekilde iki farklı sonlandırma kuralına göre gerçekleşmiştir. Bu durumda her bir katılımcı için simülasyonlar 100 (10x10) tekrarda gerçekleşmiştir. Tüm döngülerde en fazla bilgi sunan üç maddeden rastgele çekilen bir madde başlangıç maddesi olarak kullanılmıştır. En fazla bilgi sunan maddelerin belirlenmesinde katılımcılara ait önsel yetenek puanları kullanılmıştır. Uygulama başladıktan sonraki maddeler randomesque yöntemi ile en fazla bilgi sunan beş maddeden rasgele bir maddenin seçilmesi şeklinde belirlenmiştir. Hem geçici yetenek puanları hem son yetenek puanının hesaplanmasında beklenen sonsal dağılım yöntemi (BSD, EAP) kullanılmıştır. Sonlandırma kuralı olan kesinlik (precision) için kesme noktası olarak 0.3, minInfo kuralı için 0.24 kullanılmıştır.

## Değerlendirme Kriterleri

Bu araştırmada, Rasch modelden elde edilen madde güçlükleri ve yetenek parametreleri referans değerler kabul edilerek diğer modellerden edilen parametreler bu modelden elde edilenle karşılaştırılmıştır. Madde havuzlarında elde edilen madde güçlükleri ve yetenek puanları aynı maddelere ve bireylere ait tekrarlı ölçümler olarak ele alınmıştır. Bir madde havuzunda her bir madde için dört model ile kestirilmiş madde güçlükleri bulunmaktadır. Bu madde güçlükleri tekrarlı ölçümler olarak düşünülüp tekrarlı ölçümler için varyans analizi ile karşılaştırılmıştır. BBT simülasyonlarında ise tüm madde havuzları ve döngülerden elde edilen bireysel yetenek puanlarının ortalamaları tekrarlı ölçümler olarak ele alınmıştır. Bu tekrarlı ölçümlerin karşılaştırılmasında bağımlı örneklemeler için t-testi ve tekrarlı ANOVA kullanılmıştır. Tekrarlı tek yönlü varyans analizinde manidar bulunan farkın hangi alt gruplar arasında olduğunu belirlemek için ikili karşılaştırma yöntemi olarak LSD kullanılmıştır. SPSS ile gerçekleştirilen tekrarlı tek yönlü varyans analizinde değişkenlerin grup-İçi etki (within subject effect) olarak tanımlanması gerekmektedir. Gruplar arası etkinin (between subject effect) hesaplanmadığı bu analizde tek yönlü varyans analizinde sık kullanılan Tukey, Scheffe ve benzeri ikili karşılaştırma yöntemleri kullanılamamaktadır (Pallant, 2016).

Ortalamalara ilişkin hipotez testlerinin (tekrarlı ANOVA ve bağımlı örneklemeler için t-testi) yanında, tüm madde havuzundaki madde güçlükleri, önsel yetenek puanları ve BBT simülasyonlarında kestirilen yetenek parametreleri hata kare ortalamasının kökü (RMSE) yardımı ile karşılaştırılmıştır. Karşılaştırmalarda Pearson momentler çarpımı korelasyon katsayısı (r) da kullanılmıştır. Madde güçlüklerine ilişkin RMSE değerleri Eşitlik 11, yetenek puanlarına ilişkin RMSE değerleri Eşitlik 12 ve yanlılık değeri Eşitlik 13'te gösterildiği şekilde hesaplanmıştır.

$$RMSE(b) = \sqrt{N^{-1} \sum_{i=1}^N (b_i^* - b_{Rasch})^2} \quad (11)$$

$$RMSE(\theta) = \sqrt{N^{-1} \sum_{i=1}^N (\theta_i^* - \theta_{Rasch})^2} \quad (12)$$

$$yanlılık(\theta) = N^{-1} \sum_{i=1}^N (\theta_i^* - \theta_{Rasch}) \quad (13)$$

Eşitlik 11’de gösterilen RMSE hesaplama yöntemine ait işlem basamakları şu şekildedir: öncelikle her bir maddenin ilgili modelde (örneğin örtük regresyon) kestirilen güçlük parametresinden, Rasch modelden kestirilen parametre çıkartılır ve karesi alınır. Ardından, elde edilen tüm kareler toplanır ve madde sayısına bölünür. Son olarak bu değer karekökü alınarak RMSE değerine ulaşılır. Benzer bir durum yetenek puanları için de geçerlidir: tüm katılımcıların ilgili modelde kestirilen yetenek puanından Rasch modelden elde edilen puanı çıkartılarak karesi alınır. Ardından karelerin toplamı katılımcı sayısına bölünür ve karekökü alınır. Yanlılık RMSE’den farklı bir istatistiktir. Yanlılık farkların toplamının kişi sayısına oranıdır. RMSE ise farkların karelerinin kişilere oranının kareköküdür. Bu durumda RMSE sürekli pozitif sonuçlar verirken yanlılık pozitif veya negatif sonuçlar verebilir. Post-hoc simülasyonlarından elde edilen yetenek puanlarına ilişkin bulguların değerlendirmesinde bunların dışında kriter olarak ortalama test uzunluğu, ortalama madde kullanım sıklığı (exposure rate), başarı ile sonlandırılan test oranı ve ortalama test örtüşme (overlap) oranı kullanılmıştır.

Bu araştırmada bağımlı örneklemeler için t-testine ilişkin etki büyüklüğü olarak Cohen’in d indeksi ile hesaplanmışken, tekrarlı ANOVA için etki büyüklüğü olarak kısmi eta-kare hesaplanmıştır. Tekrarlı varyans analizinde etki büyüklüğü olarak genellikle genelleştirilmiş eta kare ve kısmi eta kare önerilmektedir. Tek yönlü tekrarlı varyans analizinde genelleştirilmiş eta kare ve kısmi eta kare eşit etki büyüklükleri kestirmektedir. Buna karşın bağımlı örneklemeler için t-testinde genellikle Cohen’in d istatistiği kullanılmaktadır (Bakeman, 2005; Cohen, 1988). Cohen’in d etki büyüklüğü, evren ortalamaları arasındaki farkın (ortak) standart sapmaya oranıdır. Başka bir deyişle ortalamalar arasındaki farkın standart sapma cinsinden karşılığıdır ve Eşitlik 14’te gösterildiği şekilde hesaplanır. Cohen’in d istatistiği negatif veya pozitif değerler alabilir. d indeksi 0.2-0.49 aralığında olduğu durumlarda etki büyüklüğünün küçük, 0.5-0.79 aralığında olduğu durumlarda orta, 0.8’e eşit veya daha büyük olduğu durumlarda büyük olduğu kabul edilir (Cohen, 1988).

$$d = \frac{\mu_a - \mu_b}{\sigma} \quad (14)$$

Kısmi eta-kare etki büyüklüğü varyans analizinde elde edilen F değerine, grup sayısına ve örneklem büyüklüğüne bağlı bir istatistiktir. Eşitlik 15'te kısmi eta-karenin nasıl hesaplandığı görülmektedir. Eşitlikte F, varyans analizinde elde edilen F değerini, k grup sayısını ve N örneklem büyüklüğünü temsil etmektedir. Kısmi eta-kareden hesaplanan f değeri ile etki büyüklüklerinin düzeyine karar verilir. Bu değer Eşitlik 16'da gösterildiği gibi hesaplanmaktadır. f değerinin 0.1-0.24 aralığında olması etki büyüklüğünün küçük, 0.25-0.39 aralığında olması orta ve 0.4'e eşit veya daha büyük olması durumunda etki büyüklüğünün büyük olduğu anlamına gelir (Cohen, 1988).

$$\eta^2 = \frac{[F(k - 1)]}{[F(k - 1) + (N - k)]} \quad (15)$$

$$f = \sqrt{\left[\frac{\eta^2}{(1 - \eta^2)}\right]} \quad (16)$$

## Bölüm 4

### Bulgular ve Yorumlar

Bu bölümde araştırmanın 10 alt problemine ilişkin bulgular, tablolar ve şekiller yardımı ile sunulmuş ve açıklanmıştır.

#### Birinci Alt Probleme İlişkin Bulgular ve Yorumlar

Çalışmanın birinci alt problemi olan “Örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinden elde edilen madde güçlük parametreleri Rasch modeli ile elde edilen madde güçlük parametrelerinden farklılaşmakta mıdır?” sorusu betimsel istatistikler, tekrarlı ANOVA ve RMSE değerleri yardımı ile incelenmiş ve bulgular raporlanmıştır. Araştırma kapsamında geliştirilen 10 adet soru havuzunun her biri ile ayrı ayrı Rasch model, örtük regresyon analizi, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modeli ile madde parametreleri elde edilmiş ve model bazında madde parametrelerine ilişkin betimsel istatistiklere ulaşılmıştır. Elde edilen istatistikler Tablo 6’da görüldüğü gibidir.

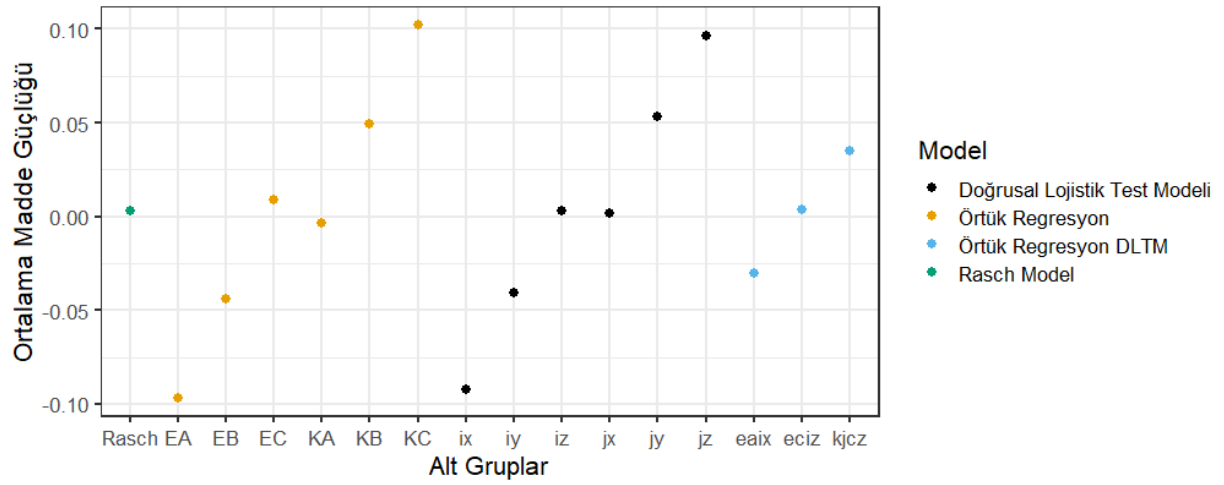
Tablo 6

#### *Madde Güçlük Parametrelerine (b) İlişkin Betimsel İstatistikler*

Model	Alt Grup	Havuz Sayısı	Madde Sayısı	$\bar{x}_b$	$\sigma_b$	En Küçük b	En Büyük b
Rasch Model		10	100	.00	.80	-2.12	1.97
Örtük Regresyon	EA	10	100	-.10	.87	-2.13	1.98
	EB	10	100	-.04	.84	-2.13	1.98
	EC	10	100	.01	.87	-2.13	1.98
	KA	10	100	.00	.87	-2.13	1.98
	KB	10	100	.05	.84	-2.13	1.98
	KC	10	100	.10	.86	-2.13	1.98
Doğrusal Lojistik Test Modeli	ix	10	100	-.09	.85	-2.13	1.98
	iy	10	100	-.04	.84	-2.13	1.98
	iz	10	100	.00	.87	-2.13	1.98
	jx	10	100	.00	.87	-2.13	1.98
	jy	10	100	.05	.84	-2.13	1.98
Örtük Regresyon Doğrusal Lojistik Test Modeli	jz	10	100	.10	.87	-2.13	1.98
	eaix	10	100	-.03	.93	-4.29	3.25
	eciz	10	100	.00	.92	-2.63	2.13
	kcjz	10	100	.03	.95	-2.53	2.80

Tablo 6 incelendiğinde tüm soru havuzlarından Rasch modeli ile elde edilen b parametrelerinin ortalamasının 0, standart sapmasının 0.8 olduğu belirlenmiştir. Parametrelerin -2.12 ile 1.97 arasında değiştiği ve standart hatanın 0.03 olduğu sonucuna ulaşılmıştır. Bu durumda madde havuzu geliştirme kriterlerine uygun madde cevap örüntülerinin elde edildiği söylenebilir. Örtük regresyon analizi kapsamında katılımcı alt gruplarının her biri için elde edilen madde parametrelerinin ortalamasının -0.1 ile 0.1 arasında değiştiği görülmektedir. Alt gruplar için elde edilen ortalamalardan elde edilen genel ortalamanın 0.003 olduğu ve Rasch modele çok yakın kestirildiği sonucuna ulaşılmıştır. Birinci birey değişkeni bakımından E grubunda, ikinci birey değişkeni bakımından A grubunda bulunan katılımcılar için elde edilen ortalama madde güçlüğü -0.1 olması, bu gruptaki bireyler için maddelerin ortalama olarak 0.1 logit daha kolay olduğu anlamına gelmektedir. Buna karşın KC grubundaki bireyler için ise madde güçlüklerinin ortalamadan 0.1 logit yüksek olduğu, başka bir ifade ile bu gruptaki bireylerin testlerde daha fazla zorlandığı söylenebilir. Örtük regresyon analizinden elde edilen standart sapmaların 0.84 ile 0.87 arasında değerler aldığı görülmektedir. Standart sapmaların Rasch modelden elde edilen değere (0.8) oldukça yakın olduğu söylenebilir. Bu durumun, örtük regresyon analizinin sadece birey değişkenlerine göre DMF gösteren ilk 10 maddenin parametreleri üzerinde etkili olmasından kaynaklandığı söylenebilir. Bu maddeler dışında kalan maddelerin parametreleri üzerinde birey değişkenlerinin etkisinin modellenmemesi ile bu maddeler için Rasch modele çok yakın kestirimler yapıldığı gözlenmiştir.

Alt gruplar bazında madde güçlüklerinin dağılımları Şekil 5'te sunulmuştur. Şekil 5 incelendiğinde madde değişkenlerinin kategorilerinin oluşturduğu alt gruplarda madde güçlükleri arasında farklılaşmaların olduğu görülmektedir. Örneğin 11-15. maddelerin i formunu, 16-20. maddelerin x formunu alan bireyler için ortalama madde güçlüğü -0.09 iken jz formunun uygulandığı bireyler için ortalama güçlük 0.1'dir. Bu durumda ix grubu için testlerin ortalamadan 0.1 logit daha kolay, jz alt grubu için ise 0.1 logit daha zor olduğu söylenebilir. Yine örtük regresyon analizinde olduğu gibi standart sapmaların 0.84 ile 0.87 arasında değiştiği, en küçük madde güçlüğü -2.13, en büyük değerin 1.98 olduğu görülmüştür. Standart sapmanın Rasch modelden bir miktar farklı kestirilmesine karşın en küçük ve en büyük güçlük değerlerinin Rasch modele çok yakın değerler olduğu söylenebilir.



Şekil 5. Alt gruplarda ortalama madde güçlüğü

Dört farklı modelden elde edilen ve betimsel istatistiklerde ortaya konan madde güçlük ortalamalarının, modeller arasında farklılaşıp farklılaşmadığı hipotez testi ile ortaya konmuştur. Rasch modelden elde edilen madde güçlükleri ile diğer üç modelden her bir alt grup için kestirilen madde güçlüklerinin ortalaması tekrarlı ANOVA testi ile karşılaştırılmış ve sonuçları Tablo 7’de sunulmuştur. Tekrarlı varyans analizinde Mauchly küresellik testinin manidar olduğu ( $w=.00$ ,  $p<.05$ ) ve verilerin küresellik varsayımını karşılamadığı gözlenmiştir. Greenhouse-Geisser düzeltmesine dayalı parametrelerin kullanıldığı analiz sonucunda en az bir madde güçlük ortalamasının manidar bir biçimde farklılaştığı ( $F_{4.41, 436.36}=4.02$ ,  $p<.05$ ) sonucuna ulaşılmıştır. Kısmi eta-kare değeri incelendiğinde farkın küçük etki büyüklüğüne sahip olduğu (Cohen, 1988) görülmektedir. Manidar bulunan farkın hangi alt gruplar arasında olduğunun tespit edilebilmesi için LSD yöntemi ile ikili karşılaştırmalar yapılmış ve sonuçları Tablo 8’de sunulmuştur.

Tablo 7

*Madde Güçlük Parametrelerine İlişkin Tekrarlı ANOVA Sonuçları*

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p	Kısmi $\eta^2$
Model (Greenhouse-Geisser)	4.777	4.41	1.08	4.02	.00*	.04
Hata	117.63	436.36	.27			

\* $p<.05$

Tablo 8

*Madde Güçlük Parametrelerine İlişkin İkili Karşılaştırma Sonuçları*

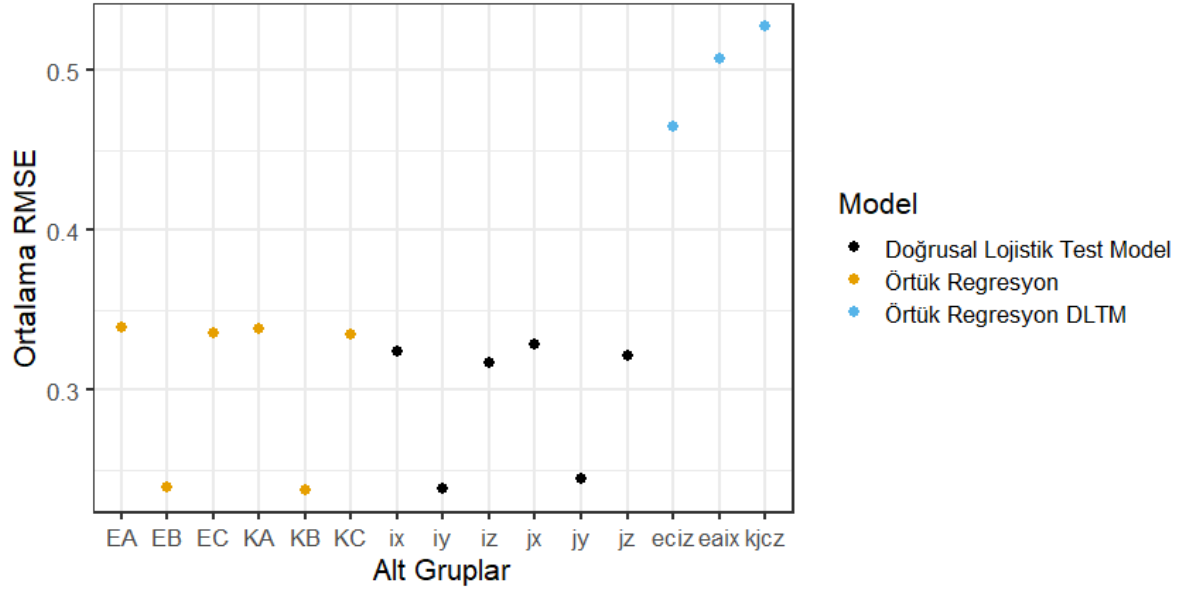
Model	Alt Gruplar	Ortalama Farkı	SH	p	Cohen d	
Örtük Regresyon	EA	.10	.03	.00*	.15	
	EB	.05	.02	.03*	-.03	
	EC	-.01	.03	.84		
	KA	.01	.03	.84		
	KB	-.05	.02	.03*	-.48	
	KC	-.10	.03	.00*	.31	
Rasch Modeli	Doğrusal Lojistik Test Modeli	ix	.09	.03	.00*	.15
		iy	.04	.02	.04*	.00
		iz	.00	.03	1.00	
	Örtük Regresyon Doğrusal Lojistik Test Modeli	jx	.00	.03	.97	
		jy	-.05	.02	.02*	-.45
		jz	-.09	.03	.00*	.12
Örtük Regresyon Doğrusal Lojistik Test Modeli	eaix	.03	.03	.30		
	eciz	.00	.03	.97		
	kcjz	-.03	.03	.35		

\*p&lt;.05

Tablo 8'e göre Rasch analizi ile elde edilen madde güçlük ortalaması ile örtük regresyon analizinde EA, EB, KB ve KC alt grupları için kestirilen madde güçlük ortalamaları arasında manidar bir fark olduğu söylenebilir. Yine Rasch analizi ile elde edilen madde güçlükleri ortalamalarının doğrusal lojistik test modelinde ix, iy, jy ve jz alt grupları için kestirilen güçlük parametrelerinin ortalamalarından istatistiksel olarak farklılaştığı görülmektedir. Rasch modeli ile örtük regresyon doğrusal lojistik test modelinden elde edilen güçlük parametreleri karşılaştırıldığında ise hiçbir alt grup ortalamasının Rasch model ortalamasından manidar bir şekilde farklılaşmadığı gözlenmiştir. Ayrıca manidar bulunan farkların etki büyüklükleri Cohen'in d katsayısı yardımı ile incelenmiştir. Buna göre Rasch modelden elde edilen ortalama madde güçlüğü ile örtük regresyon modelinde KB ve KC alt grupları için ve DLTM'de jy alt grubu için kestirilen ortalama güçlük parametreleri arasındaki farkın etki büyüklüğü küçüktür. Bu üç analiz dışında kalan tüm manidar farkların 0.2 standart sapmadan daha küçük olduğu ve istatistiksel olarak önemsiz olduğu söylenebilir (Cohen, 1988). Başka bir ifade ile bu farklar manidar bulunmuş olsa dahi pratikte bir anlam ifade etmeyebilir.

Şekil 6'da Rasch model ve üç AMTM'de alt gruplar için kestirilen madde güçlük parametrelerine ilişkin RMSE değerleri görülmektedir. Şekil 6'ya göre örtük regresyon modelinde alt gruplar için kestirilen madde güçlükleri ile Rasch modelden

kestirilen değerlerden elde edilen RMSE değerleri 0.23 ile 0.33 arasında değişmektedir. Yine doğrusal lojistik test modeli ile Rasch modelden elde edilen RMSE değerleri 0.24 ile 0.32 arasında değerler almaktadır. Buna karşın örtük regresyon doğrusal lojistik test modelden elde edilen RMSE değerlerinin diğer iki modelden daha yüksek oldukları görülmektedir. Bu modelden elde edilen RMSE değerleri 0.46 ile 0.54 arasında değişmektedir.



Şekil 6. Madde güçlüklerine ilişkin ortalama RMSE değerleri

Şekil 6 incelendiğinde örtük regresyon ve doğrusal lojistik test modeli ile sekiz alt grup için kestirilen madde güçlük değerlerinin ortalamalarının Rasch modelden kestirilen değerlerin ortalamalarından manidar bir biçimde farklılaşmasına karşın RMSE değerlerinin ÖRDLTM'den daha düşük olduğu görülmektedir. Bu durumun tekrarlı ANOVA ile RMSE değerlerinin hesaplanma biçiminden kaynaklandığı söylenebilir. RMSE değerlerinin hesaplanmasında iki değer karelerinin alınıp toplanması, negatif ve pozitif farkların birbirini nötrleştirmesine engel olabilir. Tekrarlı ANOVA hesaplamaları ise grup ortalamalarına ve genel ortalama arasındaki farkın karesine dayalıdır. Bu durumda her bir ölçümün ortalaması ile tüm puanların genel ortalamalarının yakın olduğu düşünülebilir. Her ne kadar hiçbir alt grubu için kestirilen güçlük parametrelerinin ortalaması ile Rasch modelde kestirilen değerlerin ortalaması arasında manidar bir fark bulunmasa da eşleştirilmiş madde güçlükleri arasında en büyük farkların ÖRDLTM'den elde edildiği söylenebilir.

## İkinci Alt Probleme İlişkin Bulgular ve Yorumlar

Araştırmanın ikinci alt problemi olan “Örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinden elde edilen yetenek parametreleri, Rasch modeli ile elde edilen parametrelerden farklılaşmakta mıdır?” sorununa yönelik olarak kestirilen madde güçlükleri ve madde cevap örüntüleri ile her katılımcı için beklenen sonsal dağılım (BSD) yetenek puanları elde edilmiştir. Rasch modelden ve diğer açıklayıcı madde tepki modellerinden elde edilen yetenek puanlarının karşılaştırılması betimsel istatistikler, bağımlı örneklem için t-testi, RMSE değerleri ve korelasyon analizleri yardımı ile yapılmıştır. Rasch model ve diğer üç AMTM’de her bir alt grup için kestirilen yetenek puanlarına Tablo 9’da yer verilmiştir.

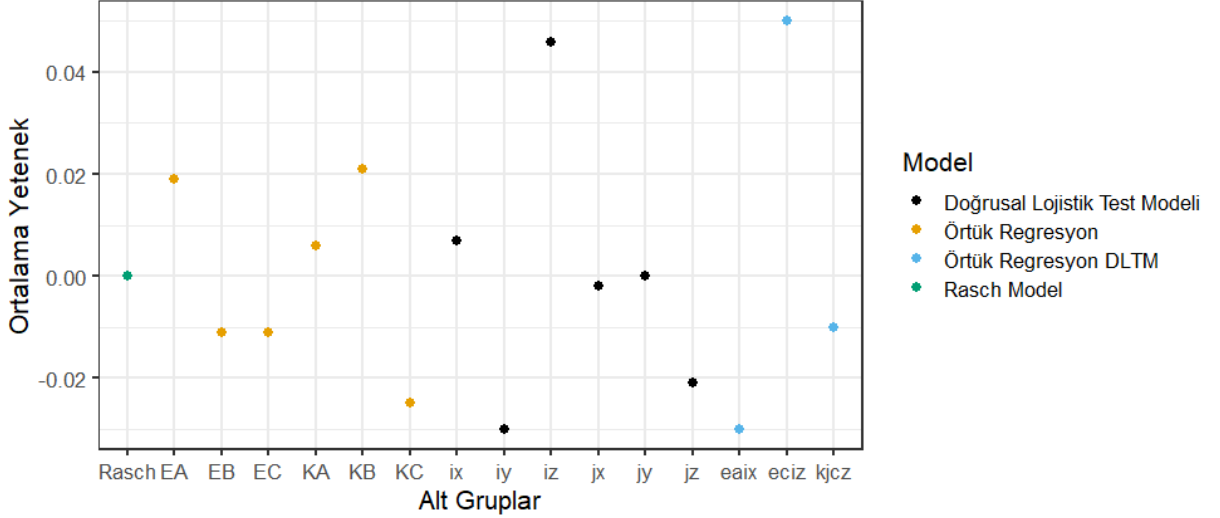
Tablo 9

*Madde Havuzundan Kestirilen Yetenek Puanlarına İlişkin Betimsel İstatistikler*

Model	Alt Grup	Havuz Sayısı	n	$\bar{x}_\theta$	$\sigma_\theta$	En Küçük $\theta$	En Büyük $\theta$	SH
Rasch Model	Rasch	10	1440	.00	.96	-3.24	3.39	.01
Örtük Regresyon	EA	10	240	.02	.97	-2.96	2.84	.02
	EB	10	240	-.01	.96	-3.30	3.28	.02
	EC	10	240	-.01	.93	-3.12	3.06	.02
	KA	10	240	.01	.94	-2.97	3.07	.02
	KB	10	240	.02	.96	-3.03	3.34	.02
	KC	10	240	-.02	.97	-2.89	2.87	.02
Doğrusal Lojistik Test Modeli	ix	10	240	.01	.94	-3.01	2.98	.02
	iy	10	240	-.03	.95	-3.11	2.87	.02
	iz	10	240	.05	.97	-3.27	3.35	.02
	jx	10	240	.00	.94	-3.11	2.94	.02
	jy	10	240	.00	.97	-2.91	3.34	.02
jz	10	240	-.02	.95	-2.89	2.84	.02	
Örtük Regresyon	eaix	10	40	-.03	.99	-2.45	2.43	.05
Doğrusal Lojistik	ecz	10	40	.05	.96	-2.52	2.74	.05
Test Modeli	kcjz	10	40	-.01	1.02	-2.84	2.97	.05

Tablo 9’a göre Rasch modelden elde edilen ortalama yetenek puanı 0 iken en küçük değer -3.24 ve en büyük değer 3.39’dur. Buna karşın örtük regresyon analizinden elde edilen ortalama yetenek puanları -0.02 ile 0.02 arasında değişmektedir. Örneğin ortalama madde güçlüğü -0.1 olduğu EA grubu için ortalama yetenek puanı 0.02 iken, madde güçlük ortalamasının 0.1 olduğu KC grubu için ise ortalama yetenek puanı -0.02 olarak kestirilmiştir. Bu durumun doğal sonucu olarak soru havuzlarında ortalamaya kıyasla daha az zorlanan katılımcıların daha yüksek yetenek puanı ortalamalarına, daha fazla zorlanan katılımcıların ise daha

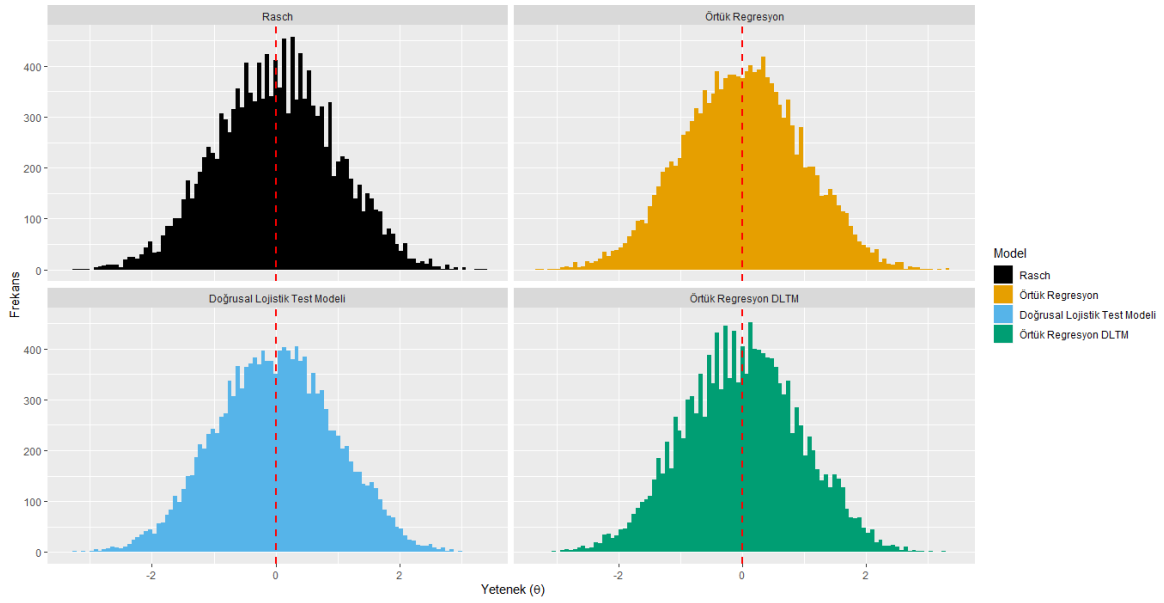
düşük yetenek puan ortalamasına sahip oldukları söylenebilir. Örtük regresyon analizine benzer bir biçimde doğrusal lojistik test modelinde de yetenek puanı ortalamaları -0.03 ile 0.01 arasında değişmektedir.



Şekil 7. Madde havuzunda alt gruplar için kestirilen ortalama yetenek puanları

Buna karşın Şekil 7’de görülebileceği üzere DLTM’de alt gruplar için kestirilen yetenek puanlarının ortalamalarının en küçüğünün -0.03, en büyüğünün ise 0.04 civarında olduğu görülmektedir. Ayrıca, örtük regresyon doğrusal lojistik test modelde alt gruplar için kestirilen ortalama yetenek puanlarının -0.03 ile 0.05 arasında değerler aldığı tespit edilmiştir. Rasch modelden kestirilen yetenek puanı ortalamalarından en fazla sapmanın ÖRDLTM’nin alt grupları için kestirilen yetenek puanlarında olduğu görülmektedir.

Tablo 9’da ayrıca Rasch modelinde yetenek puanları için ortalama standart sapmanın 0.96 olduğu, örtük regresyon analizde ortalama standart sapmaların 0.93 ile 0.97 arasında değerler aldığı, doğrusal lojistik test modelinde standart sapmaların 0.94 ve 0.97 arasında değiştiği ve ÖRDLTM’de ise 0.96 ve 1.02 arasında olduğu görülmektedir. Bu durumda, soru havuzlarında kestirilen yetenek puanlarının alt gruplar arasında ve alt grupların kendi içinde en fazla farklılaştığı modelin ÖRDLTM olduğu söylenebilir. Her üç AMTM’de madde güçlüklerinde olduğu gibi yetenek puanlarının ortalaması da alt gruplar göz ardı edildiğinde, Rasch modelin ortalamasına oldukça yakındır. Gruplar baz alındığında ise madde güçlüklerinin bir sonucu olarak avantajlı olan alt grupların, ortalamasının 0.02 logit üstünde olduğu görülmektedir.



Şekil 8. Madde havuzlarından kestirilen yetenek puanlarının dağılımı

Alt gruplara ayırmaksızın tüm katılımcıların dört modelde kestirilen yetenek puanlarının dağılımı Şekil 8'de sunulmuştur. Şekil 8 incelendiğinde tüm yetenek puanlarının her dört modelde de benzer bir şekilde normal dağılıma sahip olduğu ve ortalamalarının 0 olduğu görülmektedir.

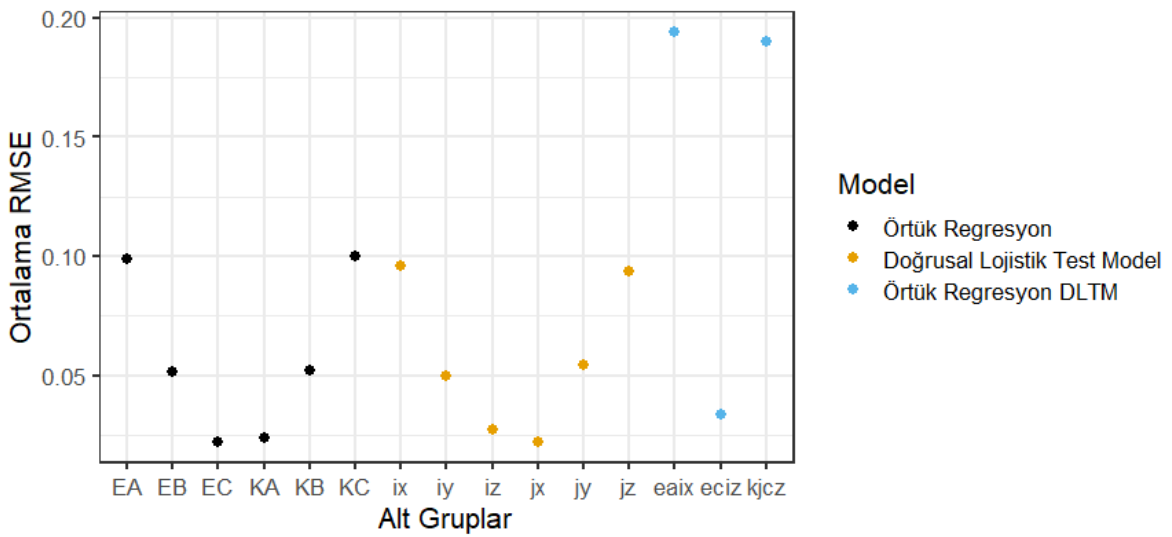
Tablo 10

Madde Havuzlarında Alt Gruplar İçin Kestirilen Yetenek Puanlarına İlişkin t-Testi Sonuçları

Model	Alt Gruplar	$\bar{x}_\theta$	Ortalama Farkı	t	sd	p	Cohen d
Rasch-Örtük Regresyon	EA	.10	.10	197.06	2399	.00*	14.31
	EB	.05	.05	91.00	2399	.00*	6.18
	EC	-.01	-.01	-13.10	2399	.00*	-1.07
	KA	.01	.01	12.20	2399	.00*	.87
	KB	-.04	-.04	-8.32	2399	.00*	-6.05
	KC	-.10	-.10	-194.81	2399	.00*	-14.36
Rasch-Doğrusal Lojistik Test Modeli	ix	.09	.09	189.03	2399	.00*	12.65
	iy	.04	.04	76.55	2399	.00*	5.52
	iz	.00	.00	3.00	2399	.00*	.21
	jx	.00	.00	3.00	2399	.00*	.21
	jy	-.05	-.05	-92.69	2399	.00*	-7.09
	jz	-.09	-.09	-178.81	2399	.00*	-13.38
Rasch-Örtük Regresyon Doğrusal Lojistik Test Modeli	eaix	.19	.19	91.72	399	.00*	29.09
eciz	-.005	-.005	-3.01	399	.00*	-.64	
kcjz	-.19	-.19	-11.34	399	.00*	-17.53	

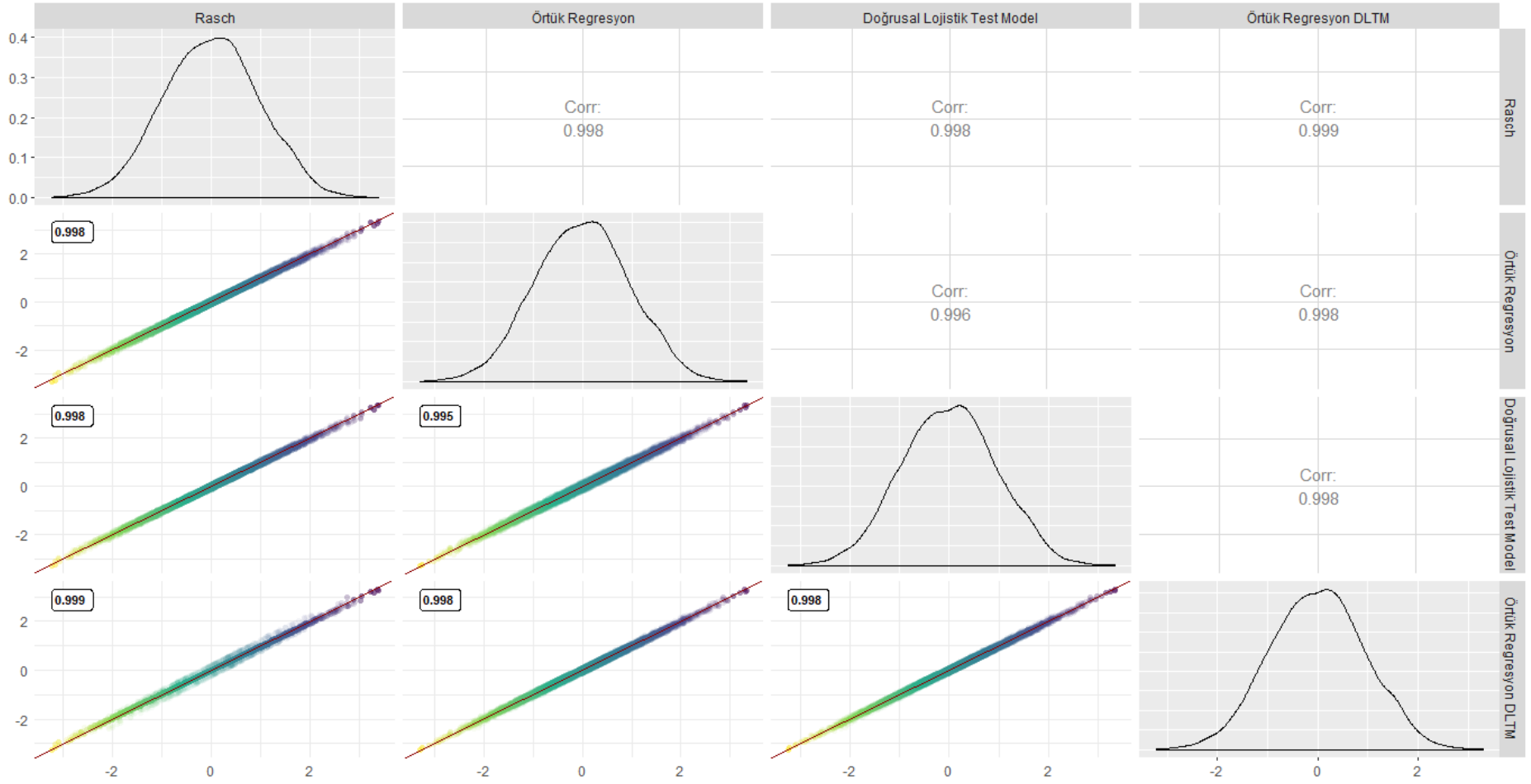
\*p<.05

Tablo 10'da alt gruplar için kestirilen yetenek puanlarının ortalaması ile Rasch modelden elde edilen yetenek parametrelerinin ortalamasının karşılaştırıldığı bağımlı örneklem için t-testi sonuçları sunulmuştur. Tablo 10'da örtük regresyon modelinde EA ( $t=197.06$ ,  $p<.05$ ), EB ( $t=91.00$ ,  $p<.05$ ), EC ( $t=-13.10$ ,  $p<.05$ ), KA ( $t=12.20$ ,  $p<.05$ ), KB ( $t=-80.32$ ,  $p<.05$ ) ve KC ( $t=-194.81$ ,  $p<.05$ ) alt grupları için kestirilen yetenek parametrelerinin ortalamalarının Rasch modelden kestirilen yetenek parametrelerinin ortalamasından istatistiksel olarak manidar bir biçimde farklılaştığı görülmektedir. Benzer bir şekilde DLTM'de ix ( $t=189.03$ ,  $p<.05$ ), iy ( $t=76.55$ ,  $p<.05$ ), iz ( $t=3.00$ ,  $p<.05$ ), jx ( $t=3.00$ ,  $p<.05$ ), jy ( $t=-92.69$ ,  $p<.05$ ) ve jz ( $t=-178.81$ ,  $p<.05$ ) alt gruplarının yetenek ortalamaları da Rasch modelden kestirilen yetenek puanı ortalamalarından farklılaşmaktadır. Örtük regresyon doğrusal lojistik test modelinde ise eaix ( $t=91.72$ ,  $p<.05$ ), eciz ( $t=-3.01$ ,  $p<.05$ ) ve kciz ( $t=-110.34$ ,  $p<.05$ ) alt grupları için kestirilen yeteneklerin ortalamaları ile Rasch modelden elde edilen değerler arasında manidar bir fark tespit edilmiştir. Manidar bulunan farkların etki büyüklüklerini incelendiğinde Rasch modelden kestirilen yetenek puanları ile DLTM'de iz ( $d=0.21$ ) ve jx ( $d=0.21$ ) alt grupları için kestirilen yetenek puanları arasındaki manidar farka ilişkin etki büyüklüğünün küçük olduğu görülmektedir. Ayrıca Rasch modele ilişkin yetenek puanları ortalaması ile ÖRDLTM'de eciz alt grubu için kestirilen yetenek puanı ortalaması arasındaki manidar farka ait etki büyüklüğü orta düzeydedir ( $d=-0.64$ ). Bunların dışında kalan tüm manidar farklara ilişkin etki büyüklükleri yüksektir.



Şekil 9. Madde havuzlarında kestirilen yetenek puanlarına ilişkin ortalama RMSE değerleri

Tüm madde havuzlarından elde edilen yetenek puanlarına dayalı ortalama RMSE değerleri Şekil 9'da görülmektedir. Şekil 9'a göre örtük regresyon analizinde en avantajlı grup olan EA ve en dezavantajlı grup olan KC için kestirilen ortalama RMSE değerlerinin 0.1, EB ve KB alt grupları için 0.05 ve EC ve KA alt grupları için yaklaşık olarak 0.025 olduğu söylenebilir. Yine benzer bulguların doğrusal lojistik test modeli için de geçerli olduğu görülmektedir. Bu modelde ix alt grubu ile jz alt grubu için kestirilen ortalama RMSE değeri yaklaşık olarak 0.1 olarak belirlenirken, iy ve jy alt grupları için bu yaklaşık 0.05 ve iz ve jx alt grupları için yaklaşık olarak 0.025 olarak hesaplanmıştır. Madde güçlük parametrelerine ait RMSE değerlerinde olduğu gibi yetenek parametresine ilişkin en yüksek RMSE değerleri, ÖRDLTm'den elde edilmiştir. Bu modelde eaix ve kcjz alt grupları için ortalama RMSE değeri yaklaşık olarak 0.2 olarak hesaplanmışken eciz alt grubu için bu değer 0.03 olarak bulunmuştur.



Şekil 10. Madde havuzlarında kestirilen yetenek puanlarına ilişkin korelasyon katsayıları

Şekil 10'da Rasch modeli ve üç açıklayıcı madde tepki modelinden elde edilen yetenek parametreleri arasındaki korelasyonun grafiği sunulmaktadır. Normal dağılım gösteren yetenek puanları ile gerçekleştirilen Pearson korelasyonu analizi sonucunda oluşturulan korelasyon matrisi incelendiğinde tüm modellerden elde edilen yetenek puanları arasında yüksek düzeyde manidar, pozitif bir ilişki olduğu görülmektedir.

### Üçüncü Alt Probleme İlişkin Bulgular ve Yorumlar

Araştırmanın üçüncü alt problemi olan “Rasch, örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinin model-veri uyumları nasıldır?” sorusuna ilişkin olarak tüm madde havuzlarında hiyerarşik (iç-içe geçmiş) olmayan modellerin karşılaştırılmasında kullanılan Akaike Bilgi Kriteri (AIC), Bayes Bilgi Kriteri (BIC) ve Sapma Bilgi Kriterleri (DIC) elde edilmiş ve bulgular Tablo 11’de sunulmuştur.

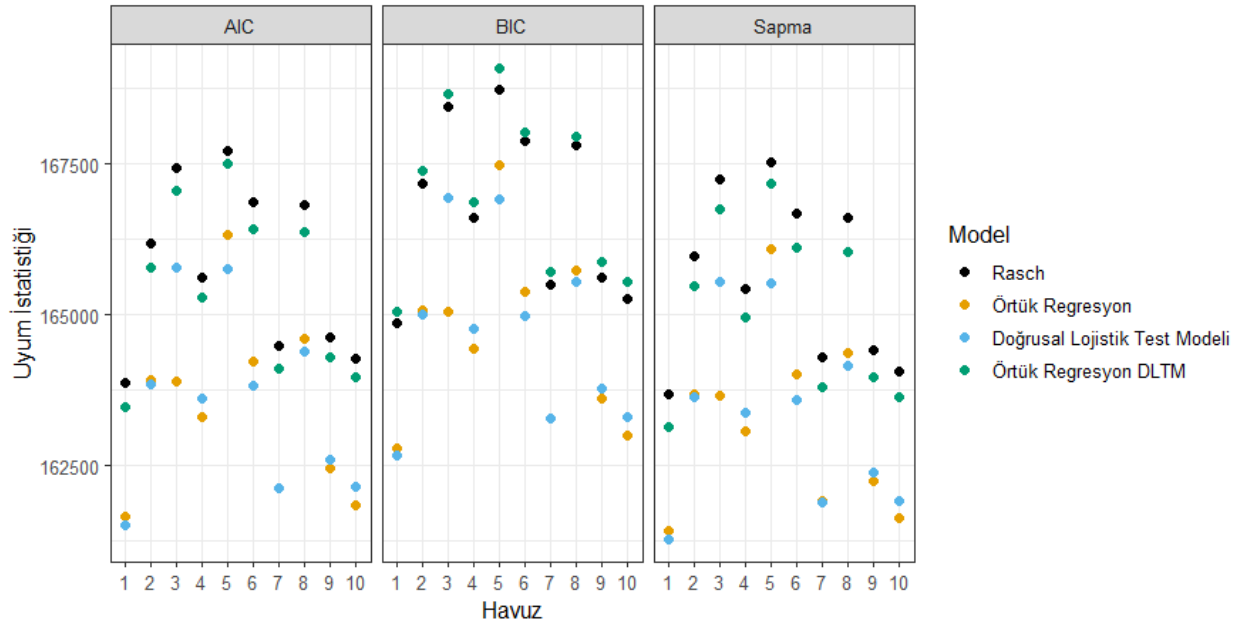
Tablo 11

#### *Madde Havuzlarındaki Kestirimlere İlişkin Model-veri Uyum İndeksleri*

İndeks	Model	n Madde Havuzu	Ortalama	En Küçük	En Büyük
AIC	Rasch	10	165791.8	163877	167721
	Örtük Regresyon	10	163439.2	161649	166330
	Doğrusal Lojistik Test Modeli	10	163562.0	161649	166330
	Örtük Regresyon DLTM	10	165425.0	163463	167497
BIC	Rasch	10	166789.4	164874	168719
	Örtük Regresyon	10	164585.0	162795	167475
	Doğrusal Lojistik Test Modeli	10	164717.6	162674	166931
	Örtük Regresyon DLTM	10	167015.3	165054	169087
Sapma (Deviance)	Rasch	10	165589.8	163675	167519
	Örtük Regresyon	10	163207.2	161417	166098
	Doğrusal Lojistik Test Modeli	10	163328.0	161284	165541
	Örtük Regresyon DLTM	10	165103.0	163141	167175

Tablo 11’e göre 10 madde havuzundan elde edilen en küçük ortalama AIC değeri örtük regresyon modelinden (163439.2) elde edilmişken en yüksek AIC değeri Rasch modelden (165791.8) elde edilmiştir. Yine DLTM’de elde edilen AIC değerinin (163562) örtük regresyondan elde edilen değere yakın olduğu; buna karşın ÖRDLTM’de hesaplanan AIC değerinin (165425) Rasch modelle benzerlik gösterdiği saptanmıştır.

AIC değerlerinde olduğu gibi BIC değerlerinde de en düşük değerin elde edildiği model örtük regresyon (164585) modeli olurken DLTM de bu değer 164717.6 olarak belirlenmiştir. Buna karşın, ortalama BIC değeri Rasch modelde 166789.4, ÖRDLTM'de 167015.3 olarak tespit edilmiştir. Sapma değerlerinde en küçük değer yine örtük regresyondan (163207.2) elde edilmişken en DLTM'de bu değer 163328'dir. En yüksek sapma değeri ise Rasch modelden (165589.8) elde edilmiştir. ÖRDLTM'de diğer uyum indekslerinde olduğu gibi sapma değeri (165103) de Rasch modele oldukça yakın kestirilmiştir. Her üç uyum indeksinde de değerin küçük olması daha iyi uyum sağlandığı anlamına gelmektedir. Bu durumda hiyerarşik olmayan dört modelde model-veri uyumları arasındaki farkın manidarlığının belirlenmesi için hipotez testleri (-2logLL değerleri ile  $\chi^2$  testi) yapılamamasına rağmen ortalama olarak madde havuzlarına en uygun modelin örtük regresyon olduğu söylenebilir. Örtük regresyona benzer bir biçimde DLTM'nin de en iyi ikinci model-veri uyumuna sahip olduğu; Rasch modelin ise en düşük uyumu gösterdiği belirlenmiştir. Tüm modellerin madde havuzlarının her birinde model-veri uyumu bakımından nasıl performans gösterdikleri Şekil 11'de sunulmuştur. Şekil 11'e göre tüm madde havuzlarında en yüksek model-veri uyumu örtük regresyon veya DLTM ile sağlanmışken, en düşük uyum AIC ve sapma indeksine göre Rasch modelde; BIC indeksine göre ÖRDLTM'de gözlenmiştir.



Şekil 11. Madde havuzu bazında model-veri uyum indeksleri

## Dördüncü Alt Probleme İlişkin Bulgular ve Yorumlar

Çalışmanın dördüncü alt problemi olan “Rasch, örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinden elde edilen parametreler ile iki farklı sonlandırma kuralına göre gerçekleştirilen BBT uygulamalarından kestirilen yetenek parametreleri nasıldır?” sorusuna yönelik olarak her bir modelden elde edilen madde ve önsel yetenek parametreleri kullanılarak kesinlik ve minInfo sonlandırma kuralları dayalı 10 döngüden oluşan post-hoc simülasyonları gerçekleştirilmiştir. Her bir model için gerçekleştirilen simülasyonlardan elde edilen yetenek puanlarının karşılaştırılması, betimsel istatistikler, yoğunluk grafiği ve tekrarlı tek yönlü varyans analizi yardımı ile gerçekleştirilmiştir.

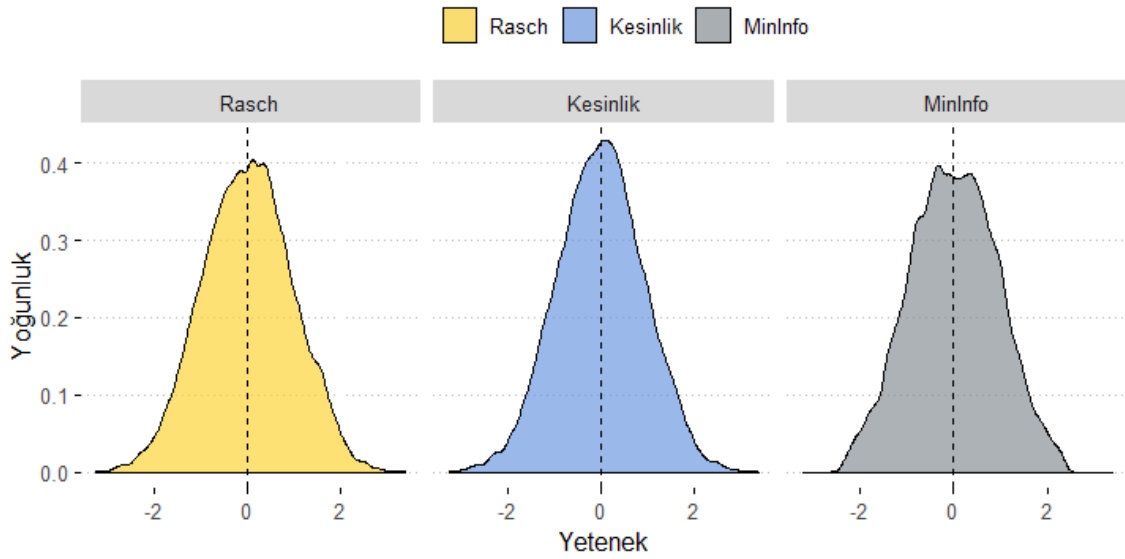
**Rasch modele dayalı BBT simülasyonu.** 10 ayrı madde havuzunda Rasch modelinden elde edilen parametreler ile 10 döngü olacak şekilde iki farklı sonlandırma kuralına göre yapılan simülasyonlardan elde edilen yetenek parametrelerine ilişkin betimsel istatistikler Tablo 12’de sunulmuştur.

Tablo 12

*Rasch Model Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Betimsel İstatistikler*

Sonlandırma Kuralı	Madde Havuzu	Döngü Sayısı	n	$\bar{x}_\theta$	$\sigma_\theta$	En Küçük $\theta$	En Büyük $\theta$	Çarpıklık	Basıklık
Kesinlik (Precision)	1	10	1440	-.01	.95	-3.24	2.91	.00	-.04
	2	10	1440	.00	.94	-2.80	3.29	-.01	-.29
	3	10	1440	.00	.89	-3.02	2.94	-.01	.23
	4	10	1440	.00	.95	-3.18	2.90	-.11	-.20
	5	10	1440	.00	.91	-2.72	3.21	-.07	.02
	6	10	1440	.00	.91	-3.10	3.28	.00	.12
	7	10	1440	.00	.91	-2.77	2.85	.02	-.12
	8	10	1440	.00	.91	-2.72	2.68	.02	-.28
	9	10	1440	.00	.95	-2.89	3.22	.05	-.07
	10	10	1440	-.01	.94	-2.95	3.05	-.01	-.03
minInfo	1	10	1440	.00	.95	-2.64	2.46	.00	-.39
	2	10	1440	.00	.95	-2.24	2.39	-.02	-.56
	3	10	1440	.00	.90	-2.44	2.43	.02	-.25
	4	10	1440	.00	.95	-2.37	2.16	-.06	-.59
	5	10	1440	.01	.91	-2.33	2.40	-.06	-.41
	6	10	1440	.00	.91	-2.21	2.43	-.06	-.40
	7	10	1440	.00	.92	-2.32	2.28	.03	-.33
	8	10	1440	.00	.93	-2.41	2.30	.04	-.48
	9	10	1440	.00	.95	-2.41	2.29	.08	-.46
	10	10	1440	-.01	.94	-2.43	2.33	-.01	-.42

Tablo 12'ye göre her iki sonlandırma kuralında da iki madde havuzu dışındaki tüm havuzlardan kestirilen ortalama yetenek 0'a eşittir. Kesinlik sonlandırma kuralında birinci ve onuncu madde havuzunda kestirilen ortalama yetenek -0.01 iken, minInfo sonlandırma kuralında beşinci madde havuzu için kestirilen ortalama yetenek 0.01, onuncu madde havuzu için bu değer -0.01'dir. Her iki sonlandırma kuralında da kestirilen standart sapmalar yakın olmakla birlikte, kesinlik sonlandırma kuralı için standart sapmalar 0.89 ile 0.95 arasında değişirken minInfo için sapmalar 0.90 ve 0.95 arasında değerler almaktadır.



Şekil 12. Rasch model parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarının dağılımı

Madde havuzları göz önünde bulundurulmadığında ise Şekil 12'deki yoğunluk grafiğinde görülebileceği üzere her iki sonlandırma kuralı ile kestirilen ortalama yetenek 0 iken standart sapma 0.93 olarak tespit edilmiştir. Ancak kesinlik sonlandırma kuralı (-3.24, 3.29) ile Rasch modelden (-3.24, 3.39) elde edilen en küçük ve en büyük yetenek puanları birbirine oldukça yakınken, minInfo kuralından kestirilen bu değerler -2.64 ve 2.46'dır. Buna bağlı olarak kesinlik sonlandırma kuralı ile elde edilen yetenek parametrelerinin dağılımının Rasch modele benzer olduğu söylenebilir.

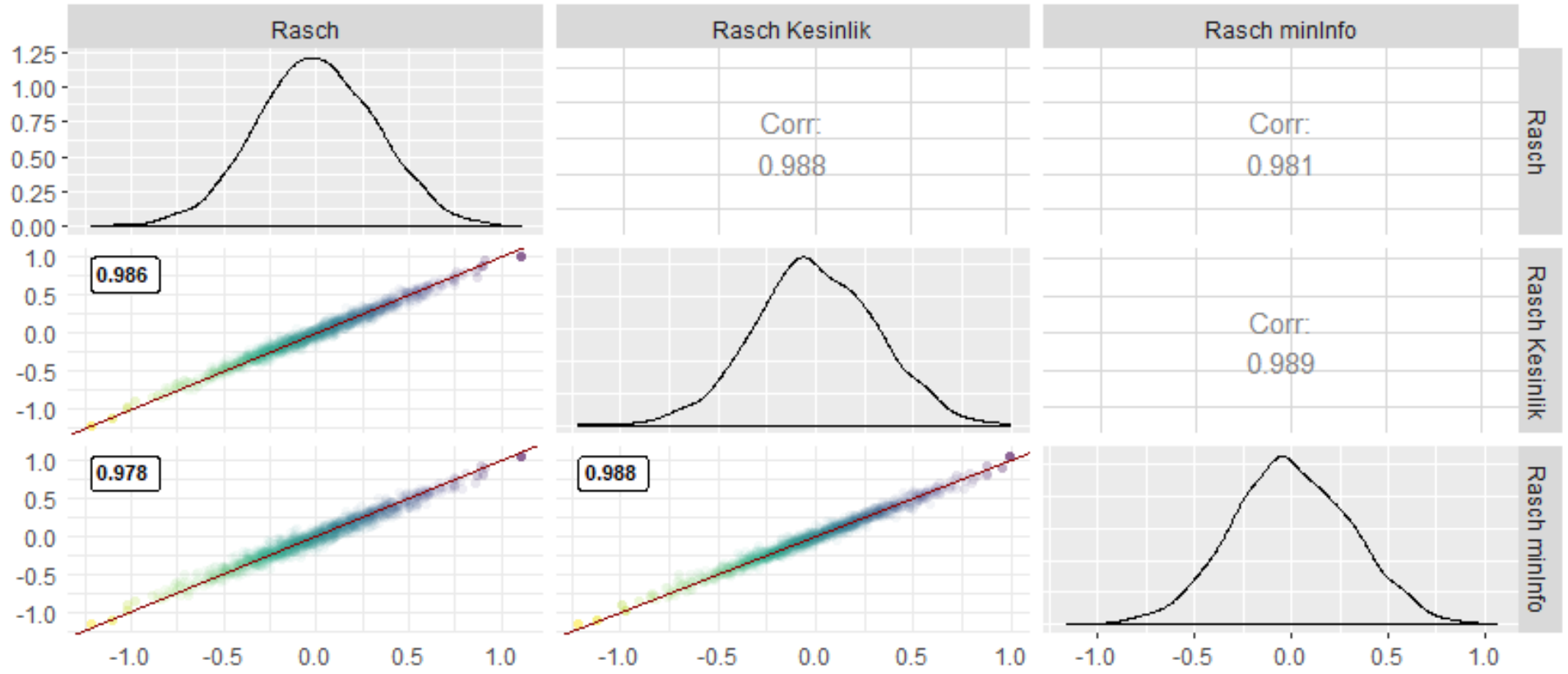
Tablo 13

*Rasch Model Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları*

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p
Model (Greenhouse-Geisser)	.00	1.74	.00	1.47	.23
Hata	4.26	2499.75	.00		

Tablo 13'te tüm madde havuzundan Rasch model ile elde edilen yetenek parametreleri ile BBT simülasyonlarından elde edilen ortalama yetenek puanlarının karşılaştırılması için gerçekleştirilen tekrarlı ANOVA sonuçları sunulmuştur. Analiz sonucunda Mauchly küresellik testinin manidar olduğu ( $w=0.85$ ,  $p<.05$ ) belirlendiğinden Greenhouse-Geisser yöntemi ile yapılan düzeltilmiş parametreler kullanılmıştır. Tablo 13'e göre Rasch modelden elde edilen önsel yetenek puanları ortalaması ile bu modelden elde edilen parametrelerle kesinlik ve minInfo sonlandırma kuralına göre yapılan BBT simülasyonlarından elde edilen ortalama yetenek puanları arasında manidar bir fark gözlenmemektedir ( $F_{1.7, 2499.8}$ ,  $p>.05$ ). Başka bir ifade ile her iki sonlandırma kuralına göre yapılan simülasyonlarda elde edilen ortalama yetenek puanları ile Rasch modelden elde edilen yetenek puanlarının ortalaması istatistiksel olarak eşittir. Tekrarlı tek yönlü varyans analizinde eşleştirilmiş bireysel puanlar arasında manidar düzeyde fark gözlenmemesi BBT simülasyonlarının daha az madde ile madde havuzunda kestirilen ortalama yetenek puanına yakın kestirimler yapabildiği şeklinde yorumlanabilir.

Şekil 13'te Rasch modeli ile madde havuzlarından kestirilen yetenek puanı ile BBT simülasyonlarında kestirilen yetenek puanları arasındaki korelasyonlar görülmektedir. Şekil 13 incelendiğinde Rasch modelden elde edilen yetenek puanları ile her iki sonlandırma kuralına göre yapılan BBT uygulamalarından elde edilen yetenek puanları yüksek ilişkili çıkmıştır. Hem yetenek puanları ortalaması arasında manidar fark bulunmaması hem de yüksek düzeyde pozitif ilişki gözlenmesi BBT sonlandırma kurallarının yakın düzeyde kestirimler sağladığı şeklinde yorumlanabilir.



Şekil 13. Rasch model parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarına ilişkin korelasyon katsayıları

**Örtük regresyon modeline dayalı BBT simülasyonu.** Örtük regresyon modelinden elde edilen madde ve birey parametrelerinin kullanıldığı iki sonlandırma kuralına göre 10 döngüde gerçekleştirilen post-hoc simülasyonları sonucunda elde edilen beklenen sonsal dağılım (BSD) puanlarına ilişkin betimsel istatistikler Tablo 14' te sunulmuştur. Tablo 14' te kesinlik sonlandırma kuralına göre birinci ve onuncu madde havuzu için ortalama yetenek puanı -0.01 olarak kestirilmişken diğer madde havuzları için bu değer 0 olarak hesaplanmıştır. Benzer bir biçimde minInfo sonlandırma kuralı için de her iki sonlandırma kuralına göre yapılan simülasyonlarda birinci ve onuncu madde havuzu dışındaki tüm madde havuzları için ortalama yetenek puanı 0 olarak kestirilmişken bu iki madde havuzu için -0.01 ve beşinci madde havuzu için 0.01 olarak tespit edilmiştir. Her iki sonlandırma kuralında kestirilen ortalama yeteneğin yanında standart sapmaların da oldukça yakın olduğu belirlenmiştir. Kesinlik sonlandırma kuralı için standart sapma, 0.92-0.96 aralığında değerler alırken, minInfo kuralında 0.91-0.96 aralığında değişmektedir.

Tablo 14

*ÖR Modeli Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Betimsel İstatistikler*

Sonlandırma Kuralı	Madde Havuzu	Döngü Sayısı	n	$\bar{x}_\theta$	$\sigma_\theta$	En Küçük $\theta$	En Büyük $\theta$	Çarpıklık	Basıklık
Kesinlik (Precision)	1	10	1440	-.01	.96	-3.28	2.95	.00	-.05
	2	10	1440	.00	.95	-2.91	3.27	-.01	-.32
	3	10	1440	.00	.92	-3.00	3.01	-.03	.19
	4	10	1440	.00	.97	-3.24	2.93	-.11	-.19
	5	10	1440	.00	.92	-2.74	3.24	-.07	.00
	6	10	1440	.00	.93	-3.09	3.33	.03	.12
	7	10	1440	.00	.92	-2.87	2.81	.03	-.13
	8	10	1440	.00	.93	-2.68	2.69	.02	-.30
	9	10	1440	.00	.96	-2.99	3.28	.05	-.07
	10	10	1440	-.01	.96	-2.92	3.09	-.02	-.06
minInfo	1	10	1440	-.01	.96	-2.61	2.52	.00	-.36
	2	10	1440	.00	.95	-2.58	2.44	-.02	-.55
	3	10	1440	.00	.91	-2.57	2.47	.00	-.21
	4	10	1440	.00	.96	-2.35	2.18	-.07	-.58
	5	10	1440	.01	.92	-2.36	2.39	-.05	-.41
	6	10	1440	.00	.92	-2.47	2.46	-.03	-.36
	7	10	1440	.00	.93	-2.34	2.33	.02	-.33
	8	10	1440	.00	.94	-2.44	2.43	.01	-.48
	9	10	1440	.00	.95	-2.42	2.30	.07	-.44
	10	10	1440	-.01	.95	-2.45	2.34	-.02	-.42

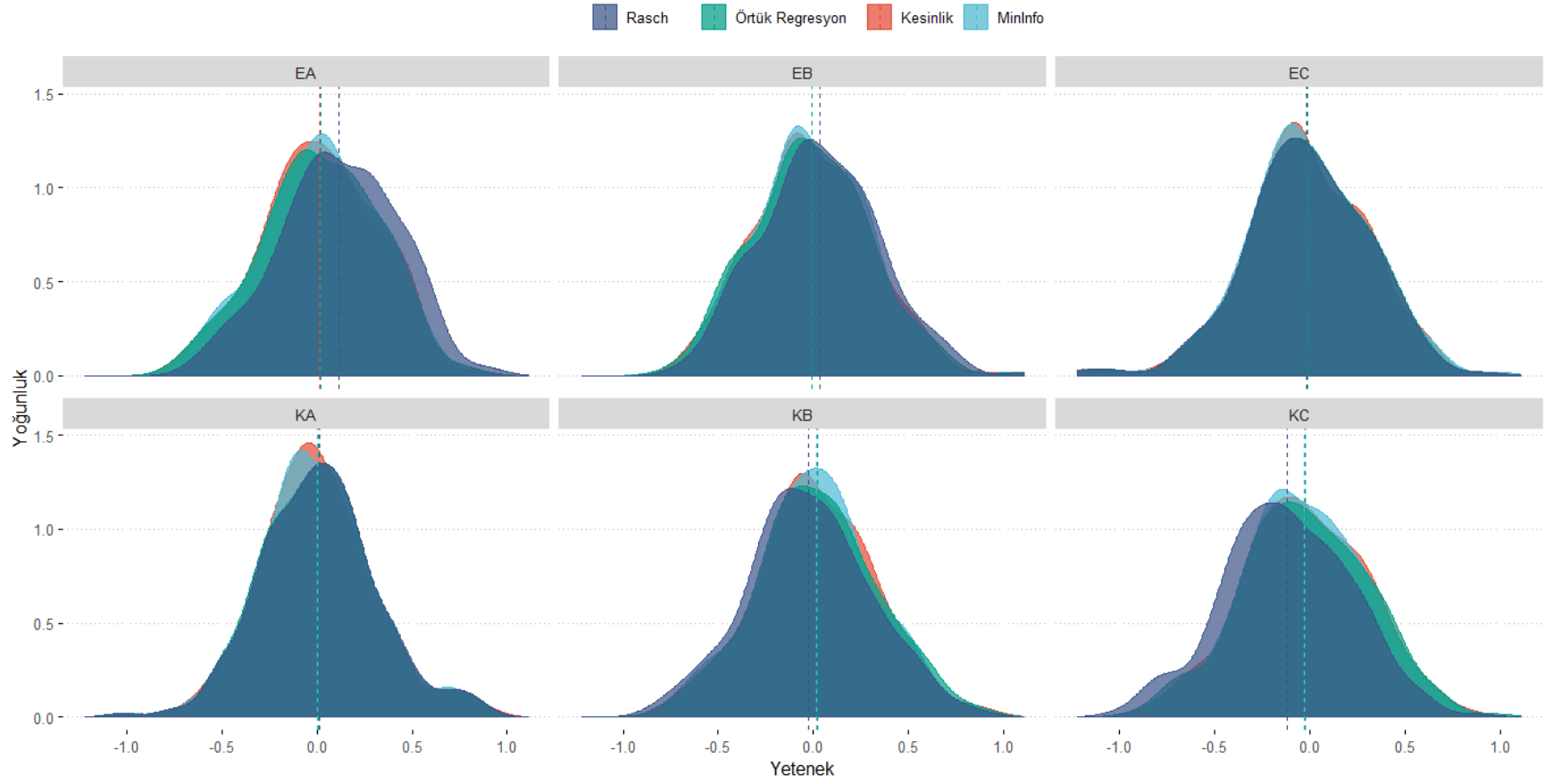
Tablo 15

*ÖR Modeli Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları*

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p
Model (Greenhouse-Geisser)	.00	1.97	.00	.53	.58
Hata	9.83	2836.84	.00		

Tablo 15'te alt gruplar göz önünde bulundurulmadan tüm katılımcılara ait Rasch ve örtük regresyon modellerinden elde edilen önsel yetenek puanları (Rasch modele ait önsel yetenek puanları referans puan olarak kullanılmıştır) ile iki farklı sonlandırma kuralına göre gerçekleştirilen post-hoc simülasyonlarından kestirilen yetenek puanlarının ortalamasına ilişkin tekrarlı ANOVA sonuçları sunulmuştur. Mauchly küresellik testinin manidar bulunmasından ( $w= 0.46$ ,  $p<.05$ ) dolayı Greenhouse-Geisser yöntemi ile düzeltilmiş parametrelerden yararlanılmıştır. Buna göre Rasch model ve ÖR modelinden elde edilen ortalama yetenek puanları ve kesinlik ve minInfo sonlandırma kuralına göre gerçekleştirilen post-hoc simülasyonlarından elde edilen yetenek puanları ortalaması arasında manidar bir fark gözlenmemektedir ( $F_{1.97, 2836.84}=0.53$ ,  $p>.05$ ).

Şekil 14'te alt gruplar baz alındığında madde havuzlarından ve BBT simülasyonlarında elde edilen yetenek puanlarının dağılımı görülmektedir. Şekil 14 incelendiğine özellikle EA alt grubu için Rasch modelde kestirilen ortalama yetenek puanının örtük regresyon modeli ile kestirilen ortalama yetenek puanından yüksek olduğu görülmektedir. KC alt grubu için Rasch modelde kestirilen ortalama yetenek puanının ise örtük regresyonda kestirilen ortalama yetenekten küçük olduğu gözlenmektedir.



Şekil 14. Örtük regresyon parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarının dağılımı

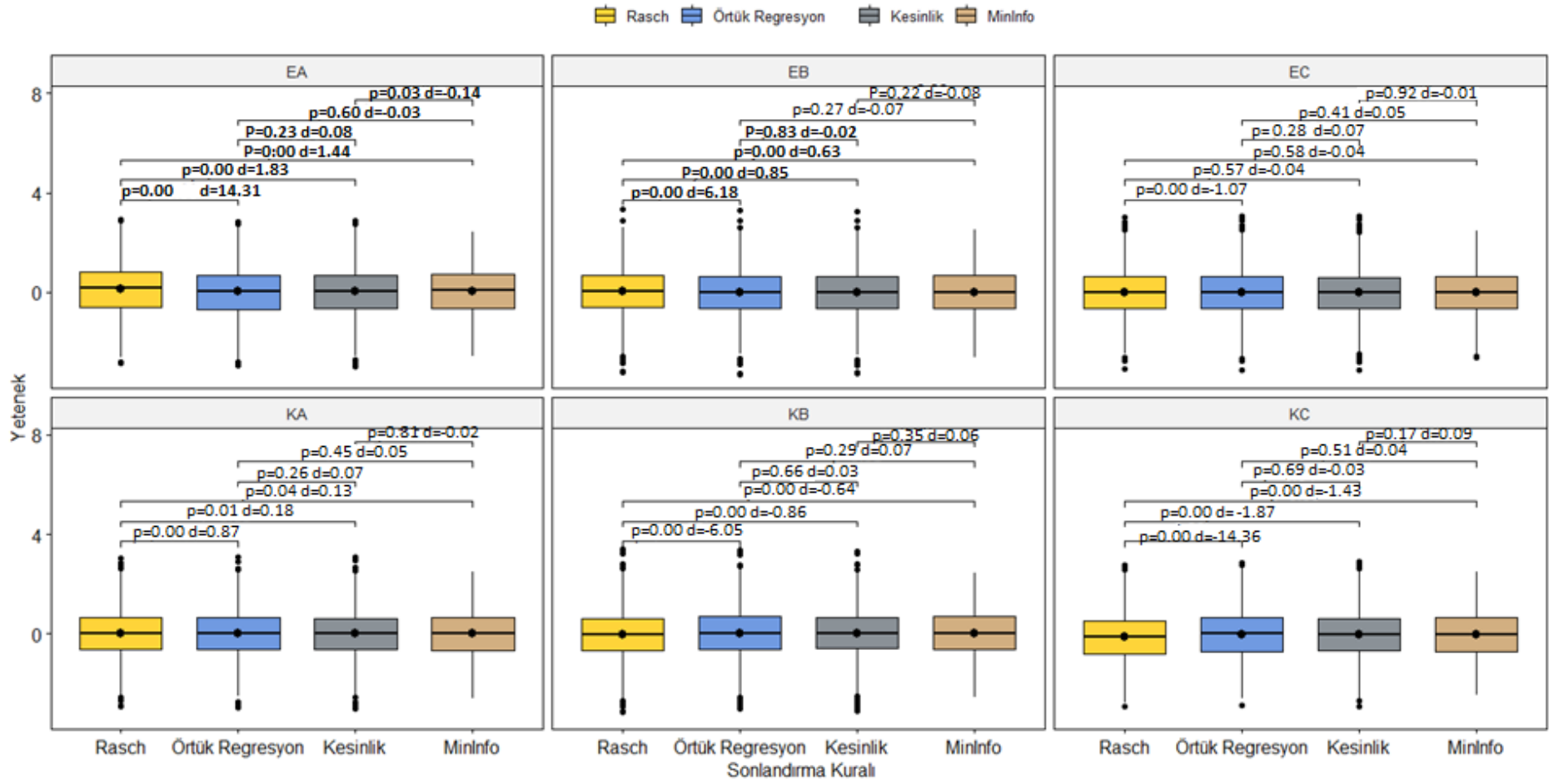
Tablo 16

*ÖR Modeli Parametrelerine Dayalı BBT Simülasyonlarında Alt Gruplar İçin  
Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları*

Alt Grup	Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p	Kısmi $\eta^2$
EA	Model (Greenhouse-Geisser)	1.68	1.58	1.07	407.02	.00*	.63
	Hata	.99	376.45	.00			
EB	Model (Greenhouse-Geisser)	.35	1.62	.21	85.46	.00*	.26
	Hata	.97	386.16	.00			
EC	Model (Greenhouse-Geisser)	.00	1.55	.00	1.01	.35	
	Hata	.95	369.57	.00			
KA	Model (Greenhouse-Geisser)	.01	1.47	.01	3.29	.05*	.01
	Hata	1.02	35.83	.00			
KB	Model (Greenhouse-Geisser)	.33	1.61	.21	88.61	.00*	.27
	Hata	.89	384.39	.00			
KC	Model (Greenhouse-Geisser)	1.67	1.57	1.06	41.68	.00*	.63
	Hata	.97	375.14	.00			

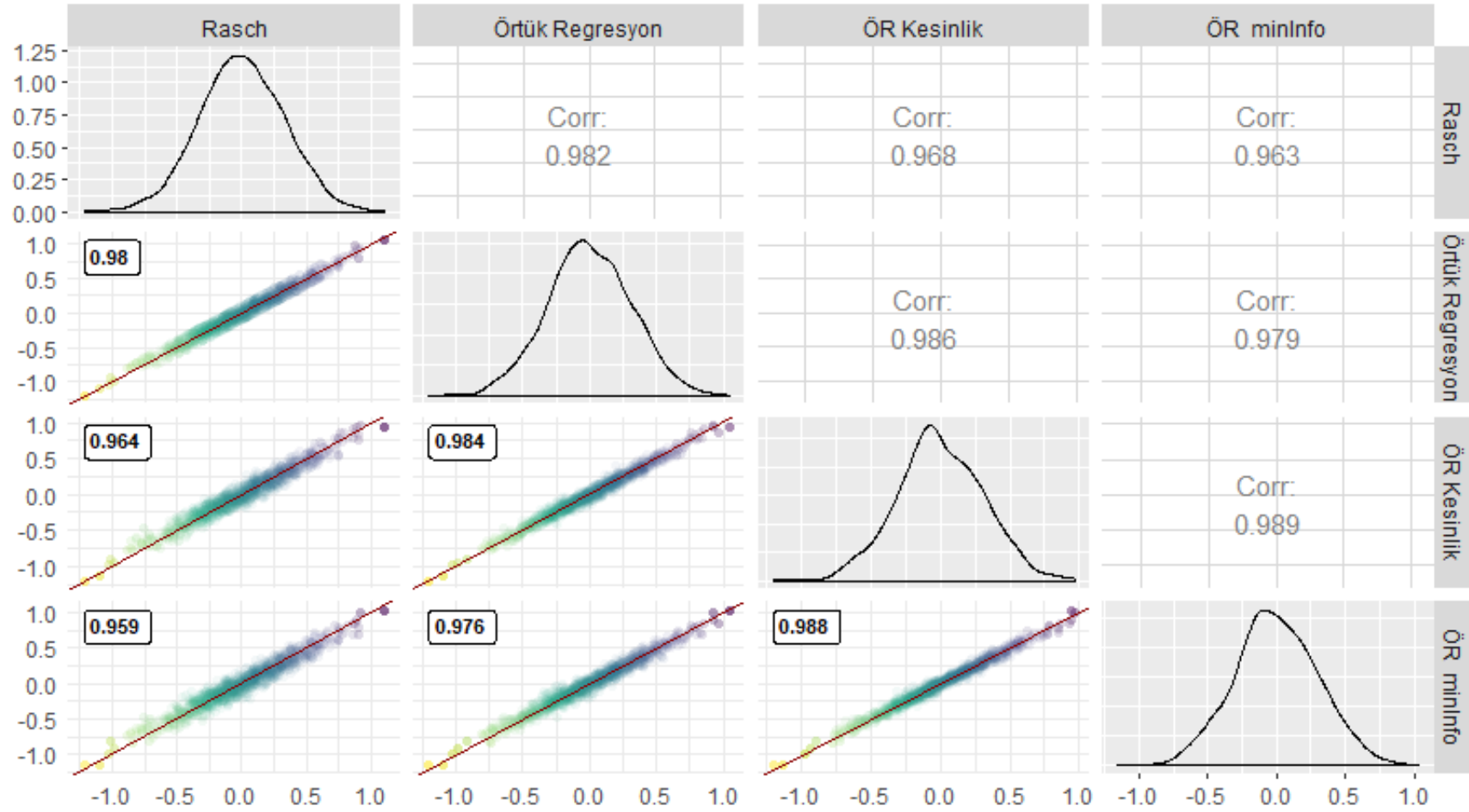
\*p<.05

Alt gruplar dikkate alındığında Rasch ve örtük regresyon modelinden kestirilen önsel yetenek puanları ile BBT simülasyonlarından elde edilen yetenek puanlarının ortalamaları tekrarlı ANOVA ile karşılaştırılmış ve bulguları Tablo 16'da sunulmuştur. Tablo 16'da EC alt grubu ( $F_{1.55, 369.57}=1.01$ ,  $p>.05$ ) dışındaki tüm alt gruplarda manidar bir fark olduğu tespit edilmiştir. Etki büyüklükleri incelendiğinde manidar farkın tespit edildiği gruplardan KA için küçük etki büyüklüğü belirlenmişken diğer alt gruplarda etki büyüklüğü büyüktür. Tekrarlı ANOVA ile tespit edilen manidar farkın hangi alt grup puan ortalamaları arasında olduğu LSD ikili karşılaştırma yöntemi ile incelenmiştir. İkili karşılaştırmada elde edilen manidarlık düzeyleri ve Cohen'in etki büyüklükleri Şekil 15'teki kutu grafiklerinde sunulmuştur. Şekil 15'te sunulan manidarlık düzeyleri incelendiğinde EC dışındaki tüm alt gruplarda, Rasch modelden kestirilen önsel yetenek puanları ortalamalarının kesinlik ve minInfo sonlandırma kuralı ile BBT simülasyonlarından elde edilen yetenek puanları ortalamalarından manidar bir biçimde farklılaştığı görülmektedir.



Şekil 15. ÖR modeli parametrelerine dayalı BBT simülasyonlarında alt gruplar için kestirilen yetenek puanlarına ilişkin ikili karşılaştırma analizi

Şekil 15'te ayrıca örtük regresyon modelinden elde edilen önsel yetenek puanları ortalamasının hiçbir alt grupta kesinlik sonlandırma kuralı ile kestirilen yetenek puanları ortalamalarından manidar bir biçimde farklılaşmadığı görülmektedir. İki sonlandırma kuralına göre gerçekleştirilen simülasyonlarda kestirilen yetenek puanları ortalamalarda ise EA dışındaki hiçbir alt grupta manidar farklılık göstermemektedir. Bu manidar farklılığa ilişkin etki büyüklüğü ( $d=0.14$ ) incelendiğinde farkın pratikte anlamlı olmadığı söylenebilir. Bu durumda madde havuzlarında alt gruplar baz alındığında Rasch model ile örtük regresyon yetenek puanları arasındaki fark BBT simülasyonlarında da gözlenmiştir. Ayrıca her iki sonlandırma kuralının da alt gruplar göz önünde bulundurulmasına rağmen örtük regresyona ilişkin önsel yetenek puanlarına yakın kestirimler yaptığı belirlenmiştir. Örtük regresyon modelinin parametreleri kullanılarak yapılan BBT uygulamalarından elde edilen yetenek puanları ile madde havuzlarından elde edilen yetenek puanları arasındaki ilişki Şekil 16'da sunulmuştur. Rasch modelde olduğu gibi bu modelde de simülasyondan elde edilen yetenek puanları ile önsel yetenek puanları arasında yüksek düzeyde manidar ilişki vardır.



Şekil 16. ÖR modeli parametrelerine dayalı BBT simülasyonlarında alt gruplar için kestirilen yetenek puanlarına ilişkin korelasyon katsayıları

**DLTM için BBT simülasyonları.** Doğrusal lojistik test modelinde (DLTM) kestirilen yetenek puanları ve cevap örüntüleri ile iki farklı sonlandırma kuralına dayalı olarak 10 döngüde gerçekleştirilen post-hoc simülasyonundan elde edilen yetenek puanlarına ilişkin betimsel istatistikler Tablo 17’de görülmektedir.

Tablo 17

*DLTM Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Betimsel İstatistikler*

Sonlandırma Kuralı	n Madde Havuzu	$\bar{x}$	ss	En Küçük	En Büyük	Çarpıklık	Basıklık
Kesinlik (Precision)	1	-.01	.96	-3.25	3.00	.01	-.06
	2	.00	.95	-2.76	3.32	-.01	-.33
	3	.01	.91	-3.08	2.92	.01	.19
	4	.00	.97	-3.22	2.88	-.11	-.21
	5	.00	.92	-2.68	3.20	-.08	-.04
	6	.00	.93	-3.15	3.33	.00	.08
	7	.00	.93	-2.79	2.82	.03	-.13
	8	.00	.93	-2.67	2.82	.03	-.29
	9	.00	.96	-2.96	3.25	.05	-.10
	10	-.01	.96	-2.92	3.10	.01	-.04
MinInfo	1	-.01	.95	-2.62	2.51	.00	-.38
	2	.00	.95	-2.23	2.39	-.03	-.55
	3	.00	.90	-2.46	2.45	.01	-.24
	4	.00	.96	-2.41	2.29	-.06	-.57
	5	.00	.92	-2.38	2.40	-.06	-.41
	6	.00	.93	-3.15	3.33	.01	.08
	7	.00	.93	-2.35	2.46	.01	-.33
	8	.00	.94	-2.42	2.38	.03	-.48
	9	.00	.95	-2.42	2.30	.08	-.45
	10	-.01	.95	-2.45	2.34	-.02	-.39

Tablo 17’ye göre her iki sonlandırma kuralında da birinci ve onuncu madde havuzu dışındaki madde havuzlarında kestirilen yetenek parametresi ortalaması 0’dır. Yine her iki sonlandırma kuralı için de birinci ve onuncu madde havuzları için kestirilen ortalama yetenek puanı -0.1’dir. Benzer bir biçimde yetenek puanlarına ilişkin standart sapmaların da oldukça yakın olduğu gözlenmektedir. Buna göre kesinlik sonlandırma kuralında standart sapma değerleri 0.91 ile 0.97 arasında değişmekteyken minInfo sonlandırma kuralında 0.90 ile 0.95 arasında değerler almaktadır. Ayrıca kesinlik sonlandırma kuralına göre kestirilen en küçük ve en

büyük yetenek puanlarına ilişkin ranjın, minInfo sonlandırma kuralına göre daha geniş olduğu saptanmıştır.

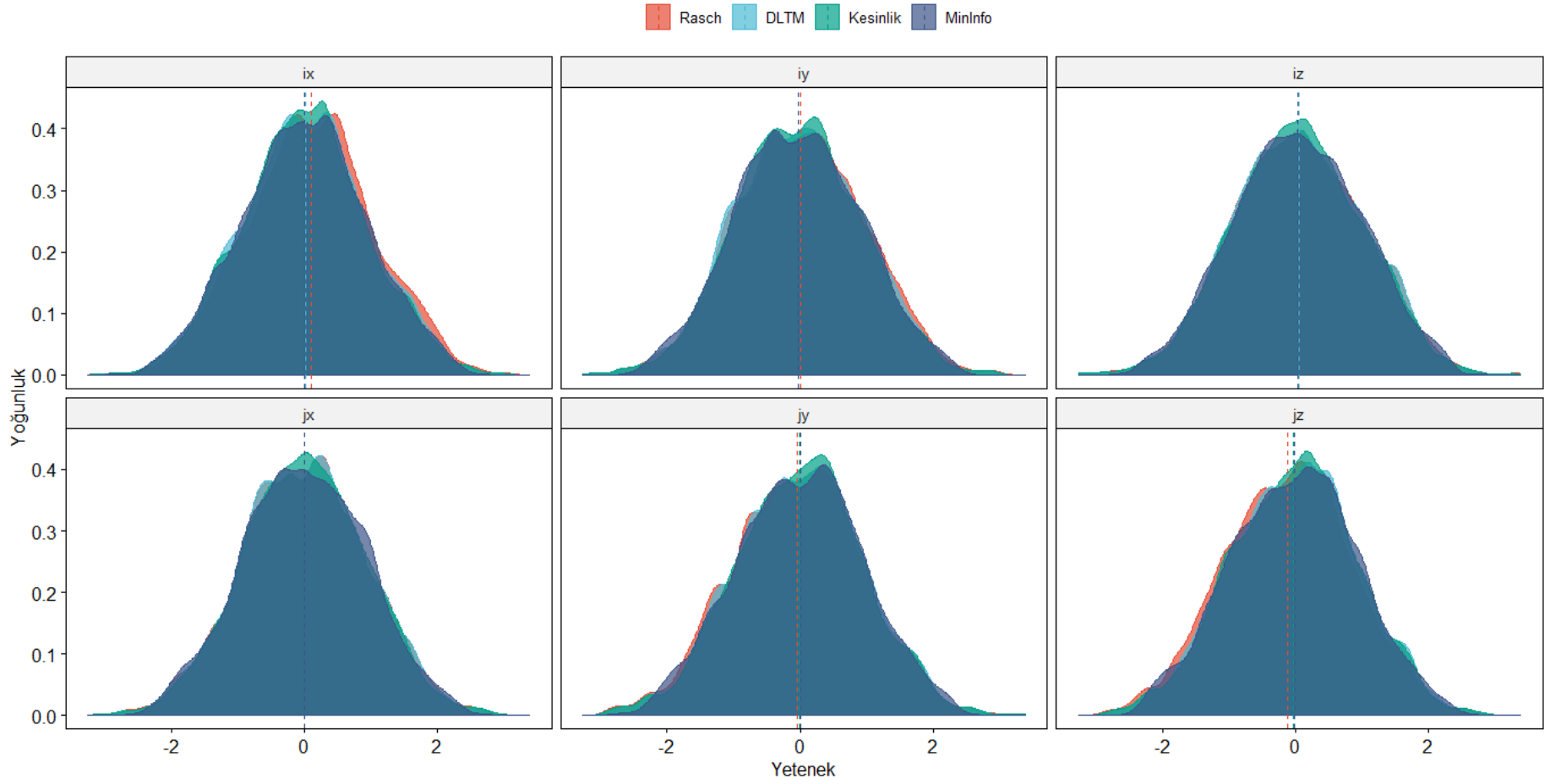
Alt gruplar göz ardı edildiğinde Rasch ve doğrusal lojistik test modelden kestirilen önsel yetenek puanları ortalamaları ve DLTM parametreleri kullanılarak iki sonlandırma kuralına göre yapılan BBT simülasyonlarından elde edilen yetenek puanları ortalamaları tekrarlı tek yönlü ANOVA yardımı ile karşılaştırılmış ve sonuçları Tablo 18’de sunulmuştur.

Tablo 18

*DLTM Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları*

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p
Model (Greenhouse-Geisser)	.00	1.86	.00	.25	.76
Hata	9.00	2678.26	.00		

Mauchly küresellik testinin manidar bulunması ( $w=0.4$ ,  $p<.05$ ) üzerinde Greenhouse-Geisser düzeltmesine dayalı parametrelerin kullanıldığı analiz sonucunda ortalama yetenek puanları arasında manidar bir farklılık olmadığı ( $F_{1.86, 2678.26}=0.25$ ,  $p>.05$ ) belirlenmiştir. Alt gruplar dikkate alındığında tüm madde havuzlarında elde edilen yetenek puanlarının ve post-hoc simülasyonlarından kestirilen yetenek parametrelerinin dağılımı Şekil 17’de sunulmuştur.



Şekil 17. DLTM parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarının dağılımı

Şekil 17’de tüm alt gruplarda bütün modellerde benzer dağılımlar gözlenirken ix ve iy alt gruplarında yetenek parametrelerinin ortalamasının DLTM modele ait önsel yetenek puanlarından ve her iki sonlandırma kuralı ile gerçekleşen simülasyondan elde edilen yetenek parametrelerinden yüksek olduğu görülmektedir. jy ve jz alt gruplarında ise Rasch modele ilişkin puan ortalamalarının daha küçük olduğu saptanmıştır. Bu durumda ix madde grubunu alan katılımcıların Rasch modele göre daha yüksek yetenek ortalamasına sahip olması bu grubun sahip olduğu haksız avantajın kontrol altında tutulmamasından kaynaklanmaktadır. Yine dezavantajlı jz alt grubunun DLTM modelde daha yüksek yetenek puanı ortalamasına sahip olması modelde bu grubun madde özelliklerinden kaynaklı dezavantajının kontrol altında tutulması ile mümkün olmuştur. Şekil 17’de gözlemlenen bu durumun istatistiksel olarak manidar olup olmadığı tekrarlı tek yönlü varyans analizi ile incelenmiş ve sonuçları Tablo 19’da sunulmuştur.

Tablo 19

*DLTM Parametrelerine Dayalı BBT Simülasyonlarında Alt Gruplar İçin Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları*

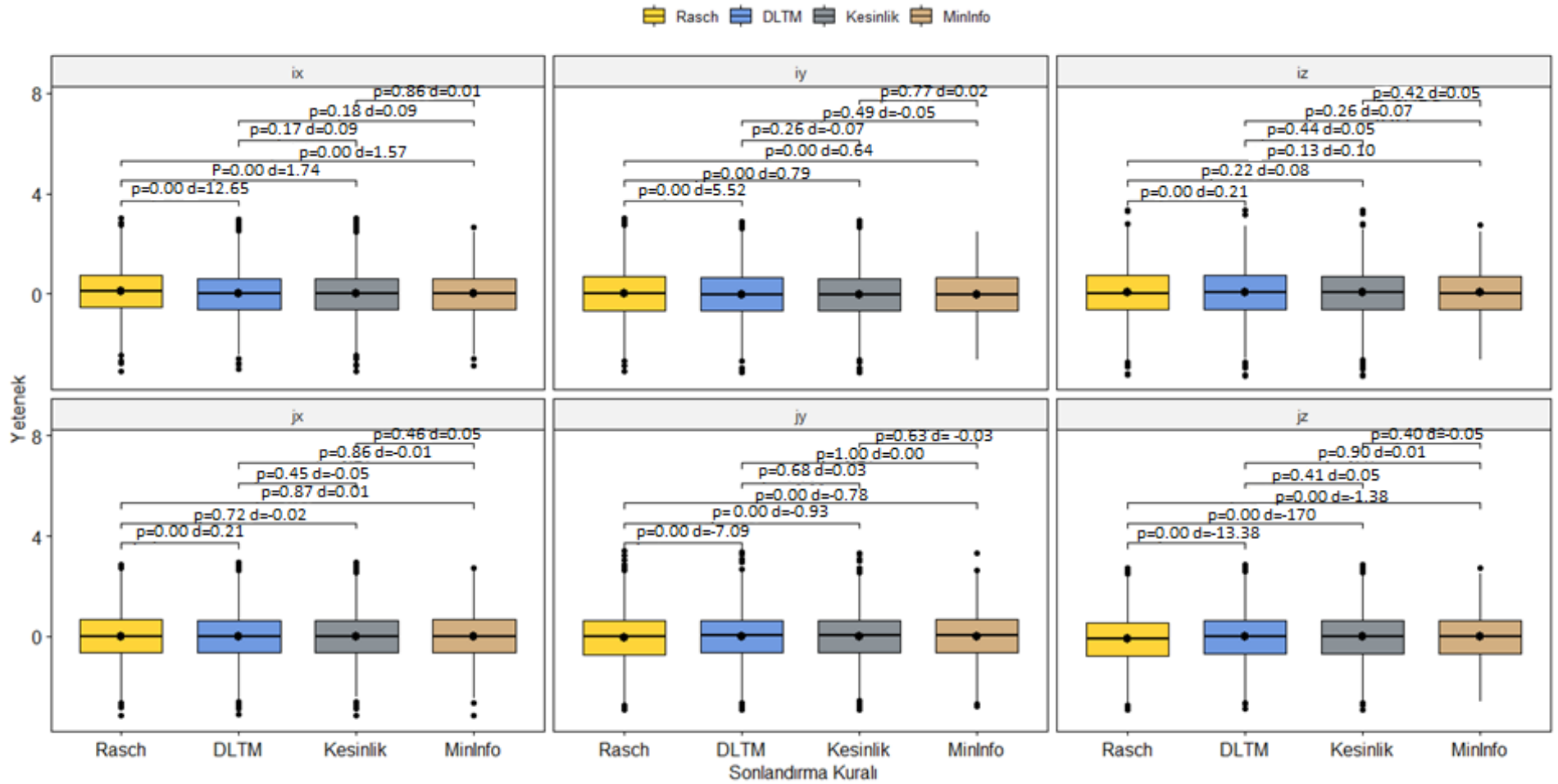
Alt Grup	Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p	Kısmi $\eta^2$
ix	Model (Greenhouse-Geisser)	1.67	1.43	1.17	435.87	.00*	.65
	Hata	.92	342.64	.00			
iy	Model (Greenhouse-Geisser)	.29	1.61	.18	81.40	.00*	.25
	Hata	.85	383.94	.00			
iz	Model (Greenhouse-Geisser)	.01	1.40	.00	1.47	.23	
	Hata	.88	335.65	.00			
jx	Model (Greenhouse-Geisser)	.00	1.47	.00	.24	.72	
	Hata	.80	351.78	.00			
jy	Model (Greenhouse-Geisser)	.41	1.54	.26	114.82	.00*	.32
	Hata	.85	369.20	.00			
jz	Model (Greenhouse-Geisser)	1.43	1.43	1.00	376.58	.00*	.61
	Hata	.90	341.92	.00			

\*p<.05

Tüm alt gruplara ilişkin tekrarlı varyans analizlerinde Mauchly küresellik testinin manidar bulunmasından dolayı Greenhouse-Geisser düzeltmesine dayalı parametreler kullanılmıştır. Buna göre iz ve jx alt grupları haricinde tüm alt gruplarda en az bir yetenek puanı ortalaması manidar bir biçimde farklılaşmaktadır. Manidar fark bulunan alt grupların tamamında farka ilişkin etki büyüklüğü yüksektir. Referans puan olan Rasch modele ait önsel yetenek puanı ortalaması, DL TM’ye ait önsel

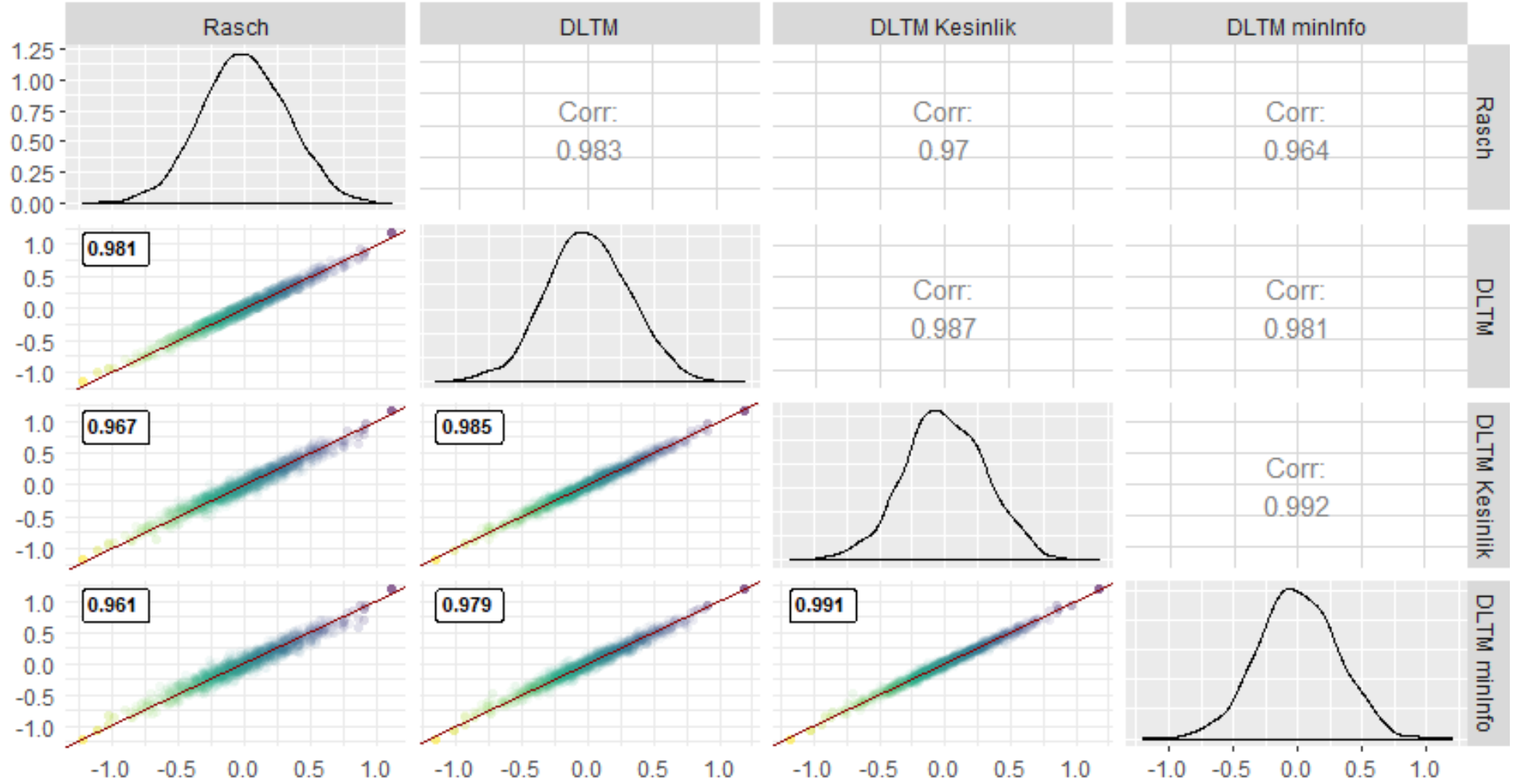
yetenek puanı ortalaması ve DLTM parametreleri kullanılarak gerekleřtirilen BBT simülasyonlarından elde edilen yetenek puanı ortalaması LSD ikili karşılařtırma yöntemi ile karşılařtırılmıřtır. Bu karşılařtırmaya ait bulgular Őekil 18’de sunulmuřtur.





Şekil 18. DLTM parametrelerine dayalı BBT simülasyonlarında alt gruplar için kestirilen yetenek puanlarına ilişkin ikili karşılaştırma analizi

LSD testine ilişkin manidarlık düzeylerinin görüldüğü Şekil 18 incelendiğinde iz ve jx alt grubu dışındaki tüm alt gruplarda, Rasch modelde kestirilen önsel yetenek puan ortalamaları her iki sonlandırma kuralıyla gerçekleştirilen BBT simülasyonlarından kestirilen yetenek puanları ortalamasından manidar bir biçimde farklılaştığı görülmektedir. Ayrıca bu alt gruplar dışındaki tüm alt gruplarda bu manidar farklara ilişkin etki büyüklükleri orta düzeyde veya büyüktür. Buna karşın, DLTM'nin önsel yetenek puan ortalaması ile her iki sonlandırma kuralı ile gerçekleştirilen BBT simülasyonlarından elde edilen yetenek puan ortalamaları hiçbir alt grupta manidar farklılık göstermemektedir. Yine benzer bir biçimde iki sonlandırma kurallarına göre kestirilen puan ortalamalarının farklılaştığı hiçbir alt grup yoktur. Alt gruplar dikkate alındığında Rasch modelde madde havuzundan elde edilen yetenek puan ortalamaları ile DLTM önsel yetenek puanları arasındaki farkın post-hoc simülasyonlarında da gözlemlendiği söylenebilir. Bu durum ÖR modelde olduğu gibi her iki sonlandırma kuralının da alt gruplardaki farklılaşmayı yansıtmada başarılı olduğu şeklinde yorumlanabilir. Şekil 19'da bu sonlandırma kuralları ile DLTM parametrelerinin kullanıldığı simülasyonlardan elde edilen yetenek puanları arasındaki ilişki görülmektedir. Şekilde yine diğer modellerde olduğu gibi tüm yetenek puanları arasında yüksek düzeyde pozitif ilişki olduğu gözlenmektedir.



Şekil 19. DLTM parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarına ilişkin korelasyon katsayıları

**ÖRDLTM için BBT simülasyonları.** Örtük regresyon doğrusal lojistik test modeline (ÖRDLTM) ait önsel madde ve yetenek parametreleri ve cevap örüntülerinin kullanıldığı iki farklı sonlandırma kuralına dayalı olarak 10 döngü olacak şekilde gerçekleştirilen post-hoc simülasyonlarından kestirilen yetenek parametrelerine ait betimsel istatistikler Tablo 20’de sunulmuştur. Tablo 20’de her iki sonlandırma kuralında da kestirilen yetenek puanlarının ortalamasının tüm madde havuzlarında 0’a yakın olduğu görülmektedir. Madde havuzu bazında karşılaştırıldığında ise sadece dördüncü ve sekizinci madde havuzlarında simülasyonlarda kestirilen yetenek parametrelerinin ortalamaları 0.01 birimlik farklılık göstermektedir.

Tablo 20

*ÖRDLTM Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Betimsel İstatistikler*

Sonlandırma Kuralı	Madde Havuzu	n	$\bar{x}_\theta$	$\sigma_\theta$	En Küçük $\theta$	En Büyük $\theta$	Çarpıklık	Basıklık
Kesinlik (Precision)	1	14400	-.01	.95	-3.23	2.92	.00	-.04
	2	14400	.00	.94	-2.80	3.29	-.02	-.30
	3	14400	-.01	.97	-3.08	2.96	-.01	-.14
	4	14400	.03	1.04	-3.21	2.98	-.09	-.43
	5	14400	.01	.99	-2.76	3.25	-.08	-.34
	6	14400	.01	.99	-3.12	3.35	.01	-.23
	7	14400	.00	1.01	-2.80	2.93	.05	-.48
	8	14400	-.03	1.01	-2.81	2.68	-.02	-.56
	9	14400	.00	1.02	-2.90	3.27	.11	-.32
	10	14400	.00	1.02	-3.02	3.13	.03	-.31
MinInfo	1	14400	-.01	.95	-2.64	2.45	.00	-.40
	2	14400	.00	.95	-2.37	2.38	-.02	-.55
	3	14400	-.01	.99	-2.60	2.46	-.02	-.40
	4	14400	.04	1.07	-2.50	2.28	-.06	-.62
	5	14400	.01	1.02	-2.38	2.38	-.06	-.62
	6	14400	.01	1.00	-2.36	2.96	-.01	-.56
	7	14400	.00	1.03	-2.35	2.63	.06	-.57
	8	14400	-.04	1.03	-2.48	2.37	-.04	-.66
	9	14400	.00	1.02	-2.43	2.34	.14	-.58
	10	14400	.00	1.03	-2.48	2.87	.06	-.50

Alt gruplar göz ardı edildiğinde Rasch ve ÖRDLTM’den kestirilen önsel yetenek puanları ortalamaları ve ÖRDLTM parametreleri kullanılarak iki sonlandırma kuralına göre yapılan BBT simülasyonlarından elde edilen yetenek

puanları ortalamaları tekrarlı tek yönlü ANOVA yardımı ile karşılaştırılmış ve sonuçları Tablo 21’de sunulmuştur. Mauchly küresellik testinin manidar bulunması ( $w=0.45$ ,  $p<.05$ ) üzerinde Greenhouse-Geisser düzeltmesine dayalı parametrelerin kullanıldığı analiz sonucunda ortalama yetenek puanları arasında manidar bir fark olmadığı tespit edilmiştir ( $F_{1.91, 275.94}=0.06$ ,  $p>.05$ ).

Tablo 21

*ÖRDLTM Parametrelerine Dayalı BBT Simülasyonlarında Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları*

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p
Model (Greenhouse-Geisser)	.00	1.91	.00	.06	.94
Hata	8.49	275.94	.00		

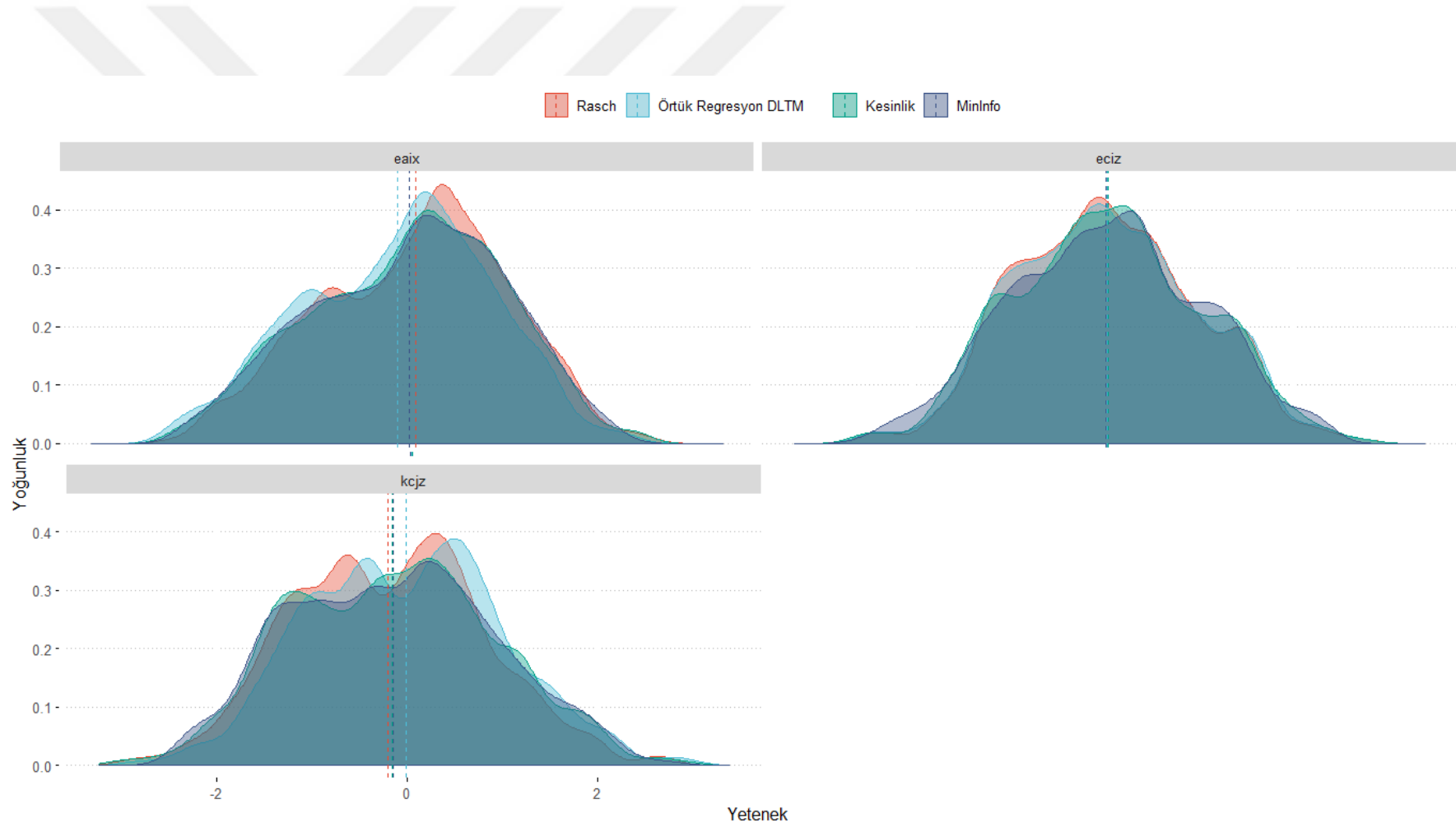
Yetenek parametrelerinin ÖRDLTM’ye ilişkin alt gruplardaki dağılımı Şekil 20’de sunulmuştur. Şekil 20 incelendiğinde alt gruplar göz ardı edildiğinde ortalamalar arasında fark belirlenememişse de alt gruplar dikkate alındığında özellikle Rasch modele ait önsel yetenek puanları ortalamasının diğer ortalamalardan farklılaştığı görülmektedir. Bu farklılığın eaix ve kcjz alt gruplarına ait bireyler için Rasch ve örtük regresyon DLTM’ye ilişkin önsel yetenek puanlarında gözlenmesine rağmen simülasyonlardan kestirilen yetenek puanı ortalamalarının farklılık gösterip göstermediği hipotez testi yardımı ile test edilmiş ve sonuçları Tablo 22’de sunulmuştur. Tüm alt gruplarda Mauchly küresellik testinin manidar bulunması sonucunda Greenhouse-Geisser düzeltmesi ile ortaya çıkan sonuçlar dikkate alınmıştır. Buna göre ECiz alt grubu dışındaki iki alt grupta en az bir yetenek puanı ortalaması diğerlerinden manidar bir şekilde farklılaşmaktadır. Ayrıca bu manidar farka ilişkin etki büyüklüklerinin büyük olduğu ( $\eta^2 = 0.75$ ,  $\eta^2 = 0.82$ ) söylenebilir.

Tablo 22

*ÖRDLTM Parametrelerine Dayalı BBT Simülasyonlarında Alt Gruplar İçin  
Kestirilen Yetenek Puanlarına İlişkin Tekrarlı ANOVA Sonuçları*

Alt Grup	Varyans Kaynağı	Kareler		Kareler		F	p	Kısmi $\eta^2$
		Toplamı	Sd	Ortalaması				
EAix	Model (Greenhouse-Geisser)	.76	1.17	.65	119.17	.00*	.75	
	Hata	.25	45.49	.01				
ECiz	Model (Greenhouse-Geisser)	.01	1.30	.01	1.50	.23		
	Hata	.20	5.65	.00				
KCjz	Model (Greenhouse-Geisser)	.77	1.35	.57	173.31	.00*	.82	
	Hata	.17	52.51	.00				

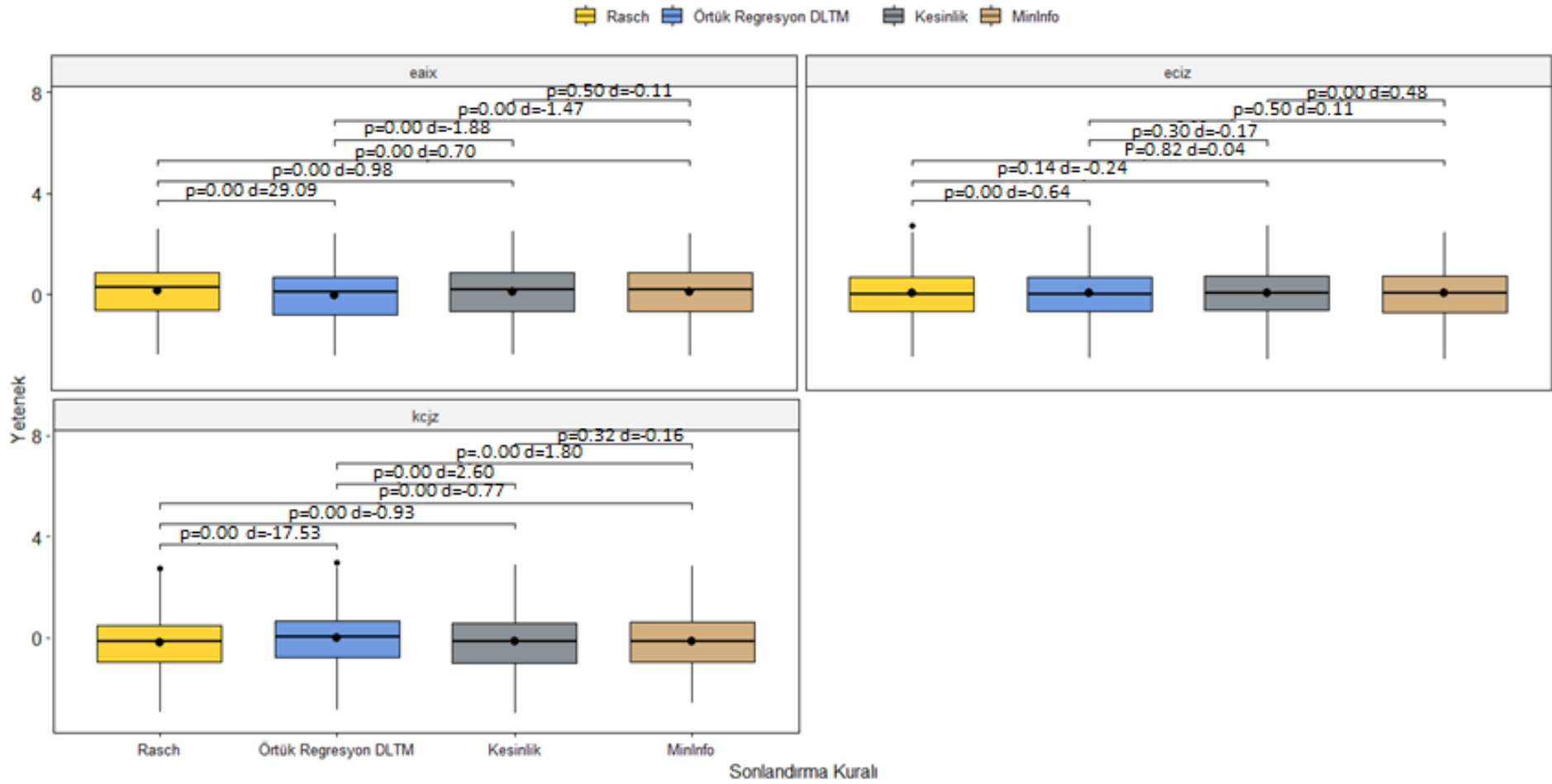
\*p&lt;.05



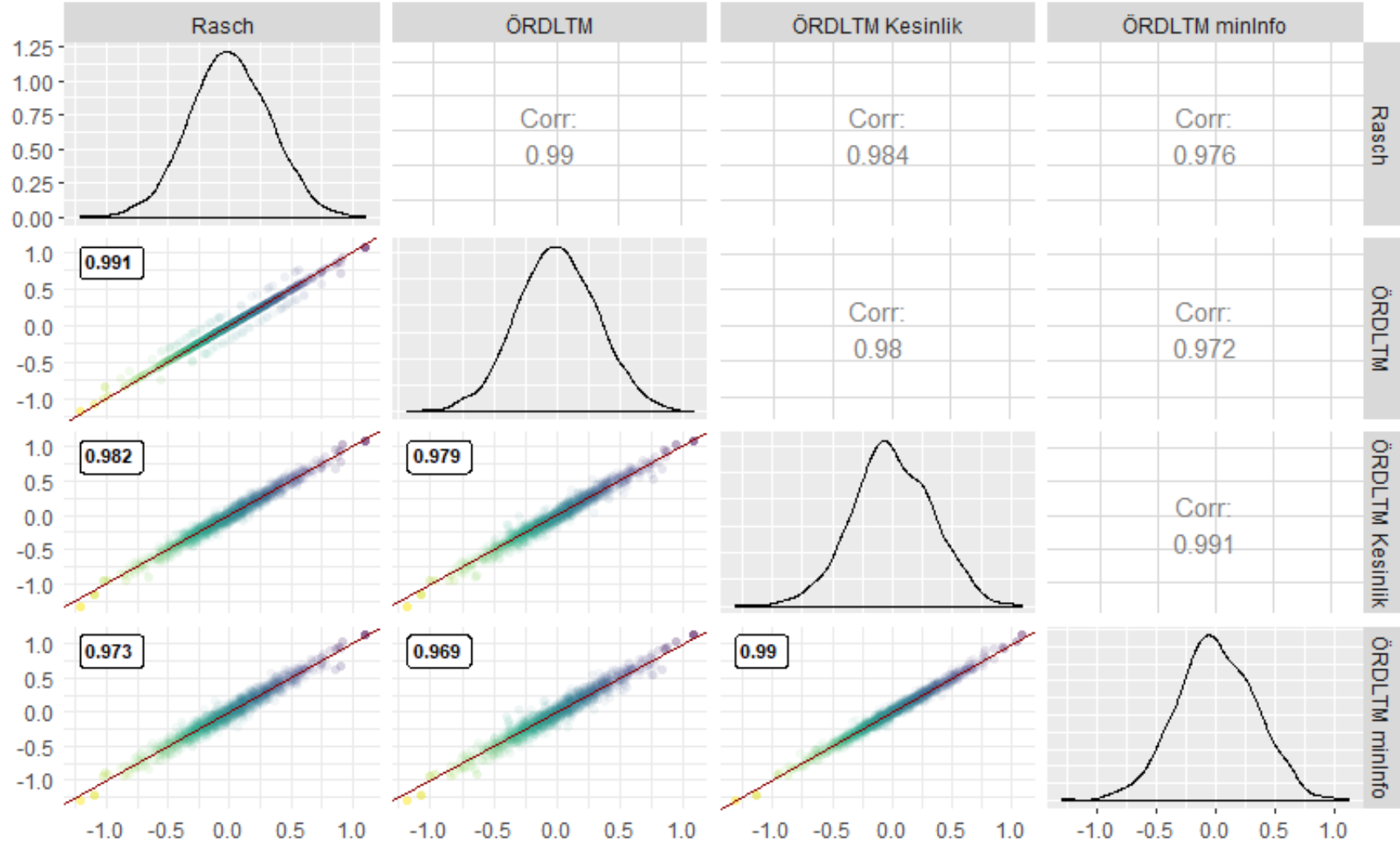
Şekil 20. ÖRDLTM parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarının dağılımı

Tekrarlı ANOVA'da yetenek puanları ortalamaları arasındaki manidar farkın hangi puan çiftleri arasında olduğunu belirlemek için yapılan ikili karşılaştırma analiz sonuçları Şekil 21'de kutu grafikleri yardımı ile sunulmuştur. Şekle 21'e göre Rasch modelde kestirilen önsel yetenek puanları ortalaması, ECiz dışındaki iki alt grupta da BBT simülasyonlarından kestirilen yetenek puan ortalamalarından manidar bir biçimde farklılaşmaktadır. ÖRDLTm'ye ait önsel yetenek parametrelerinin ortalaması ise EAix ve KCjz alt gruplarında her iki sonlandırma kuralına göre yapılan post-hoc simülasyon ortalamalarından farklılaşmaktadır. Yine iki sonlandırma kuralına göre yapılan simülasyondan elde edilen yetenek parametrelerinin ortalaması EAix ve KCjz alt grupları için farklılaşmazken, ECiz alt grubunda farklılaşmaktadır. Manidar olduğu tespit edilen bu farka ilişkin etki büyüklüğü ( $d = 0.48$ ) incelendiğinde, farkın orta düzeyde etki büyüklüğüne sahip görülmektedir.

Şekil 21'de betimlenen bulgulara dayalı olarak ÖRDLTm'den kestirilen önsel yetenek puanları ile BBT kestirimlerinin ortalamasının yakın çıkması beklenirken her iki sonlandırma kuralında da ortalamalar arasında fark olduğu görülmüştür. Bu durumun model-veri uyumu bakımından daha kötü uyum gösteren ÖRDLTm'nin önsel yetenek puanlarının kestiriminde yüksek hatalara sebep olmasından kaynaklandığı söylenebilir. Veya her iki sonlandırma kuralının bu modelde diğer modellere göre daha düşük başarı gösterdiği söylenebilir. Şekil 22'de BBT simülasyonları ile Rasch ve ÖRDLTm'ye ilişkin önsel yetenek puanları arasındaki korelasyon matrisi görülmektedir. Korelasyon matrisine göre tüm kestirimler arasında pozitif ve yüksek düzeyde ilişki vardır.



Şekil 21. ÖRDLTM parametrelerine dayalı BBT simülasyonlarında alt gruplar için kestirilen yetenek puanlarına ilişkin ikili karşılaştırma analizi



Şekil 22. ÖRDLTM parametrelerine dayalı BBT simülasyonlarında kestirilen yetenek puanlarına ilişkin korelasyon katsayıları

## Beşinci Alt Probleme İlişkin Bulgular ve Yorumlar

“Araştırmanın beşinci alt problemi olan “Rasch, örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinden elde edilen parametreler ile iki farklı sonlandırma kuralına göre gerçekleştirilen BBT uygulamalarında kullanılan ortalama madde sayısı nasıldır?” sorusuna ilişkin olarak tüm madde havuzlarında ve tüm döngülerde kullanılan madde sayılarına ait betimsel istatistikler Tablo 23’te sunulmuştur. Tablo 23’e göre her dört modelde de kesinlik sonlandırma kuralına göre yapılan simülasyonlarda kullanılan ortalama madde sayısı 47’ye yakın iken minInfo sonlandırma kuralına göre yapılan simülasyonlarda bu değer en az 39.47 iken en fazla 42.22’dir. Yine tüm modellerde kesinlik sonlandırma kuralının kullanıldığı simülasyonlarda en az 42 madde kullanılmışken en fazla 100 madde, yani madde havuzunun tamamı kullanılmıştır. MinInfo sonlandırma kuralına göre yapılan simülasyonlarda ise bazı bireylerin testleri sadece altı madde ile sonlandırılmışken madde havuzunun tamamının uygulandığı bireylere de rastlanmıştır. Uygulanan madde sayıları arasındaki farklılığın bir sonucu olarak tüm modellerde kesinlik sonlandırma kuralına göre yapılan simülasyonlarda kullanılan madde sayısına ilişkin ortalama standart sapma 10 civarında iken minInfo sonlandırma kuralında bu değer 12.73 ile 15.61 arasında değerler almaktadır.

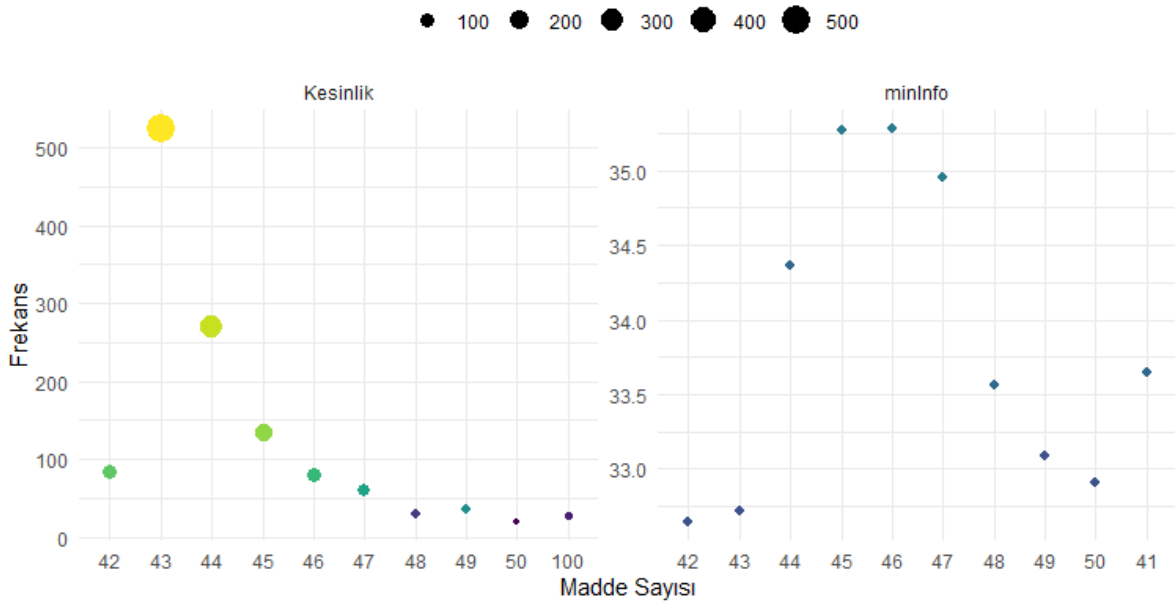
Tablo 23

### *BBT Simülasyonlarında Model Bazında Kullanılan Ortalama Madde Sayısı*

Model	Sonlandırma Kuralı	Ortalama Madde Sayısı	ss	En Küçük	En Büyük
Rasch	Kesinlik	47.20	9.94	42	100
	minInfo	42.22	15.61	6	100
Örtük Regresyon	Kesinlik	47.11	9.91	42	100
	minInfo	39.97	13.02	6	100
Doğrusal Lojistik Test Model	Kesinlik	47.11	9.92	42	100
	minInfo	4.77	12.73	6	100
Örtük Regresyon DLTM	Kesinlik	47.59	1.26	42	100
	minInfo	39.47	13.14	7	100

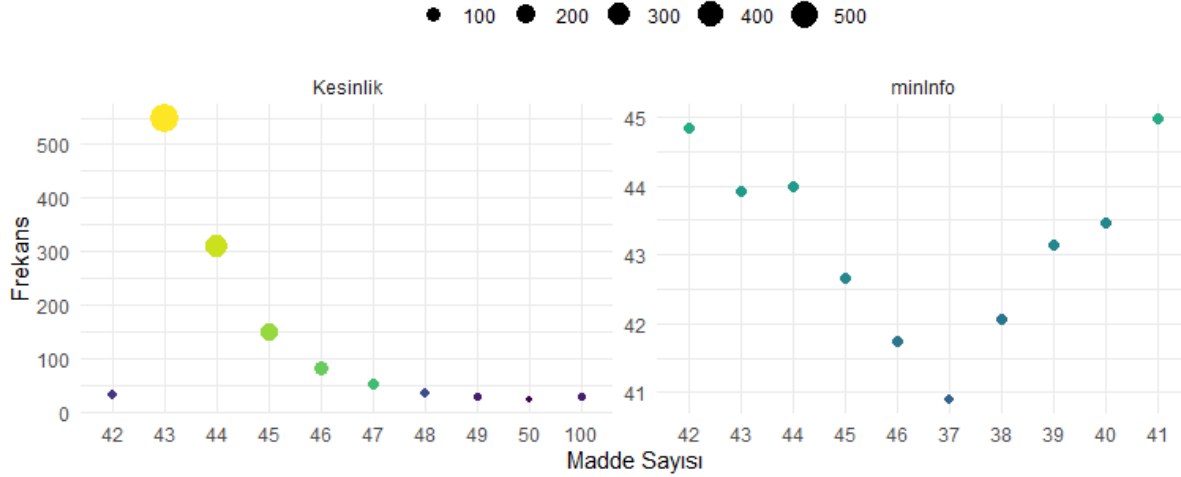
Simülasyonlarda model ve sonlandırma kuralına göre katılımcıların ne kadarının ortalama kaç madde ile test edildiğine ilişkin detaylar balon grafikleri

yardımı ile sunulmuştur. Şekil 23'te Rasch model parametreleri ile iki sonlandırma kuralına göre yapılan simülasyonlarda en fazla rastlanan 10 test uzunluğu görülmektedir. Buna göre kesinlik sonlandırma kuralında 500'den fazla katılımcının ortalama test uzunluğu 43 madde iken 250'den fazla katılımcıya 44 madde uygulanmıştır. Öte yandan minInfo sonlandırma kuralında en sık karşılaşılan test uzunluğu 45 ve 46'dır. Bu sonlandırma kuralında, kesinlik sonlandırma kuralının aksine herhangi iki test uzunluğundaki birey sayısı oldukça yakındır. Örneğin 42 maddenin uygulandığı birey sayısı yaklaşık olarak 33 iken, 45 maddenin uygulandığı birey sayısı yaklaşık 36'dır.



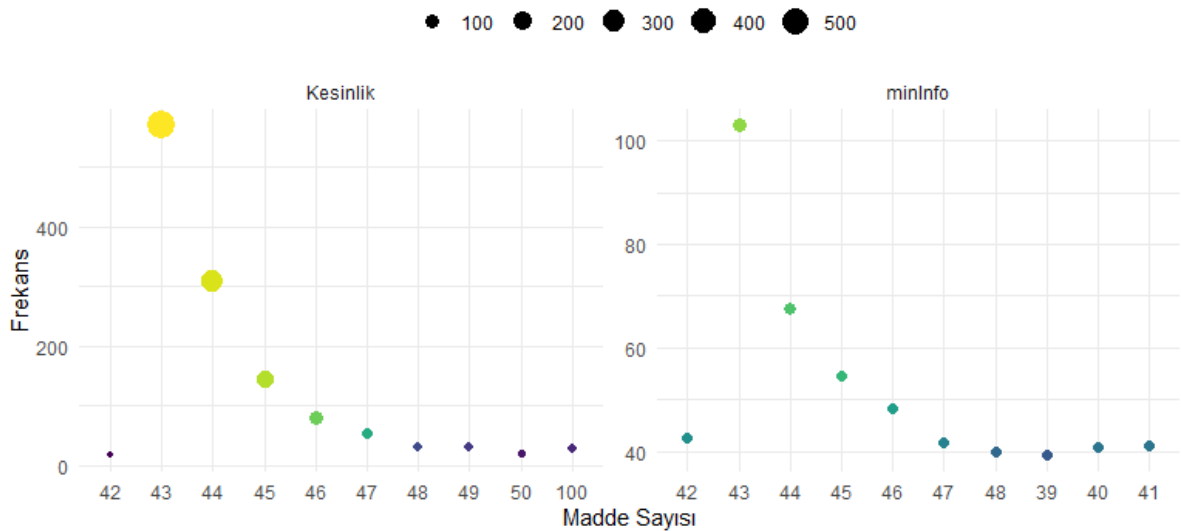
Şekil 23. Rasch model parametrelerine dayalı BBT simülasyonlarında en sık karşılaşılan test uzunluğu frekansları

Rasch modele benzer bir duruma örtük regresyon modeline ait parametrelerin kullanıldığı simülasyon sonuçlarında da karşılaşılmaktadır. Bu sonuçlara Şekil 24'te yer verilmektedir. Şekil 24 incelendiğine bir önceki modele benzer sonuçlar göze çarpmaktadır. Kesinlik sonlandırma kuralında yine en sık rastlanan test uzunluğu yaklaşık 550 katılımcı ile 43 madde iken, ikinci en sık karşılaşılan test uzunluğu yaklaşık 300 kişi ile 44 maddedir. MinInfo sonlandırma kuralında ise en fazla karşılaşılan test uzunluğu yaklaşık 45 kişi ile 41 maddedir. Yine farklı uzunluklarda maddenin uygulandığı birey sayıları birbirine yakındır.



Şekil 24. ÖR modeli parametrelerine dayalı BBT simülasyonlarında en sık karşılaşılan test uzunluğu frekansları

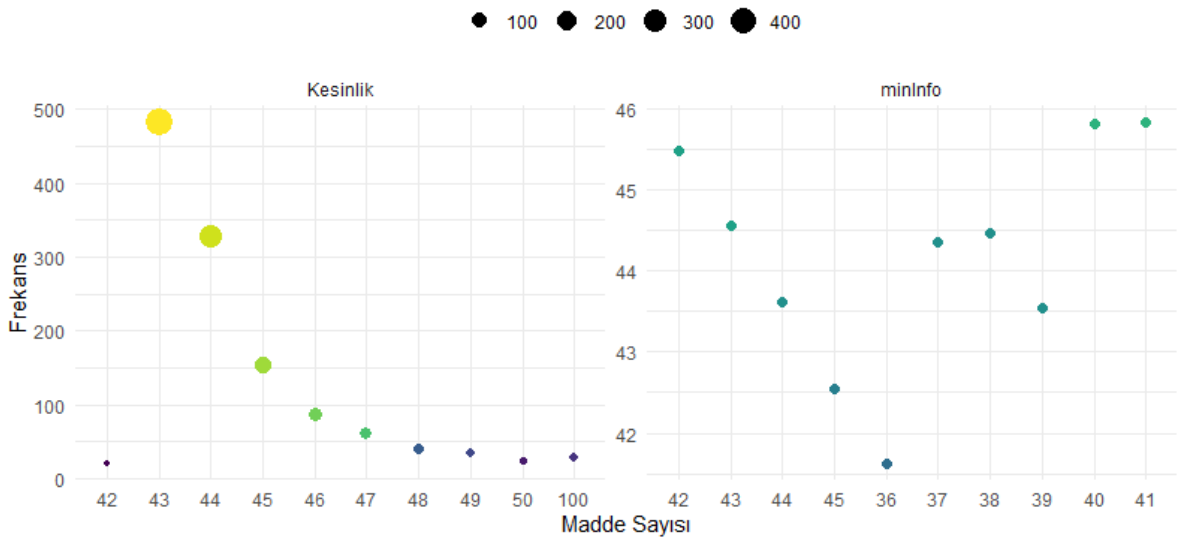
Şekil 25'te doğrusal lojistik test modelin parametreleri ile gerçekleştirilen simülasyonlardaki ortama test uzunlukları görülmektedir. Bu modelde de kesinlik sonlandırma kuralında en sık karşılaşılan test uzunluğu yaklaşık 550 kişi ile 43 maddedir. Buna karşın minInfo sonlandırma kuralında 100'den fazla kişinin testi 43 madde ile sonlandırılmışken yaklaşık 70 katılımcıya ortalama 44 madde uygulanmıştır. Yine farklı test uzunluklarındaki katılımcı sayıları kesinlik sonlandırma kuralına kıyasla birbirine yakındır.



Şekil 25. DLTM parametrelerine dayalı BBT simülasyonlarında en sık karşılaşılan test uzunluğu frekansları

Son model olan örtük regresyon doğrusal lojistik test modelden elde edilen ortalama test uzunlukları Şekil 26'da sunulmuştur. Diğer modellerde kesinlik

sonlandırma kuralında de en sık karşılaşılan 43 maddeden oluşan test uzunluğu bu model için de geçerlidir. Yani katılımcıların yaklaşık üçte birine ortalama 43 madde uygulanmışken, yaklaşık 350 katılımcının testi 44 maddeden oluşmaktadır. MinInfo sonlandırma kuralında ise en sık karşılaşılan test uzunluğunun 40, 41 ve 42 olduğu görülmektedir. Bu modelde de farklı test uzunluklarına sahip birey sayıları oldukça yakındır. Dikkat çeken bulgulardan biri ortalama test uzunluğunun modeller arasında çok fazla değişim göstermediği, bunda daha çok sonlandırma kuralının etkili olduğudur. Örneğin kesinlik sonlandırma kuralı her dört modelde de benzer ortalama test uzunluğu sağlamışken, minInfo da paralel bir davranış sergilemiştir.



Şekil 26. ÖRDLTm parametrelerine dayalı BBT simülasyonlarında en sık karşılaşılan test uzunluğu frekansları

### Altıncı Alt Probleme İlişkin Bulgular ve Yorumlar

Çalışmanın altıncı alt problemi olan “Rasch, örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinden elde edilen parametreler ile iki farklı sonlandırma kuralına göre gerçekleştirilen BBT uygulamalarındaki madde kullanım sıklığı nasıldır?” sorusuna ilişkin olarak elde edilen bulgular Tablo 24’te sunulmuştur. Tablo 24’e göre ilk üç modelde kesinlik sonlandırma kuralına göre yapılan simülasyonlarda herhangi bir madde, döngülerin ortalama %47’sinde kullanılmışken son modelde bu oran %48’dir. Buna karşın minInfo sonlandırma kuralına dayalı simülasyonlarda madde havuzlarındaki herhangi bir maddenin kullanım sıklığı örtük regresyon ve ÖRDLTm için %40, DLTM için %41 ve Rasch model için %43’tür. Yine kesinlik sonlandırma kuralına göre

yapılan simülasyonlarda en az uygulanan madde/maddeler simülasyonların yaklaşık %16'sında tekrar kullanılmışken, minInfo sonlandırma kuralına göre gerçekleşen simülasyonlarda en düşük madde kullanım sıklığı oranı %9 ile %11 arasında değişmektedir. Ayrıca en fazla tekrar kullanım oranına sahip madde/maddeler kesinlik sonlandırma kuralında simülasyon tekrarlarının %78 ile %80'i arasında değişen oranlarda kullanılmışken, minInfo'ya göre yapılan simülasyonlarda Rasch modelde %80, diğer modellerde %73 oranında kullanılmıştır.

Tablo 24

*BBT Simülasyonlarında Model Bazında Madde Kullanım Sıklığı*

Model	Sonlandırma Kuralı	Havuz Sayısı	Madde Sayısı	$\bar{x}_{MKS}$ %	$\sigma_{MKS}$ %	En Küçük %	En Büyük %
Rasch	Kesinlik	10	100	47.00	15.00	16.00	78.00
	minInfo	10	100	43.00	13.00	11.00	80.00
Örtük Regresyon	Kesinlik	10	100	47.00	16.00	16.00	80.00
	minInfo	10	100	40.00	13.00	9.00	73.00
Doğrusal Lojistik Test Model	Kesinlik	10	100	47.00	16.00	16.00	80.00
	minInfo	10	100	41.00	13.00	10.00	73.00
Örtük Regresyon DLTM	Kesinlik	10	100	48.00	16.00	17.00	79.00
	minInfo	10	100	40.00	12.00	11.00	73.00

Model ve sonlandırma kuralı bazında eşdeğişkenlerin madde güçlükleri üzerinde etkisinin olduğu ilk yirmi maddenin kullanım sıklığına ilişkin detaylar Tablo 25' te ve Şekil 27'de sunulmuştur. Tablo 25'e göre her iki sonlandırma kuralına göre ilk yirmi maddenin ortalama kullanım sıklığının en yüksek olduğu model Rasch modeli iken en düşük ortalama kullanım sıklığı ÖRDLTM modelinden elde edilmiştir.

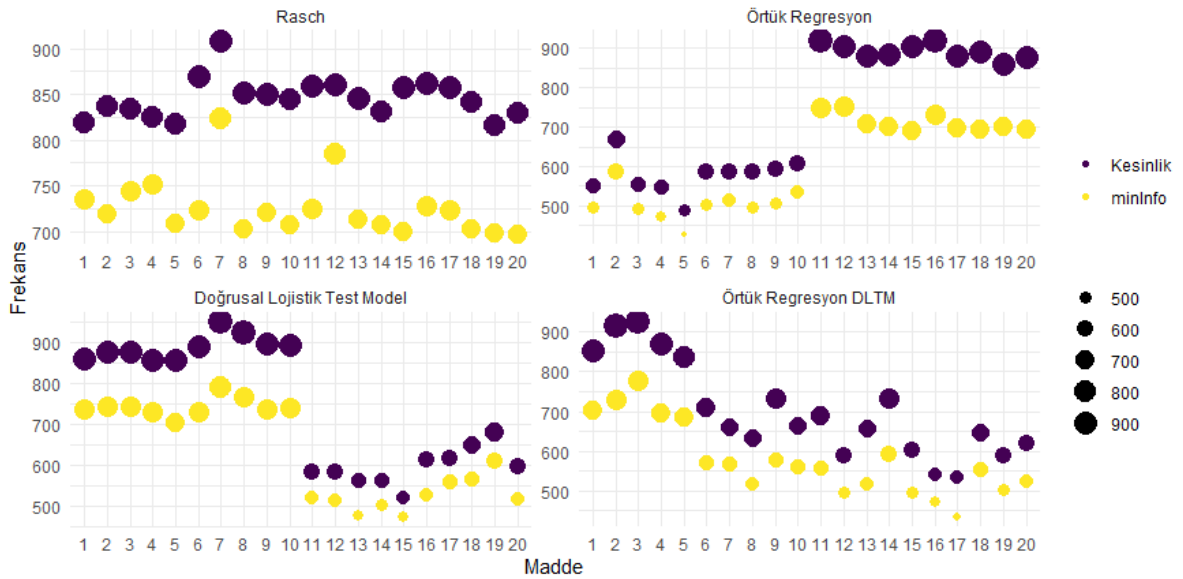
Tablo 25

*BBT Simülasyonlarında Model Bazında İlk 20 Maddeye Ait Madde Kullanım Sıklığı*

Model	Sonlandırma Kuralı	Havuz Sayısı	Madde Sayısı	$\bar{x}_{MKS}$ f	$\sigma_{MKS}$ f	En Küçük f	En Büyük f
Rasch	Kesinlik	10	20	845.61	21.82	815.96	908.8
	minInfo	10	20	725.76	31.38	696.06	823.38
Örtük Regresyon	Kesinlik	10	20	734.53	165.27	489.37	919.74
	minInfo	10	20	606.96	111.93	425.87	751.69
Doğrusal Lojistik Test Model	Kesinlik	10	20	742.75	154.98	52.38	95.78
	minInfo	10	20	633.75	115.57	472.54	789.53
Örtük Regresyon DLTM	Kesinlik	10	20	7.04	12.55	535.01	924.72
	minInfo	10	20	576.51	93.90	434.45	778.83

Şekil 27’de ayrıca tüm modellerde kesinlik sonlandırma kuralına göre yapılan simülasyonlarda madde kullanım sıklığı minInfo sonlandırma kuralına göre gerçekleşen simülasyonlardakinden daha yüksek olduğu görülmektedir. Ayrıca her iki sonlandırma kuralına göre yapılan simülasyonlarda birey değişkenlerinin etkili olduğu ilk 10 maddeye ilişkin madde kullanım sıklığı örtük regresyon modelinde daha küçük iken, madde eşdeğişkenlerinin etkili olduğu 11-20 maddelere ait kullanım sıklıkları doğrusal lojistik test modelde Rasch modele kıyasla daha küçüktür.

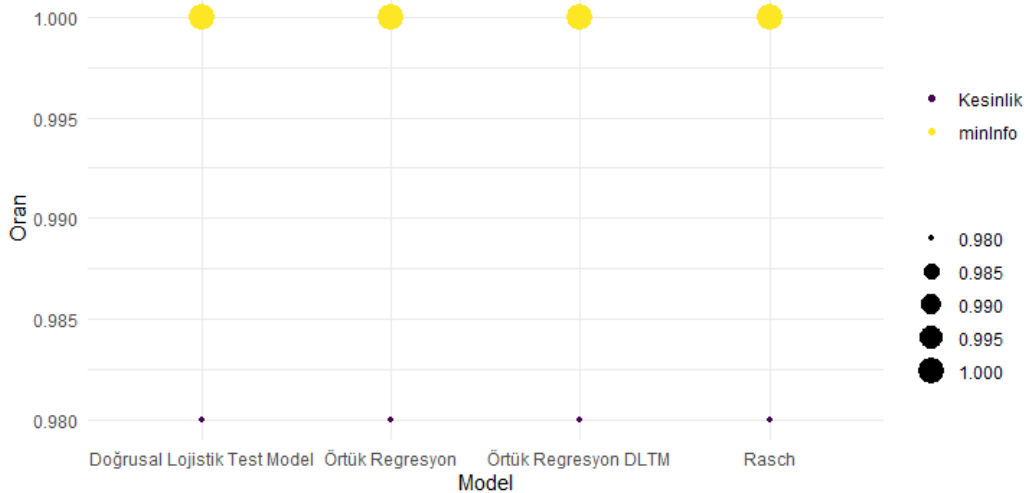
Test uzunluğunda olduğu gibi madde kullanım sıklığı da genel anlamda kestirim modeline göre farklılık göstermezken sonlandırma kuralından etkilendiği söylenebilir. MinInfo sonlandırma kuralının yaklaşık %7 oranında daha düşük madde kullanım sıklığına sebep olduğu görülmektedir. Bu durumda minInfo sonlandırma kuralının daha başarılı olduğu söylenebilir. Genel ortalamada gözlenmeyen kestirim modeli etkisinin ilk 20 maddede bariz bir şekilde gözlemlendiği görülmüştür. Örneğin birey değişkenlerinin etkili olduğu ilk 10 maddenin ÖR modelindeki kullanım sıklığı bariz bir şekilde Rasch modelden daha küçüktür. Yine madde değişkenlerinin etkili olduğu 11-20. maddelerin kullanım sıklığının DLTM modelde Rasch modele göre düştüğü gözlenmiştir. Bu bakımdan AMTM’nin en azından iki modelinde eşdeğişkenlerin etkili olduğu maddelerin kullanım sıklığı üzerinde pozitif etki yarattığı söylenebilir.



Şekil 27. BBT simülasyonlarında model bazında ilk 20 maddeye ait madde kullanım sıklığı

## Yedinci Alt Probleme İlişkin Bulgular ve Yorumlar

“Araştırmanın yedinci alt problemi “Rasch, örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinden elde edilen parametreler ile iki farklı sonlandırma kuralına göre gerçekleştirilen BBT uygulamalarında testi başarı ile sonlandırılan birey oranları nasıldır?” problemi kapsamında elde edilen bulgular Şekil 28’de sunulmuştur. Şekil 28’e göre tüm modellerde kesinlik sonlandırma kuralına dayalı simülasyonlarda testi başarılı bir biçimde sonlandırılan katılımcı oranı ortalama olarak 0.98’dir. MinInfo sonlandırma kuralına göre yapılan simülasyonlarda ise tüm katılımcıların testleri tüm madde havuzu kullanılmadan yani başarı ile sonuçlandırılmıştır. Başka bir ifade ile testi başarı ile sonuçlanan bireylerin oranı 1’dir. Her ne kadar minInfo sonlandırma kuralı daha başarılı gözükse de bu durum tamamen bu yöntemlerin doğasından kaynaklıdır. MinInfo sonlandırma kuralının her durumda %100 başarı göstermesi doğaldır çünkü kural gereği belirlenen kesme puanından yüksek bilgi sunan madde kalmayınca test otomatik olarak sonlandırılır. Yani madde havuzunun tamamının kullanılabilmesi için kesme puanının en küçük madde bilgi düzeyine göre ayarlanması gerekmektedir. Bu bakımdan modeller arasındaki farklılık kesinlik kuralı üzerinden yorumlanırsa bu oranın modellerden kaynaklı farklılaşmadığı söylenebilir. Yani başarı ile testi sonlandırılan birey oranı kestirim modelinden etkilenmemiştir.



Şekil 28. BBT simülasyonlarında model bazında başarı ile sonlandırılan test oranı

## Sekizinci Alt Probleme İlişkin Bulgular ve Yorumlar

“Rasch, örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinden elde edilen parametreler ile iki farklı sonlandırma kuralına göre gerçekleştirilen BBT uygulamalarındaki ortalama test örtüşme (overlap) oranı nasıldır?” alt problemi kapsamında 10 madde havuzunun her birinde dört istatistik model ile elde edilen parametreler ve cevap örüntüleri ile 10 tekrar olacak şekilde yapılan simülasyonlarda kullanılan ortalama ortak madde oranları belirlenmiş ve Tablo 26’da sunulmuştur.

Tablo 26

### BBT Simülasyonlarında Test Örtüşme Oranı

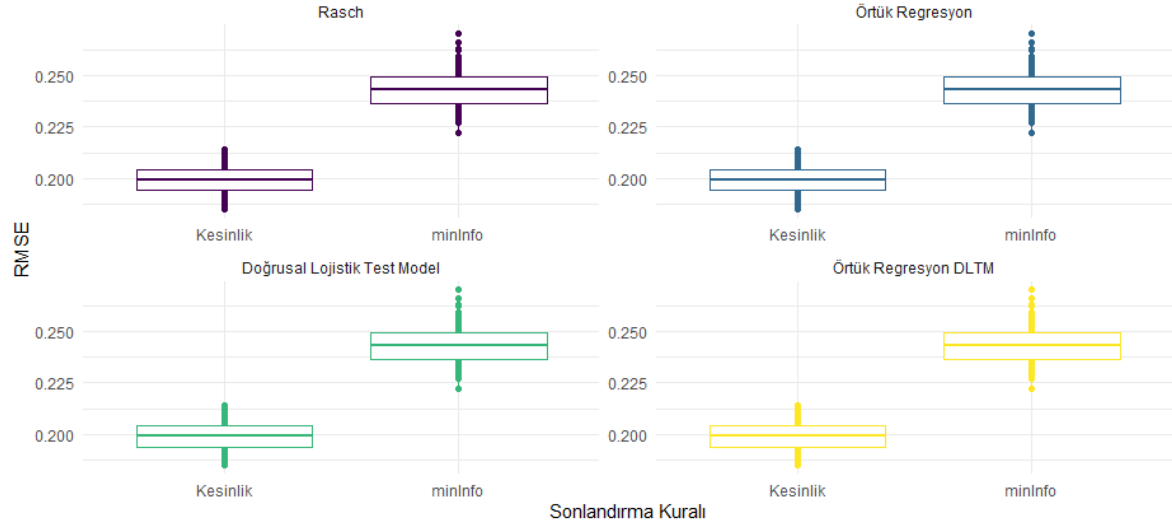
Model	Sonlandırma Kuralı	Havuz Sayısı	Madde Sayısı	$\bar{x}_{T00}$ %	$\sigma_{T00}$ %	En Küçük %	En Büyük %
Rasch	Kesinlik	10	100	52.00	1.00	51.00	53.00
	minInfo	10	100	46.00	1.00	44.00	48.00
Örtük Regresyon	Kesinlik	10	100	53.00	1.00	51.00	54.00
	minInfo	10	100	44.00	1.00	42.00	45.00
Doğrusal Lojistik Test Model	Kesinlik	10	100	53.00	1.00	52.00	54.00
	minInfo	10	100	45.00	3.00	42.00	53.00
Örtük Regresyon DLTM	Kesinlik	10	100	53.00	1.00	51.00	54.00
	minInfo	10	100	43.00	1.00	41.00	45.00

Tablo 26’ya göre tüm modellerde kesinlik sonlandırma kuralına göre yapılan simülasyonlarda maddelerin yaklaşık olarak ortalama %53’ü ortak kullanılmışken, minInfo sonlandırma kuralında bu oran %43 ile %46 arasında değişmektedir. Modeller arasında sonlandırma kuralına bağlı olmaksızın test örtüşme oranları yaklaşık %1-%2 farklılık gösterirken, sonlandırma kuralına göre orandaki farklılık %10’u bulabilmektedir. Test örtüşme oranı üzerinde modellerin büyük bir etkisinin olmadığı, sonlandırma kuralının daha fazla etkili olduğu söylenebilir.

## Dokuzuncu Alt Probleme İlişkin Bulgular ve Yorumlar

“Rasch, örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinden elde edilen parametreler ile iki farklı sonlandırma kuralına göre gerçekleştirilen BBT uygulamalarında ortalama RMSE değerleri nasıldır?” alt problemine ilişkin olarak elde edilen ortalama RMSE değerleri Şekil 29’da görülmektedir. Her dört modelde de ortalama RMSE değerleri kesinlik

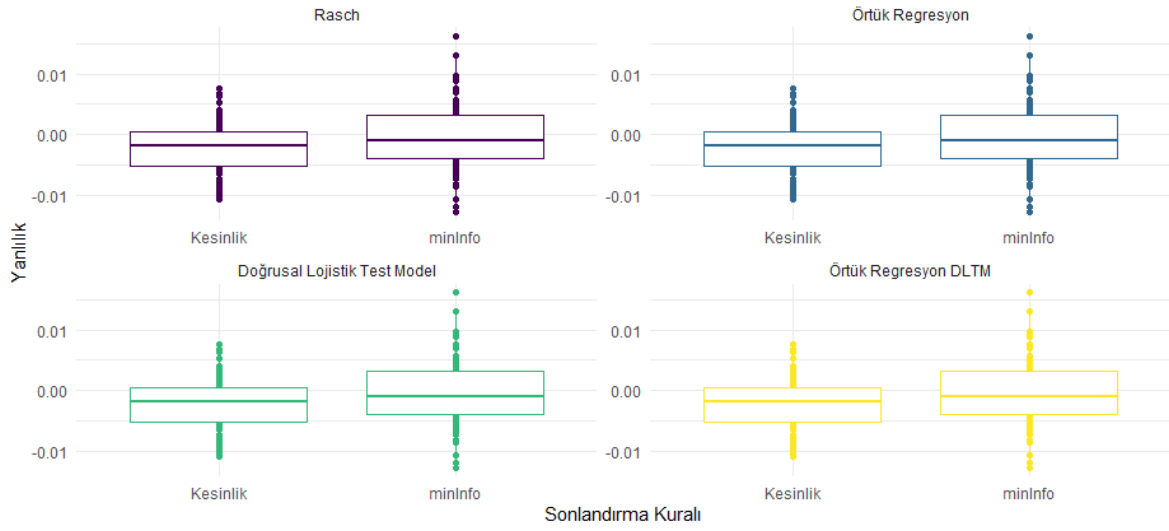
sonlandırma kuralında 0.2 civarında iken minInfo için her dört modelde de 0.24 kestirilmiştir. Bu durumun yine sonlandırma kurallarının işleyişinden kaynaklandığı söylenebilir. Çünkü kesinlik sonlandırma kuralında tüm katılımcılar için 0.2 kesme noktası olarak kullanıldığından, algoritma testleri bu değerden daha küçük hataya ulaştığı ana kadar devam ettirir. MinInfo sonlandırma kuralında ise sonlandırma bilgi fonksiyonuna bağlıdır.



Şekil 29. BBT simülasyonlarında model bazında ortalama RMSE değerleri

### Onuncu Alt Probleme İlişkin Bulgular ve Yorumlar

Araştırma kapsamında “Rasch, örtük regresyon, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerinden elde edilen parametreler ile iki farklı sonlandırma kuralına göre gerçekleştirilen BBT uygulamalarında ortalama yanlılık (bias) değerleri nasıldır?” alt problemi doğrultusunda tüm modellerde ve sonlandırma kurallarında elde edilen ortalama yanlılık değerleri Şekil 30’da görselleştirilmiştir. Şekle göre tüm kestirimlerde minInfo sonlandırma kuralı kesinlik sonlandırma kuralından daha düşük yanlılık göstermiştir. Kestirim modeli fark etmeksizin tüm modellerde kesinlik sonlandırma kuralı ile daha küçük yanlılık elde edilmiştir. Yine kesinlik sonlandırma kuralındaki kestirimlerin yanlılık değerlerinin hepsinin negatif olduğu gözlenmektedir. Bu durumda BBT simülasyonlarında kestirilen yetenek puanlarının önsel puanlardan daha küçük kestirildiği söylenebilir.



Şekil 30. BBT simülasyonlarında model bazında ortalama yanlilik değerleri



## Bölüm 5

### Sonuç, Tartışma ve Öneriler

#### Sonuç

1. Rasch model ve diğer üç açıklayıcı madde tepki modelinden (AMTM) kestirilen madde güçlükleri ortalamaları oldukça yakındır. Alt gruplar bazında ise 0.1 logite kadar farklar gözlenmektedir. En fazla orta düzeyde etki büyüklüğüne sahip bu farklar tüm puanlarda gözlemlenememesine rağmen alt gruplardan bazılarında avantaj sağlarken diğerlerine dezavantaja sebep olabilmektedir.
2. Bu çalışmada ortalama madde güçlükleri arasında fark olmasına karşın RMSE değerlerinin ortalamalar arasında fark olmayan durumlara göre daha küçük kestirilebildiği gözlenmiştir. Belirli koşullar altında en büyük RMSE değerleri, ortalamalar arasında fark olmadığı durumlarda elde edilebilir.
3. Ortalama madde güçlükleri arasında manidar bir fark olmamasının bir sonucu olarak tüm modellerdeki yetenek puanlarının ortalamalarının yakın kestirildiği belirlenmiştir. Alt gruplar bazında ise özellikle en fazla avantaj sağlayan gruplar ile en dezavantajlı grup ortalamalarının RM ortalamalarından manidar bir şekilde farklı olduğu görülmüştür.
4. Madde havuzunun tamamında elde edilen yetenek puanlarının ortalamalarının farklılaşmamasında alt gruplardaki katılımcı sayısının eşit olması etkili olabilir. Örneğin avantaj sağlayan kişi sayısının yüksek olduğu durumlarda genel ortalamalarda da farklılıklar gözlenebilir.
5. Tüm modellerde kestirilen yetenek puanları yüksek pozitif ilişkili kestirilmiştir. Ancak alt gruplarda modellerin yetenek puan ortalamaları arasında manidar fark tespit edilmiştir. Bu durumda yüksek korelasyon her zaman birey bazında çok yakın kestirimler anlamına gelmeyebilir.
6. Madde havuzunda belirli bir oranda birey veya madde değişkenlerinden etkilenen madde bulunması halinde parametre kestirimlerinde bu etkilerin modellenmesi model-veri uyumunu yükseltebilmektedir.
7. Kesinlik ve minInfo sonlandırma kurallarının her ikisi ile de tüm modellerle yapılan BBT simülasyonlarında alt gruplar dikkate alındığında ortalama

yetenek puanlarının Rasch modelden kestirilen yetenek puanlarından farklılaştığı belirlenmiştir. Madde havuzlarından elde edilen yetenek puanlarında da gözlenen bu durum BBT uygulamalarında tekrar etmiştir. Madde havuzlarında gözlenen ortalamalar arasındaki fark, BBT simülasyonları ile daha az madde kullanılarak elde edilebilmektedir.

8. Örtük regresyon doğrusal lojistik test modelinde her iki sonlandırma kuralda da alt gruplara ilişkin yetenek puanları ortalaması önsel yetenek puanlarından farklılaşmaktadır. Bu durumun önsel yetenek parametrelerin kestiriminde model-veri uyumundan kaynaklandığı düşünülmektedir.
9. BBT simülasyonlarında ortalama test uzunluğu ve hangi test uzunluğunun kaç kişiye uygulandığı kestirim yönteminden etkilenmezken sonlandırma kuralına göre farklılaşmaktadır.
10. Madde kullanım sıklığı kestirim modelinden etkilenmezken sonlandırma kuralına göre farklılaştığı gözlenmiştir. Ancak madde ve birey eşdeğişkenlerinin etkili olduğu ilk 20 maddede örtük regresyon ve DLTM'nin madde kullanım sıklığını düşürmede başarılı olduğu belirlenmiştir.
11. MinInfo sonlandırma kuralının testi başarı ile sonlandırılan birey oranındaki gerçek performansı kesme noktasının madde havuzundaki en küçük bilgi düzeyine göre ayarlanması ile ortaya konabilir. Ayrıca bu oranın bu çalışma kapsamında kestirim modelinden etkilenmediği sonucuna ulaşılmıştır.
12. Bu çalışmada minInfo sonlandırma kuralının test örtüşme oranı bakımından kestirim modeli fark etmeksizin kesinlik sonlandırma kuralından daha iyi sonuçlar verdiği gözlenmiştir.
13. MinInfo sonlandırma kuralında kestirim modeli fark etmeksizin daha büyük RMSE ve yanlılık değerleri elde edilmiştir. Ancak, hipotez testi sonuçları ile birlikte düşünüldüğünde eşleştirilmiş yetenek puan farklarının ortalamasının bu sonlandırma kuralında daha küçük olduğu söylenebilir.

## **Tartışma**

Bu araştırmada açıklayıcı madde tepki modellerinin bilgisayar ortamında bireye uyarlanmış testlerde kullanımının belirli parametreler üzerindeki etkisi incelenmiştir. Bu amaç doğrultusunda yürütülen çalışmanın literatürde ilk

olmasından dolayı bulgularının diğer çalışmalarla karşılaştırılması mümkün olmamıştır. Bundan dolayı tartışmada bulguların olası nedenleri daha çok metin içi yoruma dayandırılmıştır.

Alt gruplara ilişkin madde güçlük parametrelerinin ortalamasının RM ortalamalarından farklılaşmasına rağmen gruplar göz ardı edildiğinde her üç modelin ortalama madde güçlüğü'nün 0'a yakın olduğu ve RM'ye benzer kestirimler yaptığı söylenebilir. Bu durum, Atar'ın (2011) TIMSS verileri ile her dört modelden yapılan kestirimleri karşılaştırdığı çalışmanın bulguları ile örtüşmektedir. Manipüle edilen madde sayıları eşit olduğundan her üç AMTM'de alt gruplar için kestirilen güçlük parametreleri simetri göstermektedir. Örneğin ÖR'de en avantajlı grup olan EA için ortalama güçlük -0.1 iken, en dezavantajlı olan KC grubu için ortalama 0.1'dir. Toplam puan bakımından farklılık gözlenmemesinin sebebi alt gruplardaki zıt işaretli madde güçlük ortalamaları olabilir.

Madde güçlüklerinde karşılaşılan durumun benzeri yetenek puanlarında da gözlenmiştir. Alt gruplar göz önünde bulundurulmadığında tüm modellerde kestirilen ortalama yetenek puanlarının farklılaşmadığı belirlenmişken alt gruplar bazındaki ortalama yeteneklerde ise önemli farklılıklara rastlanmıştır. Bu durum araştırma kapsamında beklenen bir durumdur. Alt grup bazında madde parametrelerinin kestirilmesi, DMF gibi durumlardan kaynaklı dezavantajlı olan grupların yetenek puanını yükseltmiş, avantajlı olan grubun RM modeldeki yüksek puanlarını ise aşağıya çekmiş olabilir.

Alt gruplara ait yetenek puanı ortalamaları arasındaki küçük farkların bağımlı örneklem t-Testi sonuçlarında manidar bulunması dikkat çeken bir bulgudur. Rasch model ve diğer modellerden her bir katılımcı için kestirilen yetenek puanı bireysel anlamda farklılaşmasına karşın bu farklılaşma toplamda 0'a çekilmektedir. Örneğin ÖR'de EA alt grubu için kestirilen yetenek puanlarının ortalaması Rasch modelden kestirilen puan ortalamalarından 0.1 logit yüksek iken, KC için kestirilen yetenek puan ortalaması 0.1 logit küçüktür. Alt gruplardaki kişi sayısı eşit olduğundan genel ortalama Rasch model gibi 0'a yaklaşmaktadır. Bu durumda AMTM'nin genel yetenek puanı ortalamalarında herhangi bir etkisi yokmuş gibi görünse de alt gruplara ayrıldığında farklılıklar daha kolay gözlenmektedir. Alt gruplarda eşit sayıda katılımcının bulunmadığı durumlarda genel ortalamaların da farklı çıkma olasılığı vardır.

Yetenek puanlarının çok yüksek ilişkili çıkması her dört modelin birey bazında çok yakın yetenek puanları kestirdiği anlamına gelmeyebilir. Ortalamalar arasındaki farka rağmen yüksek ilişkili veri setleri olması matematiksel olarak mümkündür. Çünkü korelasyon birlikte değişim olarak düşünülmelidir. Örneğin Rasch modelde düşük yeteneğe sahip bir bireyin puanı ÖR'de yine düşük kestirilecektir. Aradaki farka rağmen önemli olan bireylerin eşleştirilmiş puanlarının birlikte değişimidir.

Bu araştırmada ÖR ve DLTM'nin Rasch modelden daha iyi model-veri uyumu gösterdiği belirlenmiştir. Benzer bir durum Atar ve Aktan'ın (2013) 2PL ve örtük regresyon 2PL modellerinin karşılaştırdıkları çalışmada da gözlenmiştir. Bu çalışmada da örtük regresyonun veriye daha iyi uyum sağladığı görülmüştür. Tüm maddelerde olmasa da belirli bir orandaki maddede birey veya madde değişkenlerinin, maddeyi cevaplama etkili olduğu durumlarda bu etkinin modellenmesi model-veri uyumunu arttırmış olabilir. ÖRDLTM'nin model-veri uyumunun daha kötü olmasında bu modelde 36 alt grup yerine üç alt grup üzerinden kestirimlerin yapılması etkili olmuş olabilir.

Gerek toplam puanda gerekse de alt gruplar bazında RM ile AMTM önsel puanları arasındaki fark her iki sonlandırma kuralı ile gerçekleştirilen BBT simülasyonlarında da gözlenmiştir. Çok daha az sayıda madde ile bu farklılıkların BBT simülasyonlarında tekrarlanması post-hoc simülasyonlarının başarılı olduğu şeklinde yorumlanabilir. Model-veri uyumunun sağlandığı durumlarda, doğrusal testlerde DMF, DYF veya MPK gösteren maddelerin olduğu testlerde AMTM kullanılabileceği görülmüştür. Tüm madde havuzunda gözlenen bu durumun BBT simülasyonlarında da başarı ile uygulanabileceği sonucuna ulaşılmıştır.

Çalışmada ayrıca AMTM'nin BBT simülasyonlarında ortalama test uzunluğu, test örtüşme oranı, RMSE ve yanlılık değerleri üzerinde olumlu etkileri olabileceği düşünülmesine rağmen bu durum gözlenmemiştir. Ancak birey ve madde eşdeğişkenlerinin etkili olduğu ilk 20 maddede madde kullanım sıklığını düşürdüğü görülmüştür. AMTM'nin bu istatistikler üzerinde beklenen düzeyde etkili olmamasında madde seçim algoritmalarının etkisinin olduğu söylenebilir. Fisher'in maksimum bilgisinin madde güçlüklerinin yanında madde ayıricılıklarından da oldukça fazla etkilendiği bilinmektedir. Bundan dolayı farklı madde seçim algoritmaları analizlerin tekrarlanması önemlidir.

## Öneriler

### Araştırmacılara yönelik öneriler.

1. Bu araştırmada MTK modellerinden sadece Rasch model ve uzantısı olan açıklayıcı madde tepki modelleri incelenmiştir. Farklı çalışmalar ikili puanlanan veriler için 2PL ve 3PL gibi diğer tek boyutlu modeller için de tekrarlanabilir.
2. Çalışma kapsamında yine normal dağılım gösteren tek boyutlu veriler için model karşılaştırması yapılmıştır. Çok boyutlu MTK modelleri ve açıklayıcı madde tepki modelleri incelenebilir.
3. Araştırma kapsamında sadece iki farklı sonlandırma kuralına ilişkin sonuçlara odaklanılmıştır. Benzer bir çalışma farklı sonlandırma kuralları ile tekrar yapılabilir.
4. Araştırmada madde kullanım sıklığı kontrol yöntemi olarak sadece randomesque yöntemi kullanılmıştır. Yine birden fazla yöntemin ortaya koyabileceği farklılıklar bir çalışmada incelenebilir.
5. Araştırmada hem madde kalibrasyonu aşamasında hem post-hoc simülasyonlarında yetenek kestirim yöntemi olarak BSD kullanılmıştır. Bayes istatistiğine dayanmayan yetenek kestirim yöntemleri ile araştırma tekrarlanabilir.
6. Post-hoc simülasyonlarında madde seçim algoritması olarak sadece maksimum Fisher bilgisi kullanılmıştır. Kullback-Leibler bilgisi gibi çok sayıda madde seçim algoritması ile sonuçlar karşılaştırılabilir.
7. Araştırmada model bazında alt gruplarda eşit katılımcı olacak şekilde tasarım yapılmıştır. Alt gruplardaki katılımcı sayısının farklı olması durumunda madde güçlükleri ve yetenek puanının nasıl değişeceği incelenebilir.
8. BBT uygulamalarında simülasyon esasında daha önce sorulan madde sayısı üzerinden maddenin kullanım sırası modellenme durumu ile modellenmeme durumu karşılaştırılabilir.

9. Aynı madde havuzunun engellilik durumuna göre hem görsel hem işitsel olarak BBT uygulamalarında kullanılması durumunda test formunun kestirimler üzerindeki etkisi incelenebilir.

#### **Uygulamaya yönelik öneriler.**

1. Araştırma kapsamında madde havuzu geliştirme sürecinde DMF, DYF, madde parametre kaymasını kontrol etme ve benzeri durumlarda örtük regresyon ve doğrusal lojistik test modeli kullanılabilir.
2. Çalışmada toplam yetenek puanlarının ortalamaları arasında fark olmadığı sonucuna ulaşılırken alt gruplar bazında ortalamalarda farklılık tespit edilmiştir. Bu bakımdan genel ortalamalar arasındaki farkın yanında mutlaka grup ortalamaları arasındaki fark da incelenmelidir.
3. Modeller bazında ortalamalar arasındaki farkın yanında RMSE değerleri ile eşleştirilmiş bireysel yetenek puanlarının farkları da incelenebilir.
4. MinInfo sonlandırma kuralı ile daha az madde kullanılarak daha düşük madde kullanım sıklığı elde edilebilir. Ayrıca birey ve madde değişkenlerinin etkili olduğu maddeler için örtük regresyon veya doğrusal lojistik test modeli kullanılarak bu maddelere ilişkin madde kullanım sıklığı düşürülebilir.
5. Hem doğrusal testlerde hem BBT simülasyonlarında, belirli bir gruba avantaj sağlayan maddeler bulunduğu anda örtük regresyon analizi ile kestirilmiş parametreler kullanılarak yanlışlık giderilebilir.
6. Maddelerde belirli bir alt gruba avantaj sağlayan madde özellikleri (örneğin maddenin uygulanma sırası) belirlenmesi halinde DLTM ile bu etki sabitlenebilir.
7. Birey ve madde değişkenlerinin yanında bu iki değişken türünün ortak etkisi göz ardı edilmemelidir. Tüm katılımcılarda ortaya çıkmayan bir madde değişkeni etkisi belirli bir grupta gözlenebilir. Bu durumda bu etki ÖRDLM ile kontrol edilebilir.

## Kaynaklar

- Ali, U. S., & Chang, H. (2014), An item-driven adaptive design for calibrating pretest items. ETS Research Report Series, 2014: 1-12. doi:10.1002/ets2.12044
- Ansley, T. N., & Forsyth, R. A. (2016). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement*, 9(1), 37–48. <https://doi.org/10.1177/014662168500900104>
- Atar, B. (2011). Tanımlayıcı ve açıklayıcı madde tepki modellerinin TIMSS 2007 Türkiye matematik verisine uyarlanması. *Eğitim ve Bilim*, 36(159). Erişim adresi: <http://egitimvebilim.ted.org.tr/index.php/EB/article/download/811/252>
- Atar, B., & Aktan, D. Ç. (2013). Birey açıklayıcı madde tepki kuramı analizi: Örtük regresyon iki parametrelili lojistik modeli. *Eğitim ve Bilim*, 38(168). Erişim adresi: <http://egitimvebilim.ted.org.tr/index.php/EB/article/download/942/578>
- Ayala, R. J. de. (2009). *The theory and practice of item response theory. Methodology in the social sciences*. New York: Guilford Press.
- Babcock, B., & Weiss, D. J. (2013). Termination criteria in computerized adaptive tests: Do variable-length CATs provide efficient and effective measurement? *Journal of Computerized Adaptive Testing*, 1(1), 1–18. <https://doi.org/10.7333/1212-0101001>
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3), 379–384. <https://doi.org/10.3758/BF03192707>
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park Md.: ERIC Clearinghouse on Assessment and Evaluation.
- Baker, F. B., & Kim, S. H. (2017). *The basics of item response theory using R. Statistics for social and behavioral sciences*, 2199-7357. Cham, Switzerland: Springer.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>

- Bejar, I. I. (1983). *Achievement testing: Recent advances. Quantitative applications in the social sciences, 0149-192x: no.07-036*. Beverly Hills, London: Sage.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443–459.
- Boeck, P. de., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software, 39*(12). <https://doi.org/10.18637/jss.v039.i12>
- Boeck, P. de, Cho, S. J., & Wilson, M. (2011). Explanatory secondary dimension modeling of latent differential item functioning. *Applied Psychological Measurement, 35*(8), 583–603. <https://doi.org/10.1177/0146621611428446>
- Boeck, P. de., & Wilson, M. (2004). *Explanatory item response models*. New York, NY: Springer New York.
- Braeken, J. & Tuerlinckx, F. (2009). Investigating latent constructs with item response models: A MATLAB IRTm toolbox. *Behavior Research Methods, 41*(4), 1127–1137. <https://doi.org/10.3758/BRM.41.4.1127>
- Briggs, D. C. (2008). Using explanatory item response models to analyze group differences in science achievement. *Applied Measurement in Education, 21*(2), 89–118. <https://doi.org/10.1080/08957340801926086>
- Brown, A., & Iwashita, N. (1996). Language background and item difficulty: The development of a computer-adaptive test of Japanese. *System, 24*(2), 199–206. [https://doi.org/10.1016/0346-251X\(96\)00004-8](https://doi.org/10.1016/0346-251X(96)00004-8)
- Bulut, O., Palma, J., Rodriguez, M. C., & Stanke, L. (2015). Evaluating measurement invariance in the measurement of developmental assets in Latino English language groups across developmental stages. *SAGE Open, 5*(2). <https://doi.org/10.1177/2158244015586238>
- Chen, F., Yang, H., Bulut, O., Cui, Y., & Xin, T. (2019). Examining the relation of personality factors to substance use disorder by explanatory item response modeling of DSM-5 symptoms. *PloS One, 14*(6), e0217630. <https://doi.org/10.1371/journal.pone.0217630>

- Chen, S. K., Hou, L., & Dodd, B. G. (2016). A Comparison of maximum likelihood estimation and expected a posteriori estimation in CAT using the partial credit model. *Educational and Psychological Measurement*, 58(4), 569–595. <https://doi.org/10.1177/0013164498058004002>
- Chen, S. K., Hou, L., Fitzpatrick, S. J., & Dodd, B. G. (2016). The Effect of population distribution and method of theta estimation on computerized adaptive testing (CAT) Using the rating scale model. *Educational and Psychological Measurement*, 57(3), 422–439. <https://doi.org/10.1177/0013164497057003004>
- 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. <https://doi.org/10.2307/1165285>
- Cho, S. J., Boeck, P. de., & Lee, W. Y. (2017). Evaluating testing, profile likelihood confidence interval estimation, and model comparisons for item covariate effects in linear logistic test models. *Applied Psychological Measurement*, 41(5), 353–371. <https://doi.org/10.1177/0146621617692078>
- Choe, E. M., Kern, J. L., & Chang, H. H. (2017). Optimizing the use of response times for item selection in computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, 43(2), 135–158. <https://doi.org/10.3102/1076998617723642>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.
- Debelak, R., & Koller, I. (2019). Testing the local independence assumption of the rasch model with Q3-based nonparametric model tests. *Applied Psychological Measurement*, 6, 014662161983550. <https://doi.org/10.1177/0146621619835501>
- DeMars, C. (2010). *Item response theory. Understanding statistics*. Oxford, New York: Oxford University Press.
- DiTrapani, J., Rockwood, N., & Jeon, M. (2018). Explanatory IRT analysis using the SPIRIT macro in SPSS. *The Quantitative Methods for Psychology*, 14(2), 81–98. <https://doi.org/10.20982/tqmp.14.2.p081>

- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists. Multivariate applications book series*. Mahwah N.J.: L. Erlbaum Associates.
- Erođlu, M. G., & Keleciođlu, H. (2015). Bireyselleřtirilmiř bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının ölçme kesinliđi ve test uzunluđu açısından karřılařtırılması. *Uludađ Üniversitesi Eđitim Fakóltesi Dergisi*, 28(1). <https://doi.org/10.19171/uuefd.87973>
- Fagginger, M. F., Hickendorff, M., & van Putten, C. M. (2018). Training can increase students' choices for written solution strategies and performance in solving multi-digit division problems. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01644>
- Field, A. P. (2009). *Discovering statistics using SPSS (3rd ed.)*. Los Angeles: SAGE Publications.
- Finch, W. H., & French, B. F. (2015). *Latent variable modeling with R*. New York: Routledge Taylor & Francis Group.
- Fox, J. P. (2010). *Bayesian item response modeling: Theory and applications. Statistics for social and behavioral sciences*. New York, NY: Springer.
- Glas, C. A. W. (1988). The derivation of some tests for the rasch model from the multinomial distribution. *Psychometrika*, 53(4), 525–546.
- Glas, C. A. W., & van der Linden, W. J. (2003). Computerized adaptive testing with item cloning. *Applied Psychological Measurement*, 27(4), 247–261. <https://doi.org/10.1177/0146621603027004001>
- Goldhammer, F., Martens, T., & Lüdtke, O. (2017). Conditioning factors of test-taking engagement in PIAAC: An exploratory IRT modelling approach considering person and item characteristics. *Large-scale Assessments in Education*, 5(1), 92. <https://doi.org/10.1186/s40536-017-0051-9>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory. Measurement methods for the social sciences series*. Newbury Park, Calif.: Sage Publications.
- Han, K. C. T. (2018). Components of the item selection algorithm in computerized adaptive testing. *Journal of Educational Evaluation for Health Professions*, 15(7). <https://doi.org/10.3352/jeehp.2018.15.7>

- Hartig, J., Frey, A., Nold, G., & Klieme, E. (2012). An application of explanatory item response modeling for model-based proficiency scaling. *Educational and Psychological Measurement*, 72(4), 665–686. <https://doi.org/10.1177/0013164411430707>
- Hecht, M., Weirich, S., Siegle, T., & Frey, A. (2015a). Effects of Design properties on parameter estimation in large-scale assessments. *Educational and Psychological Measurement*, 75(6), 1021–1044. <https://doi.org/10.1177/0013164415573311>
- Hecht, M., Weirich, S., Siegle, T., & Frey, A. (2015b). Modeling booklet effects for nonequivalent group designs in large-scale assessment. *Educational and Psychological Measurement*, 75(4), 568–584. <https://doi.org/10.1177/0013164414554219>
- Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*, 38(8). <https://doi.org/10.18637/jss.v038.i08>
- Jeon, M., & Rijmen, F. (2016). A modular approach for item response theory modeling with the R package flirt. *Behavior Research Methods*, 48(2), 742–755. <https://doi.org/10.3758/s13428-015-0606-z>
- Kahraman, N. (2014). An explanatory item response theory approach for a computer-based case simulation test. *Eurasian Journal of Educational Research*, 14(54), 117–134. <https://doi.org/10.14689/ejer.2014.54.7>
- Kim, S. H., Cohen, A. S., & Park, T. H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement*, 32(3), 261–276. <https://doi.org/10.1111/j.1745-3984.1995.tb00466.x>
- Leroux, A. J., Lopez, M., Hembry, I., & Dodd, B. G. (2013). A comparison of exposure control procedures in CATs using the 3pl model. *Educational and Psychological Measurement*, 73(5), 857–874. <https://doi.org/10.1177/0013164413486802>
- Leung, C. K., Chang, H. H., & Hau, K. T. (2003). Computerized adaptive testing: a comparison of three content balancing methods. *The Journal of Technology, Learning and Assessment*, 2(5). Retrieved from : <https://ejournals.bc.edu/index.php/jtla/article/download/1665/1507>

- Levy, R., Mislevy, R. J., & Sinharay, S. (2009). Posterior predictive model checking for multidimensionality in item response theory. *Applied Psychological Measurement, 33*(7), 519–537. <https://doi.org/10.1177/0146621608329504>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, N.J.: L. Erlbaum Associates.
- Magis, D. (2013). A Note on the Item Information Function of the Four-Parameter Logistic Model. *Applied Psychological Measurement, 37*(4), 304–315. <https://doi.org/10.1177/0146621613475471>
- Magis, D., & Barrada, J. R. (2017). Computerized Adaptive Testing with R: Recent updates of the package catR. *Journal of Statistical Software, 76*(Code Snippet 1). <https://doi.org/10.18637/jss.v076.c01>
- Magis, D., Béland, S., Tuerlinckx, F., & Boeck, P. de. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*(3), 847–862. <https://doi.org/10.3758/BRM.42.3.847>
- Magis, D., Yan, D. A., & Davier, A. A. V. (2017). *Computerized adaptive and multistage testing with R: Using packages catR and mstR* / David Magis, Duanli Yan, Alina A. von Davier. Use R! Cham: Springer.
- McDonald, R. P. (1982). Linear versus models in item response theory. *Applied Psychological Measurement, 6*(4), 379–396. <https://doi.org/10.1177/014662168200600402>
- McLeod, L. D., & Schnipke, D. L. (2006). *Detecting items that have been memorized. Law school admission council computerized testing report: 99-05*. Newtown, PA: Law School Admission Council.
- Min, H., Zickar, M., & Yankov, G. (2018). Understanding item parameters in personality scales: An explanatory item response modeling approach. *Personality and Individual Differences, 128*, 1–6. <https://doi.org/10.1016/j.paid.2018.02.012>
- Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models. Quantitative applications in the social sciences: 07-144*. Thousand Oaks, Calif., London: Sage.

- Paek, I., & Cole, K. (2019). *Using R for item response theory model applications* (1st). London: Routledge.
- Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (6th edition). Maidenhead, Berkshire, England: McGraw-Hill Education.
- Park, J. Y., Joo, S. H., Cornillie, F., van der Maas, H. L. J., & van den Noortgate, W. (2018). An explanatory item response theory method for alleviating the cold-start problem in adaptive learning environments. *Behavior Research Methods*. Advance online publication. <https://doi.org/10.3758/s13428-018-1166-9>
- Park, Y. S., Lee, Y.-S., & Xing, K. (2016). Investigating the Impact of Item Parameter Drift for Item Response Theory Models with Mixture Distributions. *Frontiers in Psychology, Quantitative Psychology and Measurement*, 7(255), DOI 10.3389/fpsyg.2016.00255
- Randall, J., Cheong, Y. F., & Engelhard, G. (2010). Using explanatory item response theory modeling to investigate context effects of differential item functioning for students with disabilities. *Educational and Psychological Measurement*, 71(1), 129–147. <https://doi.org/10.1177/0013164410391577>
- Ravand, H. (2015). Item response theory using hierarchical generalized linear models. *Practical Assessment, Research & Evaluation*, 20(7).
- Reckase, M. (2009). *Multidimensional item response theory. Statistics for social and behavioral sciences*. New York, London: Springer.
- Reckase, M. D. (1979). Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implications. *Journal of Educational Statistics*, 4(3), 207. <https://doi.org/10.2307/1164671>
- Revelle, W. (2018). *psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois. Retrieved from : <https://CRAN.R-project.org/package=psych>
- Revuelta, J., Ximénez, M. C., & Olea, J. (2016). Psychometric and psychological effects of item selection and review on computerized testing. *Educational and*

*Psychological Measurement*, 63(5), 791–808.  
<https://doi.org/10.1177/0013164403251282>

- Robitzsch, A. (2019). *sirt: Supplementary item response theory models*. Retrieved from : <https://CRAN.R-project.org/package=sirt>
- Rose, N., Nagy, G., Nagengast, B., Frey, A., & Becker, M. (2019). Modeling multiple item context effects with generalized linear mixed models. *Frontiers in Psychology*, 10(163). <https://doi.org/10.3389/fpsyg.2019.00248>
- Santi, K. L., Kulesz, P. A., Khalaf, S., & Francis, D. J. (2015). Developmental changes in reading do not alter the development of visual processing skills: An application of explanatory item response models in grades K-2. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00116>
- Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61(2), 331–354. <https://doi.org/10.1007/BF02294343>
- Segall, D. O. (2005). Computerized adaptive testing. *Encyclopedia of Social Measurement*, 1, 429–438.
- Sheu, C. F., Chen, C. T., Su, Y. H., & Wang, W.-C. (2005). Using SAS PROC NLMIXED to fit item response theory models. *Behavior Research Methods*, 37(2), 202–218. <https://doi.org/10.3758/BF03192688>
- Spray, J. A., Davey, T., Reckase, M. D., Ackerman, T. A., & Carlson, J. E. (1990). *Comparison of two logistic multidimensional item response theory models (research report onr90-8)*. Iowa City: American College Testing Program.
- Stanke, L., & Bulut, O. (2019). Explanatory item response models for polytomous item responses. *International Journal of Assessment Tools in Education*, 259–278. <https://doi.org/10.21449/ijate.515085>
- Şenel, S., & Kutlu, Ö. (2018). Görme engelli öğrencilere yönelik bilgisayar ortamında bireye uyarlanmış test tasarımı. *Eğitim ve Bilim*, 43(194). Erişim adresi: <http://egitimvebilim.ted.org.tr/index.php/EB/article/download/7515/2760>
- Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, 47(2), 175–186.

- Thompson, N. A. (2009). Item selection in computerized classification testing. *Educational and Psychological Measurement*, 69(5), 778–793. <https://doi.org/10.1177/0013164408324460>
- van den Wollenberg, A. L. (1982). Two new test statistics for the Rasch model. *Psychometrika*, 47(2), 123–140. <https://doi.org/10.1007/BF02296270>
- van der Linden, W. J. (2005). *A comparison of item-selection methods for adaptive tests with content constraints. Law school admission council computerized testing report: 04-02*. Newtown, PA: Law School Admission Council.
- van der Linden, W. J. (2006). *A formal characterization of and some alternatives to Symptom-Hetter item-exposure control in computerized adaptive testing. Law school admission council computerized testing report: 02-05*. Newtown, PA: Law School Admission Council.
- van der Linden, W. J. (2016). *Handbook of item response theory. Volume two, Statistical tools (1 bs)*. Chapman & Hall/CRC statistics in the social and behavioral sciences series. Boca Raton: Chapman & Hall/CRC.
- van der Linden, W. J., & Glas, C. A. W. (Ed.). (2000). *Computerized adaptive testing: Theory and practice*. Dordrecht, Boston: Kluwer Academic.
- van der Linden, W. J., & Glas, C. A. W. (2010). *Elements of adaptive testing*. New York, NY: Springer New York.
- Wainer, H., & Dorans, N. J. (2000). *Computerized adaptive testing: A primer (2. bs)*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Weirich, S., Hecht, M., & Böhme, K. (2014). Modeling item position effects using generalized linear mixed models. *Applied Psychological Measurement*, 38(7), 535–548. <https://doi.org/10.1177/0146621614534955>
- Weiss, D. J. (Ed.) 1983. *New horizons in testing: Latent trait test theory and computerized adaptive testing*. New York: Academic Press.
- Wilson, M., Zheng, X., & McGuire, L. (2012). Formulating latent growth using an explanatory item response model approach. *Journal of Applied Measurement*, 13(1), 1–22.

- Yao, L. (2012). 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>
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*(2), 125–145. <https://doi.org/10.1177/014662168400800201>
- Zhang, H., Lu, N., Feng, C., Thurston, S. W., Xia, Y., Zhu, L., & Tu, X. M. (2011). On fitting generalized linear mixed-effects models for binary responses using different statistical packages. *Statistics in Medicine, 30*(20), 2562–2572. <https://doi.org/10.1002/sim.4265>
- Zwick, R. (2000). The assessment of differential item functioning in computer adaptive tests. W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice*. Dordrecht, Boston: Kluwer Academic.

## Ek A. Madde Havuzlarına İlişkin Faktör Yükleri

Madde	1. Madde Havuzu	2. Madde Havuzu	3. Madde Havuzu	4. Madde Havuzu	5. Madde Havuzu	6. Madde Havuzu	7. Madde Havuzu	8. Madde Havuzu	9. Madde Havuzu	10. Madde Havuzu
M1	0,362	0,477	0,509	0,449	0,504	0,476	0,486	0,449	0,461	0,500
M2	0,464	0,529	0,523	0,512	0,533	0,414	0,454	0,516	0,561	0,461
M3	0,439	0,477	0,421	0,475	0,545	0,457	0,524	0,439	0,510	0,459
M4	0,553	0,514	0,355	0,458	0,479	0,521	0,517	0,494	0,483	0,451
M5	0,475	0,495	0,407	0,454	0,434	0,425	0,440	0,439	0,467	0,367
M6	0,485	0,493	0,362	0,491	0,514	0,448	0,515	0,450	0,519	0,510
M7	0,487	0,491	0,453	0,492	0,447	0,458	0,551	0,435	0,475	0,468
M8	0,480	0,430	0,430	0,534	0,482	0,476	0,485	0,379	0,498	0,463
M9	0,503	0,412	0,417	0,565	0,542	0,488	0,472	0,504	0,504	0,462
M10	0,427	0,502	0,485	0,500	0,516	0,551	0,478	0,548	0,538	0,470
M11	0,456	0,488	0,551	0,443	0,456	0,478	0,518	0,479	0,429	0,516
M12	0,476	0,484	0,496	0,483	0,447	0,389	0,474	0,524	0,484	0,481
M13	0,433	0,432	0,398	0,459	0,405	0,538	0,481	0,473	0,490	0,513
M14	0,541	0,404	0,462	0,531	0,443	0,440	0,360	0,386	0,512	0,545
M15	0,513	0,417	0,430	0,514	0,523	0,368	0,471	0,413	0,582	0,465
M16	0,483	0,466	0,487	0,472	0,426	0,452	0,417	0,456	0,517	0,445
M17	0,460	0,521	0,417	0,560	0,495	0,425	0,444	0,518	0,514	0,490
M18	0,481	0,555	0,421	0,529	0,526	0,369	0,510	0,461	0,454	0,488
M19	0,519	0,550	0,441	0,520	0,484	0,495	0,399	0,493	0,502	0,523
M20	0,438	0,473	0,476	0,488	0,504	0,451	0,544	0,475	0,491	0,407
M21	0,508	0,501	0,564	0,513	0,543	0,515	0,546	0,532	0,501	0,582
M22	0,603	0,503	0,508	0,540	0,508	0,536	0,504	0,471	0,524	0,542
M23	0,504	0,488	0,535	0,517	0,508	0,488	0,529	0,542	0,518	0,545
M24	0,566	0,607	0,514	0,516	0,510	0,473	0,515	0,536	0,492	0,498
M25	0,460	0,572	0,566	0,472	0,525	0,526	0,524	0,535	0,510	0,468
M26	0,516	0,532	0,545	0,539	0,517	0,566	0,505	0,599	0,514	0,496
M27	0,541	0,546	0,521	0,515	0,514	0,518	0,528	0,513	0,533	0,515
M28	0,555	0,494	0,500	0,565	0,499	0,511	0,460	0,518	0,529	0,575
M29	0,593	0,509	0,494	0,534	0,524	0,542	0,533	0,528	0,547	0,510
M30	0,573	0,571	0,506	0,508	0,517	0,518	0,512	0,535	0,521	0,500
M31	0,526	0,560	0,461	0,515	0,478	0,522	0,563	0,509	0,521	0,534
M32	0,510	0,540	0,524	0,528	0,477	0,501	0,556	0,512	0,549	0,520
M33	0,511	0,512	0,492	0,545	0,526	0,497	0,510	0,553	0,552	0,518
M34	0,525	0,566	0,501	0,544	0,511	0,498	0,484	0,494	0,548	0,553
M35	0,526	0,548	0,525	0,565	0,523	0,514	0,522	0,519	0,523	0,519
M36	0,514	0,491	0,504	0,571	0,478	0,510	0,495	0,529	0,581	0,538
M37	0,495	0,567	0,473	0,515	0,516	0,579	0,552	0,540	0,542	0,556
M38	0,555	0,519	0,515	0,478	0,498	0,498	0,509	0,485	0,514	0,494
M39	0,521	0,585	0,454	0,587	0,458	0,495	0,534	0,495	0,494	0,513
M40	0,512	0,527	0,512	0,522	0,516	0,551	0,558	0,569	0,566	0,529
M41	0,490	0,525	0,485	0,553	0,531	0,518	0,458	0,479	0,503	0,533
M42	0,569	0,490	0,462	0,570	0,506	0,510	0,536	0,517	0,593	0,535
M43	0,538	0,511	0,567	0,501	0,539	0,509	0,506	0,539	0,530	0,504
M44	0,523	0,510	0,503	0,529	0,530	0,536	0,518	0,534	0,506	0,516
M45	0,529	0,538	0,486	0,538	0,484	0,555	0,506	0,526	0,528	0,535
M46	0,584	0,557	0,518	0,510	0,538	0,485	0,491	0,495	0,561	0,499
M47	0,502	0,493	0,492	0,511	0,545	0,521	0,536	0,527	0,486	0,511
M48	0,545	0,554	0,552	0,546	0,449	0,522	0,538	0,522	0,564	0,527
M49	0,548	0,478	0,525	0,543	0,581	0,525	0,535	0,482	0,517	0,528
M50	0,487	0,502	0,535	0,540	0,550	0,524	0,498	0,535	0,517	0,558
M51	0,524	0,547	0,505	0,525	0,501	0,508	0,555	0,535	0,511	0,540
M52	0,592	0,483	0,532	0,477	0,530	0,486	0,495	0,511	0,544	0,570
M53	0,519	0,529	0,518	0,550	0,514	0,495	0,492	0,523	0,521	0,554
M54	0,520	0,563	0,509	0,507	0,500	0,533	0,501	0,481	0,531	0,534
M55	0,528	0,462	0,471	0,552	0,487	0,526	0,457	0,504	0,447	0,526
M56	0,521	0,535	0,549	0,512	0,548	0,498	0,519	0,547	0,542	0,510

M57	0,539	0,539	0,501	0,504	0,546	0,578	0,506	0,500	0,526	0,462
M58	0,505	0,485	0,473	0,540	0,535	0,518	0,521	0,486	0,541	0,549
M59	0,512	0,477	0,462	0,525	0,551	0,500	0,502	0,564	0,559	0,486
M60	0,560	0,488	0,518	0,520	0,513	0,540	0,507	0,495	0,542	0,474
M61	0,530	0,463	0,514	0,558	0,501	0,500	0,481	0,463	0,539	0,589
M62	0,498	0,533	0,516	0,557	0,550	0,457	0,510	0,556	0,541	0,519
M63	0,538	0,483	0,483	0,526	0,554	0,510	0,525	0,478	0,555	0,536
M64	0,504	0,552	0,577	0,551	0,514	0,545	0,475	0,503	0,526	0,522
M65	0,536	0,500	0,460	0,553	0,526	0,523	0,542	0,537	0,546	0,555
M66	0,547	0,540	0,481	0,488	0,466	0,535	0,498	0,545	0,494	0,527
M67	0,533	0,506	0,457	0,526	0,540	0,536	0,524	0,499	0,533	0,503
M68	0,492	0,557	0,525	0,565	0,483	0,537	0,515	0,548	0,477	0,548
M69	0,493	0,521	0,547	0,567	0,475	0,551	0,555	0,500	0,496	0,555
M70	0,544	0,522	0,552	0,544	0,523	0,478	0,500	0,511	0,519	0,523
M71	0,580	0,547	0,461	0,520	0,545	0,489	0,492	0,543	0,560	0,518
M72	0,470	0,521	0,526	0,546	0,483	0,506	0,498	0,513	0,572	0,549
M73	0,558	0,537	0,441	0,547	0,539	0,550	0,512	0,535	0,566	0,520
M74	0,473	0,551	0,547	0,517	0,492	0,548	0,554	0,545	0,499	0,501
M75	0,560	0,538	0,517	0,483	0,532	0,497	0,520	0,522	0,538	0,521
M76	0,470	0,517	0,465	0,500	0,498	0,569	0,523	0,496	0,540	0,551
M77	0,520	0,467	0,508	0,503	0,490	0,556	0,503	0,522	0,570	0,480
M78	0,541	0,584	0,488	0,460	0,493	0,538	0,522	0,510	0,540	0,536
M79	0,461	0,542	0,502	0,505	0,517	0,493	0,520	0,509	0,543	0,520
M80	0,561	0,555	0,567	0,528	0,523	0,478	0,493	0,558	0,505	0,538
M81	0,511	0,469	0,466	0,531	0,493	0,493	0,507	0,549	0,525	0,562
M82	0,510	0,538	0,516	0,560	0,448	0,535	0,510	0,533	0,544	0,527
M83	0,531	0,502	0,471	0,552	0,497	0,479	0,527	0,493	0,528	0,546
M84	0,550	0,528	0,488	0,524	0,532	0,543	0,534	0,541	0,555	0,477
M85	0,587	0,570	0,552	0,540	0,493	0,522	0,498	0,515	0,534	0,518
M86	0,530	0,611	0,533	0,540	0,462	0,565	0,517	0,534	0,552	0,547
M87	0,524	0,541	0,471	0,496	0,556	0,515	0,458	0,535	0,487	0,526
M88	0,562	0,517	0,544	0,556	0,517	0,521	0,491	0,508	0,467	0,550
M89	0,541	0,534	0,512	0,511	0,446	0,534	0,488	0,506	0,532	0,509
M90	0,553	0,556	0,536	0,577	0,512	0,500	0,441	0,551	0,495	0,537
M91	0,543	0,580	0,499	0,550	0,480	0,509	0,516	0,487	0,510	0,525
M92	0,565	0,496	0,537	0,565	0,525	0,521	0,548	0,485	0,520	0,535
M93	0,525	0,520	0,460	0,524	0,519	0,507	0,522	0,535	0,558	0,543
M94	0,574	0,523	0,518	0,516	0,516	0,501	0,520	0,500	0,483	0,539
M95	0,489	0,550	0,517	0,547	0,486	0,515	0,499	0,526	0,505	0,548
M96	0,521	0,583	0,531	0,551	0,503	0,449	0,531	0,507	0,537	0,499
M97	0,569	0,526	0,544	0,539	0,536	0,520	0,501	0,527	0,517	0,492
M98	0,538	0,564	0,575	0,538	0,468	0,523	0,515	0,510	0,484	0,485
M99	0,537	0,517	0,490	0,577	0,542	0,494	0,457	0,539	0,543	0,593
M100	0,584	0,530	0,495	0,515	0,498	0,508	0,499	0,516	0,495	0,580

## EK B. Birinci Birey Değişkenine Göre DMF Analizi Sonuçları

Madde	Lord'un $\chi^2$ İstatistiği									
	1. Madde Havuzu	2. Madde Havuzu	3. Madde Havuzu	4. Madde Havuzu	5. Madde Havuzu	6. Madde Havuzu	7. Madde Havuzu	8. Madde Havuzu	9. Madde Havuzu	10. Madde Havuzu
<b>M1</b>	<b>438.74*</b>	<b>204.05*</b>	<b>110.69*</b>	<b>311.03*</b>	<b>55.31*</b>	<b>132.67*</b>	<b>220.90*</b>	<b>229.55*</b>	<b>360.97*</b>	<b>98.02*</b>
<b>M2</b>	<b>137.50*</b>	<b>44.95*</b>	<b>39.61*</b>	<b>169.08*</b>	<b>15.05*</b>	<b>239.07*</b>	<b>223.53*</b>	<b>47.61*</b>	<b>97.45*</b>	<b>153.80*</b>
<b>M3</b>	<b>194.96*</b>	<b>197.22*</b>	<b>307.67*</b>	<b>131.47*</b>	<b>18.98*</b>	<b>261.93*</b>	<b>211.66*</b>	<b>289.12*</b>	<b>254.33*</b>	<b>117.69*</b>
<b>M4</b>	<b>76.77*</b>	<b>149.48*</b>	<b>566.63*</b>	<b>261.26*</b>	<b>165.76*</b>	<b>141.47*</b>	<b>135.16*</b>	<b>53.38*</b>	<b>259.81*</b>	<b>204.45*</b>
<b>M5</b>	<b>177.04*</b>	<b>111.95*</b>	<b>358.81*</b>	<b>249.67*</b>	<b>227.18*</b>	<b>388.72*</b>	<b>322.58*</b>	<b>126.97*</b>	<b>185.81*</b>	<b>334.17*</b>
M6	1.70	0.26	0.01	1.44	0.00	0.29	0.07	0.01	1.54	2.01
M7	0.49	0.01	0.00	0.02	0.16	0.26	1.12	1.61	1.76	0.12
M8	0.28	0.51	1.00	0.00	0.01	2.57	0.08	0.15	0.01	1.84
M9	0.07	4.54	0.56	1.56	0.05	0.02	1.31	0.37	0.00	0.90
M10	0.00	0.44	0.39	0.63	0.58	0.15	0.00	0.01	0.33	0.28
M11	0.33	0.51	1.55	0.41	0.37	2.13	0.84	0.00	1.54	0.19
M12	2.26	0.01	0.15	0.45	0.18	1.42	0.47	1.15	0.34	0.10
M13	0.16	0.74	0.25	0.02	0.04	0.37	0.13	0.55	0.00	0.02
M14	0.00	3.87	2.59	2.03	0.08	0.17	0.21	0.04	4.73	0.09
M15	1.18	0.31	0.29	1.47	0.04	0.32	3.43	0.12	2.51	0.64
M16	0.44	0.08	0.00	0.55	3.93	0.06	0.07	0.25	0.34	0.80
M17	0.21	0.50	0.67	0.27	0.06	0.09	0.15	0.33	3.72	0.17
M18	0.86	0.18	6.49	0.27	0.29	0.27	0.01	0.52	1.44	0.08
M19	0.07	1.28	0.67	1.32	0.07	2.06	1.24	0.03	0.44	0.79
M20	0.71	4.14	2.19	0.19	0.81	0.35	0.45	0.42	0.64	1.23
M21	0.06	0.25	0.49	4.07	1.17	4.00	0.45	1.15	1.25	0.01
M22	0.04	0.62	1.13	1.97	1.99	0.18	0.01	0.36	0.60	0.40
M23	0.22	0.19	1.50	0.88	0.09	0.20	1.17	0.18	0.84	0.32
M24	0.01	0.28	0.08	1.09	1.35	0.05	0.36	0.50	0.32	0.41
M25	0.07	0.02	0.32	0.71	0.06	0.65	0.13	0.05	0.02	5.08
M26	0.01	0.01	0.55	0.01	0.00	0.68	3.33	0.18	0.01	0.00
M27	0.47	0.21	2.51	1.62	0.02	3.44	0.01	1.46	0.17	3.68
M28	0.26	0.26	0.00	1.39	3.14	2.16	0.75	1.32	0.22	0.07
M29	1.90	2.05	0.45	0.00	1.20	0.23	2.00	1.09	1.42	0.21
M30	0.01	0.19	0.71	0.65	1.09	0.03	1.78	1.04	2.16	0.09
M31	0.59	0.02	1.22	0.19	1.04	2.09	0.99	0.49	3.14	0.02
M32	1.20	0.01	2.38	1.34	7.80	0.84	0.80	0.10	1.50	1.33
M33	2.31	0.08	1.31	0.01	0.00	1.15	0.50	2.45	0.07	0.01
M34	0.05	0.49	0.93	0.00	1.40	0.07	0.00	1.31	0.11	0.17
M35	0.03	1.55	0.28	0.66	0.04	1.33	0.09	1.10	1.32	0.56
M36	0.01	0.78	0.24	0.09	0.11	0.01	2.76	0.51	0.07	0.34
M37	1.64	1.27	0.17	0.76	2.12	0.98	1.10	0.33	5.46	1.03
M38	5.20	0.00	0.57	0.02	0.01	2.47	0.40	0.30	0.49	0.17
M39	5.26	0.00	1.59	0.62	0.96	0.81	0.12	0.52	0.17	0.00
M40	1.46	2.47	0.30	0.00	0.01	0.00	0.81	0.81	0.01	0.29
M41	0.63	0.02	0.00	0.02	0.29	0.10	0.05	1.22	2.35	1.61
M42	0.08	0.10	1.40	0.68	0.13	0.49	0.00	3.22	1.29	2.24
M43	0.90	0.26	0.00	0.56	0.66	0.16	2.15	3.86	0.62	2.50
M44	0.07	0.06	5.64	0.43	0.02	0.28	3.50	0.44	0.31	3.14
M45	0.37	1.74	1.20	0.04	0.10	0.06	2.86	0.01	0.26	1.38
M46	0.02	0.20	0.01	0.27	0.81	3.41	1.91	6.74	0.16	0.96
M47	3.07	0.00	0.50	0.55	0.38	2.15	0.09	0.02	0.00	0.06
M48	0.16	0.76	2.28	0.00	1.64	0.11	1.61	1.56	0.01	1.14
M49	0.53	0.12	2.08	0.15	0.04	0.05	0.11	1.20	0.01	1.22
M50	2.63	0.00	0.04	1.18	2.05	3.17	0.12	1.28	1.41	4.03
M51	0.10	0.03	0.00	2.65	0.53	0.25	0.05	0.35	0.01	3.71
M52	0.52	0.06	0.08	0.72	0.35	0.40	0.00	0.26	0.01	2.38
M53	0.17	0.01	0.01	0.10	1.81	0.27	0.63	1.07	0.38	0.17
M54	0.19	0.00	0.00	0.48	0.59	1.98	2.75	0.15	0.72	0.98
M55	0.45	0.10	0.06	1.02	0.83	0.26	0.26	0.11	2.23	0.01
M56	0.22	0.41	3.18	0.05	1.07	0.15	0.00	0.00	0.11	2.56
M57	0.26	1.46	2.25	0.96	0.09	0.54	1.33	2.38	1.00	0.84
M58	3.07	0.68	0.21	0.61	0.06	5.18	0.05	1.73	0.03	0.12
M59	1.92	0.05	1.48	2.77	0.78	0.45	7.25	1.73	1.07	1.98
M60	0.56	0.02	0.00	0.55	0.66	4.59	0.57	0.00	0.13	2.16

M61	0.26	0.38	1.51	0.01	0.18	2.25	0.00	0.10	0.24	0.20
M62	2.30	1.88	0.96	0.37	0.00	0.58	0.14	4.03	1.96	0.71
M63	0.17	1.90	0.27	0.02	1.67	0.52	0.57	3.40	0.04	0.52
M64	1.76	0.44	0.08	0.02	0.48	2.36	0.94	0.38	0.01	0.01
M65	0.36	1.12	0.52	0.16	0.01	2.15	7.46	2.40	3.40	2.11
M66	0.20	0.00	0.02	0.42	0.44	0.11	0.75	1.11	0.48	1.03
M67	2.71	1.51	0.70	1.57	0.34	3.04	0.00	2.47	0.02	3.02
M68	0.04	0.98	1.17	5.79	2.65	1.25	0.05	1.54	0.38	0.46
M69	0.04	0.24	0.25	0.32	3.23	0.25	0.28	0.10	0.51	0.27
M70	0.25	0.76	0.08	0.06	0.24	2.01	0.30	0.09	0.37	0.78
M71	2.14	0.05	0.02	2.73	0.31	0.11	0.27	0.00	0.32	2.92
M72	0.78	0.25	0.28	2.04	0.00	0.14	0.80	0.95	0.03	0.56
M73	0.04	0.75	1.02	0.14	0.05	0.59	1.77	5.58	0.73	0.11
M74	0.13	1.08	3.61	0.15	0.03	0.00	0.00	0.46	0.01	0.30
M75	3.55	1.12	4.31	2.41	0.04	1.64	0.10	1.18	0.00	0.02
M76	0.11	0.09	0.49	1.62	0.02	0.75	0.72	0.23	0.02	0.47
M77	0.44	0.10	0.00	8.30	0.11	1.90	2.37	0.00	0.01	1.63
M78	0.45	0.04	5.30	0.29	1.17	0.65	0.01	0.90	0.65	2.03
M79	0.37	0.03	0.02	0.81	1.38	0.42	4.35	0.92	2.35	1.21
M80	1.22	0.00	0.63	0.12	0.05	0.26	0.57	1.82	0.05	2.71
M81	2.84	3.20	4.68	0.06	0.00	9.94	0.79	5.27	0.65	0.23
M82	0.15	0.11	3.01	0.34	0.02	0.04	0.00	0.44	0.01	0.12
M83	0.89	0.32	0.19	2.28	1.96	3.43	2.80	0.14	0.04	3.42
M84	0.00	2.00	0.70	3.16	0.07	0.81	0.00	2.74	0.11	0.22
M85	1.13	1.04	1.04	0.01	0.00	0.02	0.00	2.03	0.77	0.78
M86	0.80	0.48	0.44	0.39	1.20	8.72	0.34	0.38	0.10	1.58
M87	0.84	0.36	0.60	0.16	0.05	0.40	0.01	1.49	0.18	0.86
M88	0.05	3.31	3.34	0.00	0.02	0.83	0.63	0.00	0.00	0.52
M89	0.35	0.09	1.46	0.30	1.43	0.43	0.12	1.75	0.83	0.01
M90	0.50	0.02	7.63	0.81	0.21	1.06	0.70	5.61	2.44	1.19
M91	0.00	2.21	0.13	0.01	0.33	0.17	0.42	0.28	0.48	0.00
M92	0.00	0.01	3.53	0.49	0.04	0.06	1.69	0.94	1.65	0.28
M93	0.00	3.54	0.02	0.58	0.18	0.59	0.71	0.00	0.95	0.05
M94	2.36	0.05	0.44	0.00	1.60	0.01	0.07	0.70	0.78	0.93
M95	0.80	0.42	2.82	0.00	0.00	1.37	0.02	0.08	0.37	0.05
M96	1.81	0.01	2.14	0.05	0.18	1.90	0.32	3.18	0.85	0.01
M97	0.03	0.04	2.73	1.38	0.81	8.61	0.92	0.00	0.02	0.63
M98	1.36	0.69	2.43	1.18	0.10	0.01	0.04	0.37	0.01	0.77
M99	1.34	0.07	3.69	2.17	0.07	1.87	0.03	0.02	0.57	1.13
M100	0.63	3.93	1.24	0.40	0.01	0.17	4.30	0.00	1.61	0.45

\* C Düzeyinde DMF

## Ek C. İkinci Birey Değişkenine Göre DMF Analizi Sonuçları

Lord'un Genelleştirilmiş $\chi^2$ İstatistiği										
-	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Madde	Madde Havuzu	Madde Havuzu	Madde Havuzu	Madde Havuzu	Madde Havuzu	Madde Havuzu	Madde Havuzu	Madde Havuzu	Madde Havuzu	Madde Havuzu
M1	0,27	2,29	1,01	1,34	0,55	1,10	0,60	0,52	1,23	1,00
M2	6,37	1,69	1,52	0,15	0,57	0,14	0,17	0,96	1,88	0,31
M3	0,11	1,68	1,24	0,93	1,61	1,68	0,67	0,81	0,24	1,08
M4	1,31	2,01	1,61	1,37	0,11	0,75	1,39	1,78	1,06	4,84
M5	0,83	0,69	6,18	2,25	0,03	5,22	0,11	2,11	0,85	2,15
<b>M6</b>	<b>156,37*</b>	<b>233,23*</b>	<b>310,01*</b>	<b>219,72*</b>	<b>65,50*</b>	<b>305,33*</b>	<b>88,50*</b>	<b>329,40*</b>	<b>90,29*</b>	<b>122,73*</b>
<b>M7</b>	<b>49,81*</b>	<b>177,49*</b>	<b>227,26*</b>	<b>213,44*</b>	<b>259,78*</b>	<b>296,38*</b>	<b>113,30*</b>	<b>229,14*</b>	<b>167,10*</b>	<b>193,43*</b>
<b>M8</b>	<b>193,30*</b>	<b>133,77*</b>	<b>234,56*</b>	<b>60,97*</b>	<b>44,81*</b>	<b>136,08*</b>	<b>217,56*</b>	<b>383,88*</b>	<b>148,75*</b>	<b>257,32*</b>
<b>M9</b>	<b>215,56*</b>	<b>399,94*</b>	<b>100,10*</b>	<b>137,18*</b>	<b>32,89*</b>	<b>127,80*</b>	<b>333,11*</b>	<b>118,59*</b>	<b>110,06*</b>	<b>144,65*</b>
<b>M10</b>	<b>195,93*</b>	<b>208,12*</b>	<b>265,68*</b>	<b>156,02*</b>	<b>42,27*</b>	<b>100,85*</b>	<b>134,08*</b>	<b>106,04*</b>	<b>118,29*</b>	<b>222,39*</b>
M11	1,97	3,53	0,99	0,05	0,01	2,23	0,78	0,26	1,69	0,82
M12	2,77	0,82	1,79	0,93	0,43	0,04	0,09	0,15	0,98	0,45
M13	1,15	0,63	1,56	4,22	0,16	2,15	2,62	0,82	2,47	5,57
M14	0,93	1,40	0,27	0,35	1,34	3,05	0,10	0,42	1,10	2,49
M15	0,31	0,25	1,28	2,82	1,13	0,96	2,13	0,97	2,57	0,17
M16	1,82	0,28	0,69	3,12	2,47	0,25	1,62	6,98	0,15	3,25
M17	0,48	2,89	2,30	0,55	1,28	0,03	1,78	2,30	0,96	0,24
M18	4,09	0,04	0,48	0,29	2,32	2,21	1,10	5,41	3,90	2,82
M19	0,83	1,67	0,98	0,28	0,16	0,16	1,03	2,56	2,92	1,11
M20	1,65	0,05	0,80	4,00	2,73	0,80	0,88	0,58	0,45	0,12
M21	1,22	4,42	0,70	1,74	0,01	2,02	2,11	0,49	6,33	1,74
M22	2,46	0,84	2,69	0,68	1,74	0,92	4,16	2,09	0,49	5,64
M23	0,10	1,89	0,50	2,43	0,00	0,33	1,22	0,04	1,10	2,28
M24	3,33	0,99	1,98	3,32	4,47	5,45	0,63	0,02	3,37	1,94
M25	1,58	0,12	1,11	0,27	0,24	0,70	0,32	0,55	1,96	0,31
M26	0,02	0,12	3,87	3,63	1,49	0,40	1,72	1,19	0,26	2,56
M27	0,01	3,97	1,08	0,75	0,61	1,88	0,05	1,62	0,31	0,89
M28	1,23	2,28	1,10	1,81	1,45	0,60	0,57	0,12	5,19	0,79
M29	1,95	1,28	2,35	1,28	0,49	1,18	3,52	3,25	0,65	1,17
M30	0,95	2,56	2,71	4,31	1,60	1,04	1,46	0,69	1,12	2,95
M31	0,86	0,63	3,99	4,07	1,41	0,04	3,24	3,84	1,56	1,03
M32	1,98	0,12	3,94	1,81	0,32	0,46	5,51	0,86	2,79	2,65
M33	2,15	0,06	1,51	2,74	0,01	0,43	4,66	1,97	3,81	0,11
M34	0,59	0,32	0,60	0,12	0,05	0,78	3,91	5,36	0,39	0,07
M35	1,07	1,19	1,81	1,52	2,81	3,18	3,56	2,74	7,68	0,96
M36	4,18	1,20	0,45	0,56	1,05	5,44	0,70	1,72	0,05	0,83
M37	0,38	1,86	2,72	4,05	0,79	0,63	0,60	2,62	1,29	1,28
M38	7,25	1,60	1,52	0,02	0,79	0,26	1,86	4,70	1,69	1,61
M39	8,60	0,53	7,91	1,08	0,98	0,30	0,33	6,34	4,58	2,87
M40	0,57	0,47	1,33	1,87	1,53	1,53	5,18	0,55	0,06	3,84
M41	5,36	1,13	1,34	0,05	0,09	4,27	0,57	0,54	0,87	0,20

M42	1,12	7,97	2,90	0,59	2,36	12,80	0,79	1,66	2,68	1,32
M43	3,29	0,14	0,30	1,21	0,85	5,16	0,09	1,29	0,52	0,95
M44	0,81	0,27	0,61	0,35	0,08	0,52	0,87	0,12	1,28	3,09
M45	1,30	3,41	2,51	1,85	4,07	1,05	0,23	0,01	0,50	1,00
M46	1,14	2,06	3,50	3,96	0,26	1,93	3,35	1,89	1,93	0,13
M47	0,38	0,82	0,41	2,59	1,16	0,61	0,08	2,06	3,00	2,44
M48	0,99	1,41	3,30	1,68	0,07	0,84	0,02	3,25	0,22	1,23
M49	0,27	3,96	2,27	0,08	5,77	3,00	0,28	3,51	1,78	4,29
M50	0,18	3,33	4,03	0,70	0,55	1,75	0,54	1,27	3,42	4,17
M51	0,32	0,49	0,21	0,45	0,26	3,67	0,10	3,13	1,06	0,42
M52	2,40	1,15	2,55	0,98	3,98	1,01	0,90	5,71	1,21	2,48
M53	0,23	0,51	2,01	0,42	0,85	1,81	1,74	3,38	0,49	8,37
M54	0,54	0,33	3,11	0,40	1,17	4,25	0,75	1,79	1,08	1,05
M55	0,45	7,49	6,47	0,59	1,00	0,18	0,16	3,28	1,54	0,11
M56	1,75	1,92	1,66	1,96	0,39	5,80	0,61	2,66	0,65	2,24
M57	0,44	1,67	0,84	1,74	1,18	1,18	0,82	1,11	1,27	3,08
M58	0,13	1,52	0,06	3,64	0,80	0,47	2,78	0,23	0,43	0,37
M59	5,42	0,50	0,67	0,01	1,32	0,47	1,17	1,16	1,81	0,26
M60	0,72	0,84	1,17	2,52	1,81	2,13	0,37	0,50	3,38	4,49
M61	1,62	2,01	2,37	1,32	1,27	0,09	0,36	0,53	0,61	2,06
M62	1,04	4,43	0,37	0,27	0,63	1,04	6,13	4,61	4,95	1,63
M63	6,84	0,33	5,75	2,91	0,80	0,67	0,34	3,06	0,99	2,78
M64	3,20	0,21	5,03	0,85	0,87	0,46	2,41	1,20	1,17	3,52
M65	9,67	0,74	2,70	4,11	1,71	1,57	1,03	0,98	0,57	3,23
M66	0,30	7,82	0,92	0,24	0,99	0,21	0,28	2,08	0,07	0,14
M67	0,17	0,44	2,40	0,49	0,34	0,25	0,33	0,46	0,98	0,46
M68	0,38	0,01	0,01	0,24	0,69	1,53	0,46	1,47	2,27	1,69
M69	1,71	1,88	4,63	2,19	0,09	3,10	2,13	2,80	0,77	2,15
M70	1,29	1,58	1,83	3,40	0,44	0,52	1,47	1,22	0,74	0,71
M71	0,30	0,10	0,81	0,68	0,63	1,33	0,47	2,96	1,01	2,00
M72	4,72	0,74	0,13	0,47	0,06	4,23	1,32	1,16	4,28	1,75
M73	0,26	0,90	0,10	2,19	1,21	0,01	3,16	0,40	0,69	2,60
M74	0,47	4,45	5,06	2,64	0,64	4,27	1,13	0,67	4,01	1,18
M75	3,47	0,10	0,03	0,62	0,62	0,86	5,47	0,85	3,09	2,90
M76	0,72	0,93	1,69	6,23	0,43	6,77	0,22	1,77	2,47	5,37
M77	2,98	0,95	2,55	1,43	0,04	0,20	1,95	1,30	0,80	0,06
M78	0,21	0,86	0,99	0,21	0,83	1,23	2,42	0,39	1,34	4,99
M79	1,09	0,12	1,39	0,06	2,14	1,27	4,09	0,58	0,27	0,85
M80	0,68	1,88	0,20	0,32	0,98	0,72	10,29	1,89	1,32	4,01
M81	1,78	0,36	5,58	1,11	6,78	1,10	2,03	0,27	2,13	2,99
M82	0,16	4,56	1,01	1,15	0,73	1,09	0,55	10,16	0,59	0,19
M83	0,98	0,01	0,01	0,33	0,30	0,06	0,61	0,70	1,09	1,10
M84	0,68	0,07	1,58	1,03	2,77	2,46	0,03	0,23	4,55	0,02
M85	7,30	0,50	0,68	3,13	1,40	0,17	2,42	0,30	2,98	0,07
M86	0,54	0,61	0,64	4,12	0,57	0,11	1,10	0,96	0,77	0,11
M87	2,91	1,87	0,72	0,73	1,24	0,30	1,80	1,08	3,91	3,39

M88	2,42	0,45	9,61	0,26	0,86	2,06	7,13	1,10	0,25	2,21
M89	2,45	1,68	0,29	0,45	1,22	0,61	1,70	0,94	0,62	0,28
M90	3,87	3,19	2,08	0,12	0,30	3,15	2,81	1,14	0,40	1,62
M91	0,33	0,13	1,15	0,59	1,61	3,44	2,72	1,25	0,28	1,96
M92	3,03	0,99	2,44	4,50	0,38	0,64	5,58	0,54	0,30	0,31
M93	0,42	2,64	12,85	0,48	1,29	0,77	2,08	0,38	0,34	0,59
M94	1,74	0,83	3,58	5,73	0,57	2,03	1,30	2,81	1,11	0,96
M95	2,60	0,25	2,87	5,26	3,54	0,80	2,57	1,99	0,12	0,08
M96	0,34	0,96	0,26	0,07	3,60	1,60	3,72	0,87	0,66	0,68
M97	1,61	1,27	1,07	0,25	1,49	0,82	0,20	0,91	0,77	1,66
M98	1,31	0,00	1,37	0,74	0,98	0,19	2,68	5,08	1,10	1,56
M99	1,31	0,22	0,03	0,26	1,64	0,68	2,54	0,15	3,88	1,10
M100	0,88	2,61	2,66	1,59	0,25	1,38	0,40	0,59	0,44	2,09

\* C Düzeyinde DMF



## Ek Ç. Birinci Madde Değişkenine Göre DMF Analizi Sonuçları

Madde	Lord'un $\chi^2$ İstatistiği									
	1. Madde Havuzu	2. Madde Havuzu	3. Madde Havuzu	4. Madde Havuzu	5. Madde Havuzu	6. Madde Havuzu	7. Madde Havuzu	8. Madde Havuzu	9. Madde Havuzu	10. Madde Havuzu
M1	0,09	0,00	0,09	0,47	0,03	1,14	0,31	0,31	0,69	1,00
M2	2,69	0,18	0,22	0,41	0,15	0,11	0,72	0,67	0,14	0,00
M3	0,22	7,57	0,88	0,62	0,41	1,19	0,33	0,24	0,00	0,06
M4	0,00	0,93	0,17	0,18	3,06	0,01	0,33	0,01	0,00	0,12
M5	4,33	0,16	1,01	1,52	0,10	1,60	0,49	0,00	0,34	0,16
M6	0,30	0,47	0,03	0,34	0,50	1,63	0,00	0,20	0,93	0,28
M7	0,13	0,56	1,11	1,96	0,05	0,51	0,60	0,00	1,03	1,10
M8	0,18	1,35	0,04	1,20	0,64	1,25	0,10	0,64	1,25	0,25
M9	0,48	0,38	1,09	0,03	1,39	0,24	0,13	0,13	0,25	1,01
M10	0,25	0,94	0,04	1,06	0,59	1,22	2,50	2,10	0,34	0,01
<b>M11</b>	<b>313,86*</b>	<b>87,06*</b>	<b>41,18*</b>	<b>263,64*</b>	<b>174,98*</b>	<b>246,24*</b>	<b>84,45*</b>	<b>98,34*</b>	<b>388,98*</b>	<b>114,60*</b>
<b>M12</b>	<b>6,90*</b>	<b>223,66*</b>	<b>209,58*</b>	<b>174,10*</b>	<b>192,66*</b>	<b>358,91*</b>	<b>164,53*</b>	<b>97,15*</b>	<b>122,83*</b>	<b>236,66*</b>
<b>M13</b>	<b>408,28*</b>	<b>282,72*</b>	<b>438,49*</b>	<b>179,02*</b>	<b>389,54*</b>	<b>17,60*</b>	<b>195,04*</b>	<b>158,35*</b>	<b>109,39*</b>	<b>24,86*</b>
<b>M14</b>	<b>219,12*</b>	<b>305,39*</b>	<b>64,94*</b>	<b>152,20*</b>	<b>261,97*</b>	<b>151,42*</b>	<b>415,50*</b>	<b>344,36*</b>	<b>43,80*</b>	<b>42,05*</b>
<b>M15</b>	<b>140,82*</b>	<b>359,38*</b>	<b>251,17*</b>	<b>131,94*</b>	<b>87,54*</b>	<b>376,11*</b>	<b>248,37*</b>	<b>343,09*</b>	<b>69,77*</b>	<b>260,12*</b>
M16	0,03	0,82	1,96	0,24	0,24	0,84	0,00	0,08	0,02	0,69
M17	0,60	0,14	2,71	0,30	0,04	0,15	1,51	0,58	0,45	0,16
M18	1,44	1,40	1,11	6,38	0,03	0,01	0,04	0,02	0,03	0,33
M19	1,10	5,01	4,31	0,08	0,03	0,01	3,61	0,64	0,52	0,02
M20	0,27	0,02	0,99	0,28	0,36	0,29	1,60	2,89	0,17	1,44
M21	0,42	0,00	0,16	2,43	0,35	0,63	0,38	0,42	0,64	0,06
M22	1,54	0,15	0,11	0,77	0,73	1,58	0,00	2,20	0,14	4,75
M23	0,00	0,12	0,38	1,05	0,10	0,19	0,97	0,20	0,16	0,38
M24	0,01	0,04	0,01	0,10	0,23	0,43	0,03	1,76	0,04	0,06
M25	0,11	0,53	0,13	0,33	0,50	0,88	0,09	1,40	2,15	0,05
M26	0,44	0,98	0,02	0,03	2,46	0,00	0,11	0,08	0,27	1,39
M27	1,47	0,01	0,09	0,49	1,13	0,01	0,14	0,18	0,18	0,33
M28	1,34	0,04	0,33	1,52	2,05	0,75	0,14	2,22	0,61	0,97
M29	0,00	0,00	2,14	0,05	0,22	0,03	2,40	0,92	0,03	1,15
M30	0,12	0,31	0,01	0,08	1,16	0,24	5,46	0,31	0,03	0,61
M31	1,30	2,10	0,02	0,05	0,71	0,17	1,87	0,50	1,18	0,98
M32	2,05	0,81	0,07	0,08	1,34	0,90	1,33	2,43	0,16	2,31
M33	0,46	1,51	0,63	1,20	0,10	1,10	0,07	2,27	1,99	0,18
M34	0,72	0,52	0,91	1,51	1,53	0,22	0,34	0,05	0,06	0,24
M35	0,00	0,09	1,32	2,61	0,14	0,00	0,04	0,02	2,49	0,10
M36	0,03	0,87	1,23	0,05	2,34	0,04	0,00	0,56	0,02	3,62
M37	0,78	0,16	0,09	0,49	0,15	0,12	0,37	2,24	0,33	0,20
M38	0,51	2,45	0,37	0,86	0,88	1,81	0,24	0,20	1,99	1,21
M39	1,30	0,74	0,15	0,58	5,49	0,01	0,64	0,86	1,15	4,23
M40	0,81	0,20	0,27	0,03	0,66	0,11	0,51	0,81	2,12	0,84
M41	0,20	0,03	1,40	0,86	0,00	0,40	0,49	0,65	0,08	0,02
M42	0,20	0,15	0,39	0,03	0,81	1,22	0,29	2,72	1,50	0,35
M43	0,59	1,05	0,32	0,01	0,55	0,05	0,04	3,28	0,81	0,99
M44	0,07	0,00	0,02	0,37	5,78	0,13	6,67	0,05	0,00	0,25
M45	1,60	0,79	0,00	0,05	0,13	2,11	0,12	0,14	0,05	0,09
M46	1,89	5,12	0,13	0,23	0,07	0,76	0,05	0,40	0,38	0,26
M47	0,05	3,49	0,21	1,29	1,87	0,05	0,31	0,17	0,13	0,20
M48	0,76	0,00	5,21	0,06	5,35	0,00	0,89	0,03	0,49	0,16
M49	0,00	0,46	1,00	0,03	0,13	0,18	1,24	0,52	0,27	1,37
M50	1,25	0,00	0,60	1,09	0,98	0,06	0,94	1,44	0,10	0,00
M51	3,61	0,37	0,01	0,54	0,33	0,81	0,59	0,74	0,01	1,35
M52	0,00	1,13	0,01	0,23	0,12	1,51	0,04	0,15	8,25	0,63
M53	0,10	0,93	3,67	0,63	0,00	0,06	7,36	0,24	0,41	0,24
M54	1,67	0,07	0,10	0,05	1,16	0,10	0,00	0,37	1,03	0,61
M55	3,34	1,08	1,49	0,99	3,64	0,01	0,00	0,37	0,02	0,18
M56	0,19	0,73	0,28	0,17	0,00	0,12	5,77	0,25	0,35	0,03

M57	0,08	0,40	0,98	0,49	0,06	0,66	0,06	0,35	0,39	0,02
M58	0,13	1,46	0,16	0,00	0,59	0,14	0,89	0,15	0,19	3,34
M59	0,37	0,99	1,69	0,12	0,06	0,02	2,48	0,31	1,93	0,20
M60	0,00	3,45	1,04	0,01	0,00	0,22	0,09	0,02	0,41	0,49
M61	0,12	0,21	5,69	0,62	0,98	0,61	0,39	0,31	0,36	0,69
M62	0,75	1,83	0,96	1,24	1,59	0,83	0,04	0,07	0,59	0,59
M63	0,01	1,43	0,08	0,21	0,05	0,00	2,66	0,64	0,89	0,05
M64	0,22	0,12	1,32	0,78	0,13	1,95	1,40	0,12	1,93	2,93
M65	0,44	0,00	0,21	0,58	0,06	0,50	0,00	2,59	0,75	0,31
M66	1,52	0,54	1,10	0,34	0,12	0,35	2,43	0,31	0,17	0,03
M67	0,02	1,34	0,05	0,93	0,00	0,35	0,20	0,03	0,01	0,09
M68	0,56	0,81	0,00	0,03	0,21	2,61	0,91	0,77	0,02	1,11
M69	0,32	0,03	0,00	0,01	0,08	1,12	0,97	0,91	1,02	2,71
M70	0,00	0,01	0,32	0,28	3,00	0,09	0,02	0,43	0,54	0,26
M71	4,81	0,02	0,01	0,06	0,19	0,58	0,82	0,05	0,34	0,25
M72	0,19	0,22	2,66	0,63	0,67	0,00	3,53	0,28	0,07	0,69
M73	1,32	0,27	2,93	0,85	0,01	1,55	0,80	0,37	0,01	1,09
M74	0,88	2,63	0,28	0,03	0,85	0,06	0,08	0,05	0,70	1,34
M75	0,12	0,02	6,97	0,54	0,00	1,10	0,09	0,10	0,17	0,26
M76	0,00	0,71	1,91	0,08	0,30	0,72	0,97	0,66	0,01	1,02
M77	0,09	3,08	0,00	0,19	0,00	0,20	0,87	3,78	0,51	4,43
M78	2,04	0,38	0,21	0,15	0,69	0,01	0,00	0,85	5,03	0,05
M79	0,09	0,22	2,37	0,00	0,05	2,09	0,06	0,07	1,04	0,02
M80	0,01	0,06	0,82	0,02	0,32	1,47	0,69	0,20	3,48	1,23
M81	0,69	2,72	0,02	0,96	0,97	0,39	3,04	0,00	0,00	0,70
M82	0,42	0,98	0,44	0,14	0,94	0,50	1,20	0,49	0,11	1,73
M83	0,04	0,55	0,06	0,21	11,99	0,01	0,32	0,03	0,90	0,35
M84	0,00	0,95	0,15	0,69	0,09	0,00	0,84	0,10	0,53	0,17
M85	4,02	0,70	2,76	0,01	4,80	0,17	0,93	0,13	0,06	3,25
M86	0,24	2,58	0,05	1,43	2,24	0,07	0,06	0,00	1,81	1,53
M87	0,65	0,57	0,19	0,52	0,42	0,59	0,02	0,44	1,60	1,23
M88	0,01	0,70	0,62	0,17	0,09	0,03	0,28	1,29	0,09	0,61
M89	4,37	1,15	0,06	0,76	3,21	2,67	4,72	0,26	0,08	0,88
M90	5,30	0,05	0,00	0,46	0,01	0,01	0,29	2,06	0,26	0,26
M91	0,19	0,01	2,55	0,01	0,65	0,01	0,08	1,81	0,07	1,33
M92	2,05	0,29	2,07	0,05	2,05	1,38	0,51	2,07	0,18	0,54
M93	1,89	0,35	1,23	1,35	0,07	0,09	1,45	0,75	0,11	0,59
M94	0,73	0,31	0,04	0,00	0,01	0,14	0,18	0,35	0,00	0,00
M95	0,62	0,00	0,26	3,79	0,23	0,16	0,82	0,43	2,29	1,10
M96	0,67	0,22	1,30	0,98	0,13	0,27	1,39	0,09	1,83	1,87
M97	4,63	2,81	0,07	0,81	0,04	0,05	1,16	0,15	0,17	0,33
M98	5,39	0,47	1,74	1,11	0,07	3,16	0,05	0,00	0,10	0,61
M99	0,71	0,36	0,80	0,96	0,06	0,01	0,84	2,21	0,18	2,06
M100	0,95	2,95	0,11	1,75	0,00	0,22	0,20	0,89	0,52	0,70

\* C düzeyinde DMF

## Ek D. İkinci Madde Değişkenine Göre DMF Analizi Sonuçları

Lord'un Genelleştirilmiş $\chi^2$ İstatistiği										
Madde	1. Madde Havuzu	2. Madde Havuzu	3. Madde Havuzu	4. Madde Havuzu	5. Madde Havuzu	6. Madde Havuzu	7. Madde Havuzu	8. Madde Havuzu	9. Madde Havuzu	10. Madde Havuzu
M1	2.78	0.51	0.69	1.10	4.80	0.18	2.23	2.06	1.06	1.30
M2	1.45	0.39	1.59	0.11	1.43	0.93	0.29	3.28	1.20	0.93
M3	1.47	3.71	1.45	1.01	1.97	1.18	1.43	2.05	5.02	0.15
M4	0.14	0.36	0.89	1.67	0.23	0.58	0.54	0.83	2.54	1.92
M5	0.45	1.38	0.04	0.71	3.02	1.11	0.09	0.70	8.22	0.37
M6	0.65	1.68	0.31	4.56	0.75	1.57	0.04	1.05	0.26	8.65
M7	1.05	0.22	0.20	0.06	4.62	1.72	2.47	2.18	3.82	0.69
M8	1.80	2.86	0.52	1.66	0.59	4.92	0.91	0.56	1.80	0.54
M9	0.17	0.74	0.40	0.27	1.06	0.40	1.19	1.09	1.29	0.99
M10	1.50	5.85	0.51	0.61	0.83	0.16	1.39	0.62	0.06	3.34
M11	0.17	0.72	2.18	0.42	0.12	1.42	0.71	2.32	1.65	1.83
M12	1.80	3.64	0.52	0.23	0.97	0.25	0.65	4.36	0.27	0.18
M13	0.29	0.78	0.70	5.65	0.88	0.60	0.84	0.86	0.33	1.05
M14	0.06	0.55	0.30	0.97	0.90	0.25	0.07	0.49	3.46	6.48
M15	2.60	0.76	0.09	4.97	4.83	0.53	1.21	0.29	4.73	1.13
<b>M16</b>	<b>141.92*</b>	<b>297.94*</b>	<b>43.95*</b>	<b>171.51*</b>	<b>168.85*</b>	<b>183.42*</b>	<b>315.52*</b>	<b>217.64*</b>	<b>59.70*</b>	<b>298.42*</b>
<b>M17</b>	<b>176.84*</b>	<b>80.24*</b>	<b>130.16*</b>	<b>142.13*</b>	<b>150.30*</b>	<b>259.11*</b>	<b>111.87*</b>	<b>115.28*</b>	<b>130.86*</b>	<b>214.58*</b>
<b>M18</b>	<b>132.54*</b>	<b>69.53*</b>	<b>37.29*</b>	<b>64.25*</b>	<b>38.83*</b>	<b>432.33*</b>	<b>112.68*</b>	<b>229.66*</b>	<b>281.07*</b>	<b>147.85*</b>
<b>M19</b>	<b>99.63*</b>	<b>76.58*</b>	<b>131.28*</b>	<b>109.39*</b>	<b>70.83*</b>	<b>70.73*</b>	<b>287.03*</b>	<b>123.30*</b>	<b>208.25*</b>	<b>75.66*</b>
<b>M20</b>	<b>257.90*</b>	<b>198.41*</b>	<b>81.57*</b>	<b>243.73*</b>	<b>124.59*</b>	<b>153.87*</b>	<b>79.28*</b>	<b>281.33*</b>	<b>74.09*</b>	<b>361.28*</b>
M21	1.37	2.70	1.64	0.51	0.30	0.47	1.14	0.12	0.34	0.01
M22	2.18	9.68	0.27	7.24	1.53	9.13	3.10	0.36	0.22	0.61
M23	0.37	0.26	1.86	0.15	1.11	2.71	1.38	3.55	7.76	0.36
M24	2.18	0.86	0.03	0.84	0.34	0.56	8.95	0.07	1.64	1.57
M25	2.75	0.92	0.07	2.51	1.11	0.69	1.06	2.80	1.93	0.00
M26	0.19	2.42	0.43	5.11	0.36	4.35	0.54	2.20	1.80	0.38
M27	0.73	2.35	0.42	0.68	0.18	0.63	3.34	1.02	0.46	0.48
M28	0.81	0.72	3.42	0.74	0.15	0.33	12.12	0.88	0.24	2.89
M29	2.54	3.55	0.57	6.77	0.93	3.60	0.13	1.78	0.42	1.33
M30	1.06	1.95	0.16	1.41	4.31	0.95	4.02	1.86	1.45	2.00
M31	0.98	0.15	1.28	1.57	3.09	0.18	0.98	1.31	0.53	3.11
M32	0.37	0.18	0.21	2.20	5.19	2.67	0.54	3.31	0.84	0.14
M33	2.39	8.38	1.65	0.79	0.60	5.71	0.90	0.10	0.46	0.01
M34	1.30	1.30	0.21	0.49	0.04	1.04	0.16	15.50	0.13	3.09
M35	2.56	3.13	1.52	0.74	2.68	1.44	2.14	0.01	1.59	0.63
M36	0.36	2.45	2.54	0.88	0.80	0.56	1.20	1.91	4.21	1.62
M37	0.78	1.65	0.22	0.17	1.14	0.00	1.25	0.49	0.60	0.61
M38	0.34	7.39	2.18	2.85	2.52	1.01	0.71	2.59	0.12	3.00
M39	0.02	1.19	1.15	1.32	2.12	6.04	0.18	1.66	0.23	0.46
M40	8.42	2.19	1.16	0.07	4.05	0.38	3.62	0.30	1.53	0.66
M41	4.88	1.32	0.74	0.51	7.67	2.36	1.46	5.79	1.00	0.64
M42	0.38	2.52	3.05	3.26	1.08	0.04	2.67	0.85	0.14	2.63
M43	0.83	0.29	2.66	1.67	2.08	0.04	0.71	1.06	1.27	0.69
M44	2.51	2.59	1.01	1.16	0.38	1.00	0.84	0.21	0.95	2.43
M45	0.72	4.13	0.99	1.80	1.85	1.36	0.62	0.31	0.89	3.85
M46	0.88	1.79	0.27	0.09	0.90	0.24	3.03	6.66	1.27	4.30
M47	0.23	2.14	1.27	3.29	1.43	0.09	0.58	3.01	1.69	1.15
M48	4.90	0.86	1.36	0.26	0.13	2.69	2.09	1.22	3.57	0.23
M49	0.31	3.52	5.90	2.09	0.63	0.40	2.83	2.42	0.52	0.41
M50	0.92	0.79	0.74	3.18	0.02	1.62	0.15	2.43	0.28	0.51
M51	1.16	1.99	0.52	4.24	3.81	1.74	0.96	2.00	0.59	0.81
M52	8.37	1.32	1.37	1.60	0.91	0.86	2.74	4.75	0.98	2.26
M53	0.84	0.86	2.22	0.44	1.28	0.84	0.11	1.08	0.01	0.68
M54	4.61	1.49	1.07	1.30	4.23	1.89	3.59	0.61	1.12	4.03
M55	0.47	7.59	1.07	0.67	0.16	0.10	0.28	0.63	0.60	0.84
M56	1.38	0.54	1.13	2.10	10.63	0.58	5.06	1.02	2.80	1.38
M57	0.54	2.49	4.25	0.28	0.60	0.29	4.18	4.22	3.07	9.58
M58	0.34	0.37	4.52	0.14	3.69	0.26	1.23	1.40	0.58	3.30
M59	0.67	0.34	0.19	0.76	0.10	2.80	1.38	0.70	0.83	2.09
M60	0.18	0.49	0.20	0.38	0.80	0.71	0.56	3.92	2.63	0.12

M61	0.96	0.63	0.78	2.39	5.44	0.71	0.23	0.01	0.82	0.00
M62	0.94	0.40	1.06	1.36	3.44	0.50	0.79	1.66	2.61	1.55
M63	1.70	0.01	0.25	0.56	3.92	2.17	0.48	3.91	6.25	1.01
M64	0.75	2.13	0.31	5.17	1.09	1.30	0.25	0.24	1.31	0.74
M65	1.85	1.45	0.85	1.25	1.75	0.21	0.44	0.24	1.07	0.26
M66	0.06	3.76	0.33	1.61	1.70	0.12	0.60	1.11	1.43	0.63
M67	1.17	0.24	0.01	1.01	4.20	1.86	1.51	3.65	1.54	2.90
M68	2.24	2.00	1.45	0.98	0.84	0.18	0.43	1.27	1.24	0.57
M69	1.08	1.49	0.18	2.22	0.05	0.99	1.11	1.29	3.09	1.42
M70	0.90	4.25	0.17	1.38	0.07	0.01	1.23	8.22	0.27	0.42
M71	1.07	0.99	0.14	1.88	2.22	1.39	0.26	0.00	0.39	0.88
M72	2.41	0.54	0.83	2.83	0.75	0.92	1.22	1.01	1.33	0.43
M73	1.03	0.34	1.76	2.73	0.28	0.85	0.60	4.07	0.74	2.26
M74	1.77	0.22	0.27	1.03	1.93	2.98	2.77	0.49	0.52	8.33
M75	1.15	1.45	0.33	1.18	1.92	1.79	0.04	2.47	9.37	1.39
M76	6.10	5.26	0.43	2.24	0.36	0.41	0.63	2.16	3.95	0.20
M77	0.83	0.48	0.47	0.12	0.26	2.09	7.70	1.59	4.70	0.41
M78	0.68	4.69	0.12	2.48	0.58	3.72	0.27	2.44	0.09	3.18
M79	2.26	3.82	0.54	1.95	0.05	1.21	0.64	2.19	1.53	1.27
M80	1.37	1.88	0.25	2.94	0.04	0.07	2.43	0.54	0.78	0.58
M81	0.98	1.44	2.64	0.83	0.33	0.18	1.71	2.52	1.15	2.48
M82	0.55	0.42	1.16	3.29	1.82	1.28	0.59	0.16	5.66	3.21
M83	2.04	0.68	0.12	0.54	0.44	8.98	1.21	2.59	2.55	0.15
M84	0.33	6.36	0.85	0.25	1.98	1.43	0.52	7.31	0.38	1.12
M85	2.07	1.56	0.63	2.02	0.92	0.49	0.53	1.14	2.52	0.59
M86	0.48	0.76	0.97	0.75	0.52	0.71	1.68	2.69	6.82	1.86
M87	0.52	1.02	1.04	3.36	1.14	1.43	1.78	1.88	0.00	5.54
M88	3.49	0.14	2.57	0.54	2.05	1.36	1.19	0.74	1.50	0.06
M89	2.26	5.45	0.49	4.04	0.42	1.71	0.25	0.83	1.60	0.35
M90	0.75	1.89	0.25	2.21	1.46	0.94	1.86	1.17	0.13	0.87
M91	1.46	1.60	0.55	1.11	0.01	0.35	0.58	3.31	0.01	0.89
M92	0.45	0.88	0.84	1.55	1.77	1.48	1.93	0.08	0.60	0.64
M93	0.85	2.71	0.22	2.83	3.69	4.35	3.34	0.67	0.80	0.16
M94	0.39	2.04	1.58	0.18	1.25	2.09	2.30	2.04	0.62	0.59
M95	0.17	1.97	0.65	1.52	1.17	2.43	0.35	0.89	0.83	2.11
M96	2.26	1.41	1.88	0.95	3.88	2.42	0.52	0.70	0.75	1.36
M97	0.08	0.14	0.39	5.67	0.59	1.90	4.65	0.52	1.26	0.78
M98	2.76	1.34	0.11	0.79	0.75	1.76	5.23	4.63	0.75	0.22
M99	0.48	0.19	2.80	2.55	0.80	1.59	1.86	0.68	0.28	1.83
M100	0.91	0.20	0.00	2.58	0.41	1.62	2.14	0.94	4.26	1.03

\* C düzeyinde DMF

## EK-E: Etik Komisyonu Onay Bildirimi



T.C.  
HACETTEPE ÜNİVERSİTESİ  
Rektörlük

Tarih: 21.12.2018 16:43  
Sayı: 35853172-300-E.00000381495  
  
E.00000381495

Sayı : 35853172-300  
Konu : Osman TAT Hk.

### EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Enstitünüz Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı doktora öğrencilerinden **Osman TAT**'in **Prof. Dr. Nuri DOĞAN** danışmanlığında yürüttüğü "**Açıklayıcı Madde Tepki Modellerinin Bilgisayar Ortamından Bireye Uyarlanmış Testlerde Kullanımı**" başlıklı tez çalışması, Üniversitemiz Senatosu Etik Komisyonunun **18 Aralık 2018** tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini saygılarımla rica ederim.

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

Evrakın elektronik imzalı suretine <https://belgedogrulama.hacettepe.edu.tr> adresinden 66751f231e4474438ea5188453ba5728499766 kabul edilmiş ve onaylanmıştır.  
Bu belge 5070 sayılı Elektronik İmza Kanunu'na uygun olarak Güvenli Elektronik İmza ile imzalanmıştır.

## EK-F: Etik Beyanı

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

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin bütününi 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.

30/12/2019



Osman TAT

## EK-G: Doktora Tez Çalışması Orijinallik Raporu

30/12/2019

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

**Tez Başlığı: Açıklayıcı Madde Tepki Modellerinin Bilgisayar Ortamında Bireye Uyarlanmış Testlerde Kullanımı**

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
30/12/2019	105	159987	03/12/2019	% 3	1238747661

Uygulanan filtreler:

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

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

Ad Soyadı: Osman TAT

Öğrenci No.: N15141536


Ana Bilim Dalı: Eğitim Bilimleri

Programı: Eğitimde Ölçme ve Değerlendirme-Doktora

Statüsü:  Y.Lisans  Doktora  Bütünleşik Dr.

İmza  


**DANIŞMAN ONAYI**

  
UYGUNDUR.  
(Prof. Dr., Nuri DOĞAN)

## EK-Ğ: Dissertation Originality Report

30/12/2019

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

Thesis Title: Use of Explanatory Item Response Models in Computer Adaptive Tests

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using Turnitin plagiarism detection software taking 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
30/12/2019	105	159987	03/12/2019	% 3	1238747661

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: Osman TAT  
Student No.: N15141536  
Department: Educational Sciences  
Program: Educational Measurement and Evaluation  
Status:  Masters  Ph.D.  Integrated Ph.D.

Signature  


ADVISOR APPROVAL

  
APPROVED

(Prof. Dr., Nuri DOĞAN)

## EK-H: Yayınlama ve Fikrî Mülkiyet Hakları Beyanı

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

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

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

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

30 /12/2019



Osman TAT

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

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

\* Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

