



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

Eğitim Bilimleri Ana Bilim Dalı  
Eğitimde Ölçme ve Değerlendirme Programı

BİLİŞSEL TANI VE ÇOK BOYUTLU MADDE TEPKİ MODELLERİNİN SINIFLAMA  
DOĞRULUĞU VE PARAMETRELERİNİN KARŞILAŞTIRILMASI

Elif Özlem ARDIÇ

Doktora Tezi

Ankara, 2020



Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

*Daha ileriye ... En İyiyeye ...*



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

Eğitim Bilimleri Ana Bilim Dalı  
Eğitimde Ölçme ve Değerlendirme Programı

BİLİŞSEL TANI VE ÇOK BOYUTLU MADDE TEPKİ MODELLERİNİN SINIFLAMA  
DOĞRULUĞU VE PARAMETRELERİNİN KARŞILAŞTIRILMASI

COMPARISON OF CLASSIFICATION ACCURACY AND PARAMETERS OF  
COGNITIVE DIAGNOSTIC AND MULTIDIMENSIONAL ITEM RESPONSE  
MODELS

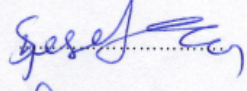
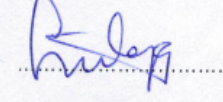
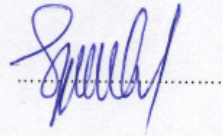
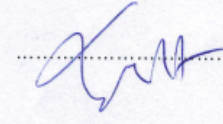
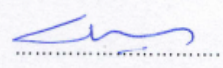
Elif Özlem ARDIÇ

Doktora Tezi

Ankara, 2020

### Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,  
Elif zlem Ardıç'ın hazırladıđı "Bilişsel Tanı ve Çok Boyutlu Madde Tepki Modellerinin Sınıflama Doğruluđu ve Parametrelerinin Karşılaştıırılması" başlıklı bu çalışma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde lçme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiştir.

J¼ri Başkanı	Prof. Dr. Şeref TAN	
J¼ri Üyesi (Danışman)	Prof. Dr. H¼lya KELECİOđLU	
J¼ri Üyesi	Prof. Dr. Selahattin GELBAL	
J¼ri Üyesi	Prof. Dr. Nuri DOđAN	
J¼ri Üyesi	Doç.Dr. Tahsin BAŞOKÇU	Ođuz 

Bu tez Hacettepe Üniversitesi Lisans¼st¼ Eđitim, đretim ve Sınav Ynetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından 16 / 01/ 2020 tarihinde uygun gr¼lm¼ş ve Enstit¼ Ynetim Kurulunca .... / .... / ..... tarihi itibarıyla kabul edilmiştir.

Prof. Dr. Ali Ekber ŞAHİN  
Eđitim Bilimleri Enstit¼s¼ M¼d¼r¼

## Öz

Bu çalışmada, 3PL çok boyutlu madde tepki kuramı (ÇB-MTK) ve bilişsel tanı modelleri (BTM) ile kestirilen birey parametrelerinin sınıflama doğruluğunu incelemek üzerine simülasyon çalışması ve gerçek veri uygulaması yapılmıştır. Modelleri karşılaştırmak için birey parametre kestirimlerine ilişkin RMSE, yanlılık ve nitelik ve nitelik profilleri için doğru sınıflama oranları hesaplanmıştır. Araştırmanın sonucunda tüm test koşulları için verinin üretildiği 3PL ÇB-MTK modelinin en düşük hata ve yanlılık değerine sahip olduğu bulunmuştur. Test uzunluğu ve nitelikler arası korelasyonun artması ile hata değerlerinin azaldığı ancak hata değerlerinin düşmesindeki en etkili değişkenin madde sayısı olduğu görülmüştür. Yanlılık değerleri ise test uzunluğu ve nitelikler arası korelasyonun artması ile birkaç istisnai durum dışında azalma eğilimi göstermiştir. Daha fazla soru ile ölçülen 1. ve 4. nitelikler için nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanılması hata ve yanlılığı arttırırken, diğer nitelikler için hata ve yanlılığın azaldığı gözlenmiştir. Tüm simülasyon koşulları için nitelik ve nitelik profili bazında en yüksek doğru sınıflama oranına 3PL ÇB-MTK'nın sahip olduğu görülmüştür. Nitelik bazında doğru sınıflama oranlarının 0.728 ile 0.890, nitelik profili bazında doğru sınıflama oranlarının ise 0.135 ile 0.711 arasında değiştiği belirlenmiştir. Testteki madde sayısı ve nitelikler arası korelasyonun artması ile tüm modeller için nitelik ve nitelik profili bazında yapılan doğru sınıflama oranları monotonik olarak artmıştır. Nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ise daha yüksek oranda doğru sınıflama yapılmasına neden olmuştur. Gerçek veri uygulamasına dayalı sonuçlar incelendiğinde, G-DINA modelinin diğer BTM'lere göre mutlak ve göreceli model uyumu ve madde uyumu indekslerine göre daha iyi uyum gösterdiği bulunmuştur. Her bir nitelik sınıfı için geçen kişi yüzdeleri incelendiğinde ise bireylerin büyük bir çoğunluğunun (1,1,1) ve (0,0,0) sınıflarında toplandığı görülmüştür.

**Anahtar sözcükler:** bilişsel tanı modelleri, çok boyutlu madde tepki kuramı, uyarılma, doğru sınıflama oranı

## Abstract

This study investigated the classification accuracy of person parameters estimated by 3PL multidimensional item response theory (MIRT) and cognitive diagnostic models (CDMs) via simulation and real data analysis. To compare the models, RMSE, bias, and correct classification rates of attributes and attribute profiles were calculated. As a result of the study, it was found that for all test conditions, 3PL MIRT had the lowest RMSE and bias values. The increase in test length and the correlation between attributes decreased the RMSE and test length had the greatest positive impact on decreasing RMSE values than the correlation between attributes. By increasing test length and the correlation between attributes, bias tended to decrease except for a few exceptional cases. The use of Q-matrix which was designed to measure the attributes and attributes profiles in equal number, increased the RMSE and bias for the attributes 1 and 4 which were measured by more items, and reduced RMSE and bias for other attributes. For all test conditions, 3PL MIRT had the highest attribute and attribute profile correct classification rates. It was found that the correct classification rates based on attributes ranged from 0.728 to 0.890 and the correct classification rates of attribute profiles were between 0.135 and 0.711. As the test length and correlation increased, attribute and attribute profile based correct classification rates increased monotonically. Similarly, the use of the Q-matrix which measured the attributes and the attribute profiles in equal number, increased the correct classification rates. In real data applications, G-DINA model had the best relative and absolute model fit and item fit among the other CDMs. When examining the attribute mastery probabilities, it was found that the majority of the examinees were clustered in (1,1,1) and (0,0,0) attribute classes.

**Keywords:** cognitive diagnostic models, multidimensional item response theory, retrofitting, correct classification rate

## Teşekkür

Doktora eğitimim boyunca bana her zaman destek olan, bilgisine, tecrübesine ve çalışma disiplinine hayranlık duyduğum değerli danışmanım Prof. Dr. Hülya KELECIÖĞLU'na;

İhtiyacım olan her durumda bana yardımcı olan, tez izleme jürimde bulunarak tezime katkı sağlayan sevgili hocam Prof. Dr. Nuri DOĞAN'a; Hacettepe Üniversitesi'nde görev yaptığım süre boyunca bana ve mesai arkadaşlarıma rahat ve huzurlu bir çalışma ortamı sağlayan değerli hocam Prof. Dr. Selahattin GELBAL'a; tez izleme jürimde yer alan kıymetli hocam Doç Dr. Tahsin Oğuz BAŞOKÇU'ya;

Bana University of Alberta'da eğitim alma imkanı tanıyan, engin tecrübe ve bilgilerini her zaman benimle paylaşan çok kıymetli hocam Prof. Dr. Mark J. GIERL'e ve bu süreçte bana her türlü konuda yardımcı olan, akademik olarak kendime örnek aldığım Doç Dr. Okan BULUT'a ve birlikte çalışmaktan keyif aldığım, bitmek bilmeyen sorularımı bıkmadan cevaplayan sevgili arkadaşım Dr. Qi Guo'ya;

Doktora eğitimim boyunca her zaman yanımda olan sevgili arkadaşım Arş. Gör. Mine DEMİRBAŞ'a ve tezimi yazarken ortaya çıkan tüm problemleri benimle birlikte çözmeye çalışan ve yardımlarını benden hiç esirgemeyen canım arkadaşım Dr. Sümeyra SOYSAL'a;

Tezimi yazdığım dönem boyunca çalışabilmem için bana her türlü kolaylığı sağlayan ve heyecanımı benimle birlikte yaşayan Trabzon Üniversitesi'ndeki değerli çalışma arkadaşlarım Arş. Gör. Yasemin KARSANTIK, Arş. Gör. Münevver ÖZDEMİR, Arş. Gör. Cansu TOSUN ve Dr. Öğr. Üyesi Fatma ALTUN'a;

Doktora eğitimim süresince yurtiçi ve yurt dışında sağladığı burs imkânları için TÜBİTAK'a,

Tüm başarılarımı borçlu olduğum, hayattaki en büyük şansım olan canım annem Kevser ARDIÇ, canım babam Hüseyin ARDIÇ ve canım kardeşim Merve ARDIÇ'a sonsuz teşekkürlerimi sunarım.

## İçindekiler

Öz.....	ii
Abstract.....	iii
Teşekkür.....	iv
Tablolar Dizini.....	viii
Şekiller Dizini.....	x
Simgeler ve Kısaltmalar Dizini.....	xiii
Bölüm 1 Giriş.....	1
Problem Durumu.....	2
Araştırmanın Amacı ve Önemi.....	4
Problem Cümlesi.....	5
Alt Problemler.....	5
Sınırlılıklar.....	6
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	7
Madde Tepki Kuramı.....	7
Çok Boyutlu Madde Tepki Kuramı.....	8
Çok Boyutluluk için Genişletilmiş Üç Parametrelili Lojistik Model.....	9
Bilişsel Tanı Modellerinin Temelleri.....	10
Q-Matrisi.....	11
Bilişsel Tanı Modelleri.....	13
Bilişsel Tanı ve Çok Boyutlu Madde Tepki Modellerinin Karşılaştırılması.....	21
Bilişsel Tanı Modellerinin Uyarlanması (Retrofitting).....	22
İlgili Araştırmalar.....	23
Bilişsel Tanı Modellerinin Uyarlanması ile İlgili Araştırmalar.....	24
Çok Boyutlu MTK ve Bilişsel Tanı Modellerinin Karşılaştırılması ile İlgili Araştırmalar.....	32



İlgili Araştırmalar Özet.....	37
Bölüm 3 Yöntem.....	39
Araştırmanın Türü.....	39
Verilerin Elde Edilmesi.....	39
Matematik Testinin Faktör Yapısı.....	41
Model Veri Uyumunun İncelenmesi.....	46
Q-Matrisinin Oluşturulması.....	48
Bilişsel Tanı Modellerinin Model Veri Uyumunun İncelenmesi.....	50
Sınıflama Doğruluğu ve Tutarlılığının Değerlendirilmesi.....	53
Simülasyon Koşulları.....	55
Verilerin Üretilmesi ve Analizi.....	58
Değerlendirme Kriteri.....	59
Bölüm 4 Bulgular ve Yorumlar.....	61
Birinci Alt Probleme İlişkin Bulgular ve Yorumlar.....	61
İkinci Alt Probleme İlişkin Bulgular ve Yorumlar.....	82
Üçüncü ve Dördüncü Alt Problemlere İlişkin Bulgular ve Yorumlar.....	112
Beşinci Alt Probleme İlişkin Bulgular ve Yorumlar.....	117
Bölüm 5 Sonuç, Tartışma ve Öneriler.....	119
Sonuç ve Tartışma.....	119
Öneriler.....	125
Kaynaklar.....	128
EK-A: TIMSS Matematik Testinin 1, 2 ve 4 Boyutlu Formlarına İlişkin Faktör Yük Değerleri.....	144
EK-B: Testin Farklı Sayıda Boyutlardan Oluşan Formlarına İlişkin EM, QMCEM ve MHRM Yöntemlerine Göre Kestirilen Mutlak ve Göreceli Model Uyumu İndeksleri.....	145

EK-C: BTM'lere göre Kestirilen Madde Uyum İstatistikleri.....	146
EK-Ç: Çalışmada Kullanılan Q-Matrisleri.....	160
EK-D: BTM'lere göre Kestirilen Model Uyum İstatistikleri .....	162
EK-E: Etik Kurul İzin Muafiyeti Formu.....	163
EK-F: Etik Beyanı .....	164
EK-G: Doktora Tez Çalışması Orijinallik Raporu .....	165
EK-Ğ: Dissertation Originality Report .....	166
EK-H: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı.....	167



## Tablolar Dizini

Tablo 1 Çalışma Kapsamında Kullanılan Bilişsel Tanı Modelleri.....	15
Tablo 2 TIMSS Matematik Testi 8. Kitapçığın Maddelerine ait Betimsel İstatistikler .....	41
Tablo 3 Testin Farklı Sayıda Boyutlardan Oluşan Formlarına İlişkin Uyum İndeksleri .....	42
Tablo 4 Testin Farklı Sayıda Boyutlardan Oluşan Formlarına İlişkin Mutlak Model Uyumu İndeksleri.....	43
Tablo 5 Testin 3 ve 4 Boyutlu Formlarına İlişkin $X^2$ 'ler İçin Ölçeklendirilmiş Fark Testi Sonuçları .....	44
Tablo 6 Testin 29 Maddelik Formunda Yer Alan Maddelere İlişkin Faktör Yük Değerleri.....	45
Tablo 7 MHRM Yöntemine göre Kestirilen 1PL, 2PL ve 3PL Modellere İlişkin Mutlak ve Göreceli Model Uyumu İndeksleri.....	47
Tablo 8 Q-Matrisi ve 3PL ÇB-MTK Modeli Madde Parametreleri .....	48
Tablo 9 Q-Matrisinin Son Hali.....	49
Tablo 10 Maddelerin Örtük Sınıflara göre Dağılımı .....	49
Tablo 11 Bilişsel Tanı Modellerine İlişkin Mutlak Model Uyumu İndeksleri.....	50
Tablo 12 Bilişsel Tanı Modellerine İlişkin Göreceli Model Uyumu İndeksleri .....	51
Tablo 13 Bilişsel Tanı Modellerine İlişkin Madde Uyum İstatistikleri.....	52
Tablo 14 Sınıflama Doğruluğu ve Tutarlığının Bilişsel Tanı Modellerine göre Dağılımı.....	54
Tablo 15 TIMSS 2015 Matematik Sınavının Alt Testlerine ait Madde Parametreleri .....	59
Tablo 16 Üç Nitelikten Oluşan Tek Boyutlu Veri Setleri için Birey Parametrelerine İlişkin Hata Değerleri .....	62
Tablo 17 Üç Nitelikten Oluşan Tek Boyutlu Veri Setleri için Birey Parametrelerine İlişkin Yanıllık Değerleri .....	72
Tablo 18 Altı Nitelikten Oluşan İki Boyutlu Veri Setleri için Birey Parametrelerine İlişkin Hata Değerleri .....	83

Tablo 19 <i>Altı Nitelikten Oluşan İki Boyutlu Veri Setleri için Birey Parametrelerine İlişkin Yanlılık Değerleri</i> .....	98
Tablo 20 <i>Üç Nitelikten Oluşan Tek Boyutlu Veri Setleri için Doğru Sınıflama Oranları</i> .....	113
Tablo 21 <i>Altı Nitelikten Oluşan İki Boyutlu Veri Setleri için Doğru Sınıflama Oranları</i> .....	115
Tablo 22 <i>Nitelik ve Nitelik Sınıfları için Geçen Kişi Oranları</i> .....	117



## Şekiller Dizini

Şekil 1. Madde uyum istatistikleri. ....	52
Şekil 2. Üç nitelikten oluşan tek boyutlu veri setleri için ÇB-MTK'ya göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	63
Şekil 3. Üç nitelikten oluşan tek boyutlu veri setleri için DINA'ya göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	64
Şekil 4. Üç nitelikten oluşan tek boyutlu veri setleri için DINO'ya göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	65
Şekil 5. Üç nitelikten oluşan tek boyutlu veri setleri için A-CDM'ye göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	66
Şekil 6. Üç nitelikten oluşan tek boyutlu veri setleri için LLM'ye göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	67
Şekil 7. Üç nitelikten oluşan tek boyutlu veri setleri için R-RUM'a göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	68
Şekil 8. Üç nitelikten oluşan tek boyutlu veri setleri için G-DINA'ya göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	69
Şekil 9. Üç nitelikten oluşan tek boyutlu veri setleri için HO-DINA'ya göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	70
Şekil 10. Üç nitelikten oluşan tek boyutlu veri setleri için ÇB-MTK'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	73
Şekil 11. Üç nitelikten oluşan tek boyutlu veri setleri için DINA'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	74
Şekil 12. Üç nitelikten oluşan tek boyutlu veri setleri için DINO'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	75
Şekil 13. Üç nitelikten oluşan tek boyutlu veri setleri için A-CDM'ye göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	76
Şekil 14. Üç nitelikten oluşan tek boyutlu veri setleri için LLM'ye göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	77
Şekil 15. Üç nitelikten oluşan tek boyutlu veri setleri için R-RUM'a göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	78

Şekil 16. Üç nitelikten oluşan tek boyutlu veri setleri için G-DINA'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	79
Şekil 17. Üç nitelikten oluşan tek boyutlu veri setleri için HO-DINA'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	80
Şekil 18. Altı nitelikten oluşan iki boyutlu veri setleri için ÇB-MTK'ya göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	85
Şekil 19. Altı nitelikten oluşan iki boyutlu veri setleri için DINA'ya göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	86
Şekil 20. Altı nitelikten oluşan iki boyutlu veri setleri için DINO'ya göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	88
Şekil 21. Altı nitelikten oluşan iki boyutlu veri setleri için A-CDM'ye göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	89
Şekil 22. Altı nitelikten oluşan iki boyutlu veri setleri için LLM'ye göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	91
Şekil 23. Altı nitelikten oluşan iki boyutlu veri setleri için R-RUM'a göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	92
Şekil 24. Altı nitelikten oluşan iki boyutlu veri setleri için G-DINA'ya göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	94
Şekil 25. Altı nitelikten oluşan iki boyutlu veri setleri için HO-DINA'ya göre kestirilen birey parametrelerine ilişkin hata değerleri. ....	95
Şekil 26. Altı nitelikten oluşan iki boyutlu veri setleri için ÇB-MTK'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	100
Şekil 27. Altı nitelikten oluşan iki boyutlu veri setleri için DINA'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	101
Şekil 28. Altı nitelikten oluşan iki boyutlu veri setleri için DINO'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	103
Şekil 29. Altı nitelikten oluşan iki boyutlu veri setleri için A-CDM'ye göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	104
Şekil 30. Altı nitelikten oluşan iki boyutlu veri setleri için LLM'ye göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	106
Şekil 31. Altı nitelikten oluşan iki boyutlu veri setleri için R-RUM'a göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	107

Şekil 32. Altı nitelikten oluşan iki boyutlu veri setleri için G-DINA'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	109
Şekil 33. Altı nitelikten oluşan iki boyutlu veri setleri için HO-DINA'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri. ....	110



## Simgeler ve Kısaltmalar Dizini

**A-CDM:** Additive Cognitive Diagnostic Model

**AFA:** Açımlayıcı Faktör Analizi

**AIC:** Akaike Information Criterion

**BIC:** Bayesian Information Criterion

**BTM:** Bilişsel Tanı Modeli

**C-RUM:** Compensatory Reparameterized Unified Model

**ÇB-MTK:** Çok Boyutlu Madde Tepki Kuramı

**DFA:** Doğrulayıcı Faktör Analizi

**DINA:** Deterministic Inputs, Noisy “and” Gate Model

**DINO:** Deterministic Inputs, Noisy “Or” Gate Model

**G-DINA:** Generalized Deterministic Inputs, Noisy “and” Gate Model

**HO-DINA:** Higher Order DINA Model

**KTK:** Klasik Test Kuramı

**LLM:** Linear Logistic Model

**MKE:** Madde Karakteristik Eğrisi

**MKF:** Madde Karakteristik Fonksiyonu

**MTK:** Madde Tepki Kuramı

**R-RUM:** Reduced Reparameterized Unified Model

**RMSE:** Root Mean Square Error

**TIMSS:** Trends in International Mathematics and Science Study



## Bölüm 1

### Giriş

Eğitim ve psikolojideki testler, genellikle, içerik alanı, nitelik ya da becerilere dayalı olarak çeşitli alanları ölçmek üzere tasarlanmıştır ve bu alt alanlara atanan puanlar alt test veya tanı puanı olarak adlandırılmaktadır (Haberman, Sinharay ve Puhan, 2009; Sinharay, Haberman ve Puhan, 2007). Ancak eğitimdeki birçok test geniş bir konu alanını kapsayacak şekilde tasarlanmasına rağmen değerlendirmelere, öğrencilerin geniş bir içerik alanındaki başarı düzeylerini gösteren toplam puanlarının rapor edildiği, tek boyutlu özetleyici testler hakimdir (de Ayala, 2009; Hambleton, Swaminathan ve Rodgers, 1991; Yao ve Boughton, 2007). Bu testler ile her bir birey, toplam madde puanı ya da tek bir sürekli gizil yeteneğe göre sıralanır ve geniş kapsamlı yetenekleri, yalnızca tek bir toplam puana göre rapor edilir (Wang, 2009). Bu modeller, kavramsal ve matematiksel olarak basit olması ve elde edilen puanlar bireyleri sıralama, karşılaştırma ve hatta gelecekteki performanslarını tahmin etmede kullanışlı olmasına rağmen bireylerin güçlü ve zayıf yönlerinin belirlenmesine ilişkin sınırlı miktarda tanılayıcı bilgi içermektedir (Choi, 2010; de la Torre ve Karelitz, 2009). Özellikle performansından mutlu olmayan bireyler için daha güçlü ve zayıf yönlerini bilmek, kendilerini bu alanlarda geliştirmeleri açısından önem taşımaktadır. Bu nedenle, yapılan değerlendirmelerin her bir bireyin hangi becerilere sahip olduğu ve olmadığı konusunda daha biçimlendirici ve bilgilendirici olması, başarılı olmaları için ne çalışmalarını ve hangi becerilerini geliştirmeleri gerektiği konusunda rehberlik etmesi büyük önem arz etmektedir. Nitekim Ulusal Araştırma Konseyi (2001) "Knowing What Students Know" raporunda da yapılan değerlendirmeler ile bireylere biçimlendirici ve tanılayıcı bilgilerin sağlanması gerektiği vurgulanmaktadır. Benzer şekilde Amerikan hükümetinin "Hiçbir Çocuk Geride Kalmasın Yasası" (2001), öğretmenlerin öğrencilerinin spesifik ihtiyaçlarını gidermelerine yardımcı olacak, tanılayıcı raporlar sağlaması gerektiğini belirtmektedir. Sonuç olarak, özetleyici tek bir toplam test puanının biçimlendirici bilgi sağlamadaki yetersizliği, bireyin her bir konu alanındaki performansına yönelik daha kapsamlı ve tanılayıcı bilgi sağlayan değerlendirmelere olan ilgiyi arttırmıştır. Bu talepleri karşılamak ve değerlendirme sonuçlarından daha fazla faydalanabilmek adına, tanılayıcı değerlendirmeler giderek daha fazla rağbet

görmeye başlamıştır (Liu, Huggins-Manley ve Bulut, 2017a; Rupp, Templin ve Henson, 2010; Sinharay, Puhan ve Haberman, 2010).

## **Problem Durumu**

Günümüzde, gelişmiş ülkelerde, eğitim ve psikolojik değerlendirmelerde kullanılan birçok büyük ölçekli standart test, çoklu yetenek ya da yapıları ölçmek üzere tasarlanmıştır. Bu testler toplam puanların yanı sıra, bireylere geliştirmeleri gereken becerileri konusunda yardımcı olmak için her bir alt konu alanındaki performansına dayalı olarak alt test puanları da sunmaktadır. Bu alt test puanları bireylerin güçlü ve zayıf yönlerine yönelik tanılayıcı bilgi sağlamanın yanı sıra, bireylerin kabul edilmesi ve sınıflandırılması gibi onlar hakkında verilecek önemli kararların alımında da kullanılabilir. Ayrıca öğretim programının etkililiğini değerlendirmek, eğitim ve öğretim programlarının geleceğiyle ilgili alınacak kararlarda da kullanılabilir (Haladyna ve Kramer, 2004). Nitekim eğitimde yapılan ölçmeler, geniş bir içerik alanı için tek bir puanın raporlandığı değerlendirmelerden birden çok daha küçük alt alana ilişkin katılımcıların başarılı olup olmadıklarına dair kararların verildiği değerlendirmelere doğru bir değişim göstermektedir (Liu, Huggins-Manley ve Bulut, 2017).

Araştırmacılar, tanılayıcı geri bildirim ihtiyacını karşılayabilmek için güvenilir ve geçerli çok boyutlu puanlar üretebilmek üzere çeşitli modellerin kullanımını araştırmıştır (Jurich, 2014). Literatürde alt test puanlarını raporlamak için klasik test kuramı (KTK), tek boyutlu madde tepki kuramı, augmented madde tepki kuramı, çok boyutlu madde tepki kuramı (ÇB-MTK), bilişsel tanı modelleri (BTM) ve faktör analizi çerçevesindeki modeller kullanılmıştır. KTK çerçevesinde popüler alt test raporlama yöntemleri Haberman'ın (2008) çalışmasına dayanmaktadır ve alt test puanı kestirimlerinin; 1) gözlenen alt test puanlarının fonksiyonu, 2) gözlenen toplam puanların bir fonksiyonu ve 3) gözlenen toplam puan ve gözlenen alt test puanlarının fonksiyonu olduğu yöntemleri içermektedir. Çok boyutlu MTK modelleri (ÇB-MTK) ise karmaşık yapıdaki performansı modelleme potansiyeline sahiptir, çünkü soruyu çözmek için birden fazla yeteneğin gerektiği durumlarda bile çoklu yetenekleri aynı anda hesaplayabilir (Hartig ve Hohler, 2009). ÇB-MTK modelleri eğitimsel ölçme alanı içerisinde tanılayıcı dönüt üretmek için ilgi uyandırmasına rağmen daha detaylı çok boyutlu dönüt verme konusundaki talepleri gidermede asıl dikkat çeken nokta,

bilişsel tanı modellerinin (BTM) hızlı gelişimi ve yaygınlaşmasıdır (Rupp ve Templin, 2008b; Rupp vd., 2010).

Tanılayıcı sınıflama modelleri (diagnostic classification models) olarak da bilinen bilişsel tanı modelleri, doğrulayıcı örtük sınıf modelleridir ve gözlenen yanıtların genel olarak nitelikler olarak adlandırılan kategorik örtük/gizil değişkenlerle olan ilişkisini karakterize etmek için kullanılırlar (Templin ve Bradshaw, 2013). Geleneksel ölçme modellerinin aksine BTM, özellikle çok daha detaylı bir şekilde yeterliliklerin belirlenmesi, öğrencinin öğrenme veya anlama durumu hakkında spesifik bilgi sağlanması amacıyla kullanılır ve öğretimin iyileştirilmesi gibi daha sonraki adımlarla da ilişkilidir (de la Torre ve Karelitz, 2009).

BTM, özellikle bir bilişsel model tarafından geliştirildiğinde, biliş ve öğrenme teorilerini öğretile ilişkilendirmede önemli bir rol oynayabilir (Gierl ve Cui, 2008). BTM'nin en ayırt edici özelliği, birey profillerinin daha karmaşık ve çok boyutlu gösterimini sağlamalarıdır (Bolt, 2007). BTM, tek bir yetenek puanı yerine, bireylere testte ölçülen çoklu becerilerdeki yeterlilikleri üzerine detaylı geri bildirim sağlayan profiller oluşturmaktadır. Böylece bireyler, henüz yeterince başarılı olmadıkları becerileri görebilirler. Yeni bir ölçme yaklaşımı olarak BTM'yi alt test puanlarını raporlamada kullanmak mümkündür (Liu, Huggins-Manley ve Bulut, 2017; Rupp vd., 2010, Sinharay, Puhon ve Haberman, 2011). BTM'de testteki her bir maddeyi doğru cevaplamak için hangi özellik veya becerilerin gerekli olduğu, Q-matrisi adı verilen bir matrisle tanımlanır. BTM çerçevesinde alt test puanları, Q-matrisindeki bilişsel bilgi yapıları olarak kavramsallaştırılmıştır (Roberts ve Gierl, 2010). Bilişsel tanı modelleri MTK modelleri gibi bireyin yerini ölçekte belirleme imkânı sunmamasına rağmen kullanımı (açıkça belirtilmese de) sınıflamaya dayalı ölçümlerin daha güvenilir olduğu düşüncesine dayanmaktadır (Templin ve Bradshaw, 2013). Sürekli puanları kestirdikten sonra kesme puanlarına dayalı olarak bireyleri sınıflandıran iki aşamalı yaklaşımlarla (örneğin, MTK modelleri gibi) karşılaştırıldığında ise BTM ile sınıflama hatası azaltılabilir (Robitzsch, Kiefer, George ve Uenlue, 2019). Nitekim Templin ve Bradshaw (2013), BTM'nin MTK modellerine göre aynı uzunlukta testler için bireyin yeteneğini daha yüksek bir güvenilirlikle kestirdiğini göstermiştir. Bu durum, BTM çerçevesinde birey kestirimlerinin daha küçük gizil değişken değer aralığında olmasından kaynaklanmaktadır.

## Araştırmanın Amacı ve Önemi

Teknolojinin gelişmesiyle bilgisayar temelli testler eğitimsel değerlendirmelerin bir parçası haline gelmiştir. Geleneksel kağıt-kalem testleri ile karşılaştırıldığında ise bilgisayar temelli testler daha sık uygulanmaktadır. Dolayısıyla sınavdan başarısız olan bireylerin gelecek çalışmalarını etkili bir şekilde planlayarak, kısa bir süre içerisinde zayıf oldukları konu/konuları gidermeleri gerekmektedir. Bu nedenle bireylere yardımcı olması ve rehberlik etmesi açısından alt test puanlarının raporlanmasının önemi artmaktadır.

Alt testlerin temel amacı bireylerin güçlü ve zayıf yönlerine ilişkin güvenilir tanılayıcı bilgi sağlamak olmasına rağmen literatürdeki çalışmaların birçoğu alt test raporlarının kullanılmasından ziyade alt test puanlarının toplam test puanlarına ek bilgi sağlayıp sağlamadığı, alt testlerin sağladığı ek bilgilerin alt gruplar arasında değişip değişmediği, alt test puanlarının psikometrik nitelikleri, farklı alt test yöntemlerinin güvenilirlik ve sınıflama tutarlıklarının karşılaştırılması ve geçerliği konularına odaklanmaktadır (Haberman ve Sinharay, 2010; Haladyna ve Kramer, 2004; Longabach ve Peyton, 2017; Reckase ve Xu, 2015; Sawaki ve Sinharay, 2017; Sinharay vd., 2007; Sinharay vd., 2010; Sinharay vd., 2011; Sinharay ve Haberman, 2014; Yao ve Boughton, 2007).

Bu çalışmanın amacı, 3PL ÇB-MTK ve BTM modelleri ile kestirilen birey parametrelerinin farklı test koşulları altında sınıflama doğruluğunu karşılaştırmak ve hangi koşullar altında benzer ve farklı sonuç verdiklerini incelemektir. ÇB-MTK, BTM'lere umut verici bir alternatif gibi gözükmese de rağmen bireyleri test ile ölçülen özelliklere sahip olma durumlarına göre sınıflandırarak, tanılayıcı dönüt sağlama konusunda yetersiz kalmaktadır. Bu nedenle mevcut testlerin BTM kapsamında konu alanı puanlarının raporlanması ile ÇB-MTK ve BTM'ye göre yapılan sınıflandırmaların ve modellerin birbirlerinin yerine kullanımının incelenmesinin önemli olduğu düşünülmektedir. Alanyazında ÇB-MTK ve BTM'yi karşılaştıran çalışmaların büyük bir çoğunluğu simülatif veriye göre yapılırken, bu çalışmayı önemli kılan noktalardan biri simülasyon verilerinin gerçek verinin madde parametrelerine göre üretilmesi ve BTM'ler kapsamında gerçek verinin faktör yapısına göre oluşturulan Q-matrislerinin kullanılmasıdır. Son olarak, çalışma

kapsamındaki modellerin birey parametre kestirimleri arasındaki farklılıkları incelemek üzere gerçek veri üzerinde uygulama yapılmıştır.

### **Problem Cümlesi**

Farklı test koşulları altında üretilen verilerin bilişsel tanı ve çok boyutlu madde tepki modelleri ile analizinden elde edilen birey parametre kestirimlerinin hata ve yanlılık değerleri ve sınıflama doğruluğu nasıl değişmektedir?

### **Alt Problemler**

1) Üç nitelikten oluşan tek boyutlu veri setleri için ÇB-MTK ve BTM'ye göre kestirilen birey parametrelerine ilişkin hata ve yanlılık değerleri,

- a) testteki madde sayısına (15 ve 30)
- b) nitelikler arası korelasyona (0.1, 0.4 ve 0.7)
- c) Q-matrisi yapısına (gerçek verinin faktör yapısına göre oluşturulan Q-matrisi ve nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi) göre nasıl değişmektedir?

2) Altı nitelikten oluşan iki boyutlu veri setleri için ÇB-MTK ve BTM'ye göre kestirilen birey parametrelerine ilişkin hata ve yanlılık değerleri,

- a) testteki madde sayısına (15 ve 30)
- b) nitelikler arası korelasyona (0.1, 0.4 ve 0.7)
- c) Q-matrisi yapısına (gerçek verinin faktör yapısına göre oluşturulan Q-matrisi ve nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi) göre nasıl değişmektedir?

3) Üç nitelikten oluşan tek boyutlu veri setleri için ÇB-MTK ve BTM'ye göre kestirilen birey parametrelerinin her bir nitelik ve nitelik profili için doğru sınıflama oranları,

- a) testteki madde sayısına (15 ve 30)
- b) nitelikler arası korelasyona (0.1, 0.4 ve 0.7)
- c) Q-matrisi yapısına (gerçek verinin faktör yapısına göre oluşturulan Q-matrisi ve nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi) göre nasıl değişmektedir?

4) Altı nitelikten oluşan iki boyutlu veri setleri için ÇB-MTK ve BTM'ye göre kestirilen birey parametrelerinin her bir nitelik ve nitelik profili için doğru sınıflama oranları,

- a) testteki madde sayısına (15 ve 30)
- b) nitelikler arası korelasyona (0.1, 0.4 ve 0.7)
- c) Q-matrisi yapısına (gerçek verinin faktör yapısına göre oluşturulan Q-matrisi ve nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi) göre nasıl değişmektedir?

5) TIMSS 2015 uygulaması matematik testi verilerinin 3PL ÇB-MTK ve BTM ile analizinden kestirilen örtük sınıflar nasıl değişmektedir?

### **Sınırlılıklar**

1. Bu araştırmanın simülatif verileri, TIMSS 2015 uygulaması 8. kitapçıkta yer alan matematik testinin madde parametrelerine göre a parametresi ranjı [0.6, 2.5] arasında olacak şekilde log-normal dağılımdan,  $a \sim \ln N(1.5, 0.8)$ ; b parametresi ranjı [-2, 1.5] arasında olacak şekilde normal dağılımdan,  $b \sim N(0, 1)$ ; c parametresi ise ranjı [0, 0.30] arasında olacak şekilde beta dağılımdan,  $c \sim \text{Beta}(20, 100)$  üretilmiştir.

2. Araştırmada birey parametrelerinin sınıflama doğruluğuna ilişkin kestirimler 3PL ÇB-MTK ve BTM'lerden DINA, DINO, A-CDM, LLM, R-RUM, G-DINA ve HO-DINA modelleri ile sınırlıdır.

3. Araştırma kapsamında TIMSS 2015 uygulaması 8. kitapçıkta yer alan tüm sorular kullanılmış ve soruların hepsi yayınlanmadığı için Q-matrisinin oluşturulmasında istatistiksel yöntemlerden faktör analizi kullanılmıştır. Faktör analizi sonuçlarına göre sorular karmaşık bir yapı sergilediği için niteliklerin ölçtükleri konu alanları isimlendirilememiştir.

## Bölüm 2

### Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

#### Araştırmanın Kuramsal Temeli

Bu bölümde, ÇB-MTK ve BTM'nin kuramsal temeli ve analizlerde kullanılacak modeller hakkında bilgi verilecektir.

#### Madde Tepki Kuramı

Klasik test kuramının (KTK) birey ve test özelliklerini birbirinden ayırmaması, ölçmenin standart hatasının testi alan bireylerin yetenek düzeyi dikkate alınmaksızın tüm bireyler için eşit sayılması, farklı testleri alan bireylerin puanlarının karşılaştırılmaması, bireyin belli bir test maddesinde nasıl bir performans göstereceği hakkında tahminde bulunulamaması gibi sınırlılıklarına çözüm olarak madde tepki kuramı (MTK) geliştirilmiştir (Crocker ve Algina, 1986; Embretson ve Reise, 2000; Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991; Lord, 1980). MTK, bireylerin test performansı ile performansın altında yatan gizil özellikler arasında bir ilişki tanımlamaya çalışan model temelli bir ölçme tekniğidir (Hambleton ve Swaminathan, 1985; Henard, 2000). Hambleton ve Swaminathan (1985) tarafından MTK'nın temel iki özelliğini: 1) bireyin bir test maddesindeki performansının, örtük özellik veya yetenek olarak adlandırılan faktörler tarafından açıklanabileceği ve 2) bireyin yetenek düzeyindeki değişim ile soruyu doğru cevaplama olasılığı arasındaki ilişkinin madde karakteristik fonksiyonu (MKF) veya madde karakteristik eğrisi (MKE) olarak adlandırılan, monotonik olarak artan bir fonksiyon ile tanımlanabileceği şeklinde belirtilmiştir. KTK'nın aksine MTK, oldukça güçlü ancak test edilebilir varsayımlara dayanmaktadır (Hambleton ve Swaminathan, 1985; Embretson ve Reise, 2000). MTK modelleri, modelin uygulandığı veri hakkında iki temel varsayımının sağlanmasını gerektirmektedir: 1) MKE'nin belli bir şekle sahip olması ve 2) yerel bağımsızlık (Embretson ve Reise, 2000, s.45). MKE'nin şekli bireyin yetenek düzeyinin, soruyu doğru cevaplama olasılığına göre nasıl değiştiğini tanımlamaktadır. S şeklindeki bu eğriye göre, bireyin yetenek düzeyi arttıkça, soruyu doğru cevaplama olasılığı da artmaktadır. MKE'ler genel bir şekle sahip olmasına rağmen fonksiyonda yer alan  $b$

parametresine göre konumu, *a* parametresine göre eğimi ve *c* parametresine göre başlangıç noktası farklılık göstermektedir. MTK'nın diğer bir varsayımı olan yerel bağımsızlık ise aynı yetenek düzeyindeki bireylerin, madde puanlarının koşullu dağılımının birbirinden bağımsız olmasıdır (Lord ve Novick, 1968).

MTK modelleri arasındaki temel ayırım, olasılık fonksiyonunda maddeleri tanımlamak için kullanılan parametre sayısından kaynaklanmaktadır. MTK'da iki kategorili puanlanan maddeler kullanılan parametre sayısına göre bir parametrelili (1PL), iki parametrelili (2PL), üç parametrelili (3PL) ve dört parametrelili (4PL) model şeklinde sınıflandırılabilir.

### **Çok Boyutlu Madde Tepki Kuramı**

MTK'nın temel varsayımlarından biri, testteki tüm maddelerin aynı özellik ya da yeteneği ölçmesidir. Ancak eğitim ve psikolojide ilgilenilen birçok yapının çok boyutlu olması nedeniyle KTK ve tek boyutlu MTK gibi tek boyutluluk varsayımını temel alan modeller yetersiz kalabilmektedir. Nitekim Lee (2007), tek bir örtük özelliği ölçen bir test geliştirmenin neredeyse imkansız olduğunu belirtmiştir. Bu nedenle, birey ve madde arasındaki daha karmaşık etkileşimleri açıklamak için faktör analizi ve tek boyutlu MTK'nın bir uzantısı olarak ÇB-MTK ortaya çıkmıştır (McKinley ve Reckase, 1982). ÇB-MTK, iki ya da daha fazla örtük değişken ile bireyin belli bir test maddesini doğru cevaplama olasılığı arasındaki ilişkiyi açıklamaktadır (Ackerman, Gierl ve Walker, 2003). Tek boyutlu MTK bireyin tek bir yeteneğine ilişkin bir puan verirken, ÇB-MTK test sonucu olarak bir yetenek profili ortaya koymaktadır (Hartig ve Höhler, 2009).

ÇB-MTK modelleri veri setinin yapısına göre basit (maddeler arası) ve karmaşık (maddeler içi), boyutlar arası etkileşime göre telafi edici ve telafi edici olmayan modeller şeklinde sınıflandırılabilir. Çok boyutlu bir testte, her bir maddenin test ile ölçülen sadece bir boyut ile ilişkili olma durumu basit yapı olarak adlandırılmaktadır (McDonald, 1999). Karmaşık yapıda ise testteki maddeler, birden çok yetenek ile ilişkilidir. Bu doğrultuda maddeler, ölçülen birden fazla yeteneğe yük vermektedir (Sheng ve Wikle, 2007). Bir başka ifadeyle, basit yapıdaki maddeler sadece ilgili boyut için madde ayırt edicilik değerine sahipken, karmaşık yapıdaki maddenin ilişkili olduğu her bir boyut için madde ayırt edicilik değeri vardır. Örtük özellikler arasındaki ilişkilere göre telafi edici modellerde, bireyin soruda ölçülen bir



boyuttaki yetersizliği sorunun ölçtüğü diğer bir boyutta yüksek yeteneğe sahip olması ile telafi edilebilir ve bireyin soruyu doğru cevaplama olasılığı artar (Reckase, 2009). Başka bir ifadeyle, telafi edici modeller bireylerin soruyu doğru cevaplamak için çeşitli alternatif stratejilerden birini kullandığını varsaymaktadır (Embretson ve Yang, 2013). Telafi edici olmayan modeller ise sorunun doğru cevaplanması için gerekli tüm özelliklere sahip olunması gerektiğini, bir özellikteki eksikliğin bir başka özellikteki üstünlük ile telafi edilemeyeceğini belirtmektedir (de la Torre, 2009; Maris, 1999; Roussos vd., 2007). Telafi edici modellerde bireyin soruyu doğru cevaplama olasılığı soruyu cevaplamak için gereken yeteneklerin olasılıklarının toplamıyken, telafi edici olmayan modellerde çarpımından oluşmaktadır (Ackerman, 1996; Ackerman, Gierl ve Walker, 2003; Reckase, 2009).

Bu çalışmada kapsamında TIMSS 2015 uygulaması matematik testi verilerine en iyi uyumu 3PL ÇB-MTK modeli sergilediği için aşağıda bu modele yönelik detaylı bilgilere yer verilmiştir.

### **Çok Boyutluluk için Genişletilmiş Üç Parametrelilik Lojistik Model**

Bir sorunun tahminle doğru yanıtlanabilmesi durumunda, şans parametresi devreye girer ve MKE sıfırdan başlamaz. Bu doğrultuda Birnbaum (1968), 2 parametrelilik modele bir parametre daha ekleyerek, şans faktörünün testlerdeki etkisini kontrol altına almaya çalışmıştır. 3PL modelde 1PL ve 2PL modellerden farklı olarak, şans parametresi olarak adlandırılan  $c$  parametresi yer almaktadır. Bu parametre bireyin bir maddeyi doğru olarak cevaplama olasılığının en düşük olasılığına karşılık gelmektedir. Şans parametresi, özellikle şansla doğru cevabı bulmanın mümkün olduğu çoktan seçmeli testlerde önem taşımakta ve 0 ile 1 arasında bir değer almaktadır (Hambleton ve Swaminathan, 1985).

Çok boyutluluk için genişletilmiş üç parametrelilik lojistik model, tek boyutlu MTK kapsamındaki 3PL modelin bir uzantısıdır. Reckase (1985, 2009),  $j$  bireyinin  $i$  maddesine doğru cevap verme olasılığına ilişkin 3PL modelin matematiksel ifadesini aşağıdaki gibi tanımlamıştır:

$$P(U_{ij} = 1 | \theta_j, \mathbf{a}_i, c_i, d_i) = c_i + (1 - c_i) \frac{e^{\mathbf{a}_i \theta_j' + d_i}}{1 + e^{\mathbf{a}_i \theta_j' + d_i}}$$

$U_{ij}$ :  $j$  bireyinin  $i$  maddesine verdiği cevap (1 ya da 0)

$\theta_j$ :  $j$  bireyinin yetenek vektörü

$a_i$ :  $i$  maddesinin eğim vektörü

$d_i$ :  $i$  maddesinin kesim noktası

$c_i$ :  $i$  maddesinin şans parametresi

## **Bilişsel Tanı Modellerinin Temelleri**

Araştırmacı ve eğitimcilerin, öğrencilerin testteki başarılarına yönelik daha biçimlendirici bilgilere talebi giderek artarken (Mislevy, 2006; Robets ve Gierl, 2010), mevcut modellerin bireylerin belli bir konudaki güçlü ve zayıf yönleri hakkında ayrıntılı bilgi sağlayamaması eleştirilmelerine neden olmaktadır (Leighton ve Gierl, 2007; Snow ve Lohman, 1989). Bu eleştirilerin merkezinde, bireyler hakkında daha kapsamlı tanılayıcı bilgi sağlayacak ve bu bilgileri bireylerin öğretim ihtiyaçları ile ilişkilendirecek yöntemlere duyulan ihtiyaç yer almaktadır. Çünkü geleneksel MTK temelli modeller; bireylerin başarılarını ölçmek, gelecekteki performanslarını tahmin etmek, bireyleri sıralamak ve karşılaştırmak için çeşitli alanlarda başarıyla uygulansa bile bireylere tanılayıcı bilgi sağlama noktasında yetersiz kalmaktadır. Bu ihtiyaçlar doğrultusunda, bireyleri belli bir alandaki beceri ya da niteliklere sahip olmaları açısından doğru bir şekilde sınıflandırarak, test performansları hakkında tanılayıcı bilgi sağlamak üzere bilişsel tanı modelleri ortaya çıkmıştır. BTM'ler çok boyutlu kategorik-gizil özellik modelleridir (García, Olea ve de la Torre, 2014). Rupp, Templin ve Henson'a (2010) göre BTM, bireylerin bilgi yapısı ve çoklu nitelikleri ölçerek sınıflandırmaya dayalı kararlar vermek amacıyla kullanılmaktadır. BTM, geleneksel MTK modellerindeki gibi bireylerin test performanslarına ilişkin tek veya çoklu yeterlilik puanları vermekten ziyade, her bir bireyin ölçülen niteliklerden hangi/hangilerine sahip olduğunu göstererek kategorik bir profil oluşturan, çok boyutlu bir ölçme modelidir. Nitekim eğitimde BTM'lerin kullanımının temel amacı, testi alan bireylerin güçlü ve zayıf oldukları beceri/alt becerileri belirlemektir (George ve Robitzsch, 2015; Rupp ve Templin, 2008b). Böylece BTM'lerin eğitimde kullanılması ile öğrencilerin soruları cevaplarken kullandıkları bilgi, beceri, süreç ve stratejiler ile ilgili daha kapsamlı bilgiler elde edilebilir (Gierl ve Cui, 2008).

## Q-Matrisi

Tanılayıcı değerlendirme'nin temel bileşenlerinden biri maddeler tarafından ölçülen niteliklerdir. Nitelik, belli bir alandaki görevi yerine getirmek için gerekli olan işlemsel ya da açıklayıcı bilgidir (Gierl, Wang ve Zhou, 2008). Burada "nitelik" olarak adlandırılan gizil değişkenler; beceri (Francis vd., 2006), özellik (Başokçu, 2010; Başokçu ve Kelecioğlu, 2014), bileşen (Kruidenier, 2002) ve alt beceriler (Moseley, 2004) şeklinde de tanımlanabilir. Bireyin ilgili maddeye doğru cevap verebilmesi için madde ile ölçülen niteliklere sahip olması gerekmektedir. Maddeler tarafından ölçülen nitelikler ise analizlere başlamadan önce belirlenmeli ve teorik olarak gerekçelendirilmelidir (Rupp, 2007). Bir başka ifadeyle, BTM'lerin kullanılabilmesi için ilk adım, Q-matrisi olarak bilinen nitelik tablosuyla maddelerin eşleştirilmesidir (Tatsuoka, 1983). Q-matrisi, BTM'nin temel bileşenidir ve soruyu doğru cevaplamak için gerekli olan özellikleri göstermektedir. Matrisin elementi  $q$ , aşağıdaki gibi tanımlanmaktadır:

$$q_{jk} = \begin{cases} 1, & \text{eğer } j \text{ maddesini doğru cevaplamak için } k \text{ becerisi gerekiyorsa} \\ 0, & \text{gerekmiyorsa} \end{cases}$$

$j$  testte ölçülen madde sayısı ve  $k$  testte ölçülen nitelik sayısı olmak üzere,  $J \times K$  boyutlarındaki madde-nitelik matrisi,  $\mathbf{Q}=\{q_{jk}\}$ , madde ile ölçülen nitelikler arasındaki ilişkilerin varlık ve yokluk durumuna göre sırasıyla 1 ve 0 ile temsil edilmektedir. Q-matrisinde niteliklerin hangi maddeler ile ilişkili olduğu analizler öncesinde bilindiği için BTM'ler doğrulayıcı modellerdir.

BTM'ye göre öğrencilerin testteki maddelere verdikleri cevaplar ait oldukları örtük/gizil sınıfların bir vektörüdür. BTM, bireylerin maddelere verdikleri cevaplardan yola çıkıp her bir nitelikteki performanslarını belirleyerek, ait oldukları örtük sınıfların tanımlanmasını amaçlamaktadır (Leighton ve Gierl, 2007). Başka bir ifadeyle, Q-matrisi ve bireylerin maddelere verdikleri cevaplar kullanılarak, her bir birey için nitelik profili kestirilmektedir. Nitelikler arası ilişkiler için bir sınırlandırma yapılmadığında (nitelikler arasında hiyerarşik bir ilişki tanımlanmadığında),  $K$  tane nitelik için olası nitelik profili (örtük sınıf) sayısı  $2^K$  dir. Bu nitelik profilleri, test için belirlenen  $K$  niteliğin bir fonksiyonudur ve  $\alpha_i = \{\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iK}\}$ ,  $i$  bireyinin nitelik vektörünü temsil etmektedir (de la Torre ve Karelitz, 2009). Bu nitelik profili,  $K$  tane özelliğe ilişkin 1 ve 0'lardan oluşur ve bireyin o özelliğe sahip olup olmadığını göstermektedir. Nitelik profili kestirimleri genellikle, her bir özelliğe ilişkin kestirilen sonsal

olasılıklara kesme puanının uygulanması ile elde edilir (de la Torre, 2009). Aşağıda, 5 madde ve 4 nitelik için tanımlanan iki kategorili Q-matrisi ve  $i$  bireyinin yetenek profili verilmektedir. (1,1,0,0) nitelik profiline sahip bireyin, 1. ve 2. niteliklere ( $\alpha_1$  ve  $\alpha_2$ ) sahip olduğu, 3. ve 4. niteliklere ( $\alpha_3$  ve  $\alpha_4$ ) ise sahip olmadığı görülmektedir. Bu doğrultuda, aşağıdaki gibi tanımlanan Q-matrisine göre yetenek profili (1,1,0,0) olan bir bireyin 1., 2. ve 5. soruları doğru, 3. ve 4. soruları ise yanlış cevaplaması beklenmektedir.

$$Q \times \alpha_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

K sütunlu bir Q-matrisi  $2^K-1$  tane olası nitelik profiline karşılık olarak,  $2^K-1$  tane birbirinden farklı satıra sahip olacaktır (de la Torre, Hong ve Deng, 2010). Bu nedenle, bireyleri ölçülen niteliklere sahip olma durumlarına göre doğru bir şekilde sınıflayabilmek ve bireylerin bilişsel özelliklerini daha doğru bir şekilde inceleyebilmek için her bir niteliğin yeterli sayıda ölçülmesi gerekmektedir.

Q-matrisinin tasarımı, çeşitli özelliklere göre değişiklik göstermektedir. Q-matrisi, bir maddeye sadece bir niteliğin atandığı basit yapı ya da birden fazla niteliğin atandığı karmaşık yapıda (maddeler arası boyutluluk) olabilir. Q-matrisinin karmaşıklığı ise madde ile ölçülen nitelik sayısı arttıkça artmaktadır. Q-matrisinin oluşturulmasında bir diğer ayırım ise nitelikler arasında tanımlanan ilişkilere göre yapılmaktadır: nitelikler arası hiyerarşik ilişkilerin tanımlandığı ve tanımlanmadığı Q-matrisi. Hiyerarşik olmayan Q-matrisinde, niteliklerin birbirinden bağımsız olduğu varsayılmaktadır. Hiyerarşik yapıdaki Q-matrisinde ise maddeyi doğru cevaplamak için gereken nitelikler arası bir sıralama tanımlanarak, bilişsel becerilerin bir hiyerarşisi oluşturulur (Gierl, Wang ve Zhou, 2008). Nitelikler arası hiyerarşik ilişkilerin tanımlanmasında, düşük bilişsel yüklü niteliklerden yüksek bilişsel yüklü niteliklere doğru bir sıra izlenir ve ilk nitelik, hiyerarşinin en alt düzeyine yerleştirilir. Çünkü, bireyin daha karmaşık niteliklere sahip olabilmesi için daha temel niteliklerde yeterli olması beklenmektedir. Hiyerarşik yapıdaki Q-matrisleri; doğrusal, yakınsak, ıraksak ve yapılandırılmamış şeklinde sınıflandırılabilir (Gierl, Leighton ve Hunka, 2007; Leighton, Gierl ve Hunka, 2004; Rupp, Templin ve Henson, 2010). Doğrusal nitelik hiyerarşileri, tüm niteliklerin ardışık olarak birbirinin ön koşulu olmasını

gerektirmektedir. Yakınsak nitelik hiyerarşisinde ise tek bir nitelik, birden fazla farklı niteliğin ön koşulu olabilir. Aynı zamanda tek bir nitelik, önceki niteliklerden bir ya da daha fazlasında yeterli olmayı da gerektirebilir. İraksak nitelik hiyerarşisi, ortak tek bir nitelikten kaynaklanan farklı ayrı parçalara işaret ederken, yapılandırılmamış nitelik hiyerarşisinde tek bir nitelik, birden fazla niteliğin ön koşuludur ve bu nitelikler birbirleri ile doğrudan bir ilişkiye sahip değildir. Q-matrisinin oluşturulmasındaki bir diğer farklılık ise niteliklerin iki ya da çok kategorili olarak tanımlanmasından kaynaklanmaktadır. 1-0 şeklinde iki kategorili olarak tanımlanan Q-matrisinde, nitelik soruyla ölçülüyorsa 1, ölçülüyorsa 0 değerini almaktadır. Çok kategorili/düzeyle niteliklerde ise bireyin madde ile ölçülen niteliğe sahip olma derecesi söz konusudur. Çok kategorili nitelikler, her bir seviyedeki özelliğe sahip olma-olmama durumuna göre iki kategorili niteliklere dönüştürülebilir (Chen ve de la Torre, 2013). Ancak bu durum, önemli miktarda bilgi kaybına neden olacaktır (Chen ve de la Torre, 2018).

BTM'lerin tanısal gücü, genellikle, Q-matrisinin teorik ve ampirik doğruluğu tarafından belirlenmektedir (Lee ve Sawaki, 2009). Ancak birçok BTM uygulamasında, Q-matrisinin oluşturulma süreci öznel ve geçerlikle ilgili sorunları gündeme getirmektedir (de la Torre, 2008; Rupp ve Templin, 2008a, Rupp, Templin ve Henson, 2010). Henson (2004) ve de la Torre ve Douglas (2004) Q-matrisi kalitesinin, bireylerin nitelik profilinin kestirimini doğrudan etkilediğini belirtmişlerdir. Nitekim Q-matrisi, BTM analizlerinden elde edilecek tanılayıcı bilgiyi etkilemekte ve Q-matrisinin yanlış veya eksik tanımlanması madde ve birey parametrelerinin sınıflama doğruluğunun yanlış kestirilmesine neden olmaktadır. Başarılı bir BTM, büyük ölçüde Q-matrisinin sağlamlığı üzerine kuruludur ve iyi tanımlanmış bir Q-matrisi, çoklu kaynakların kanıtlarına dayanmaktadır (Jang, 2009). Bu doğrultuda Gierl ve Cui (2008, s.265), testte ölçülecek bilgi ve becerileri belirlemek için önce bilişsel modelin geliştirilmesini, daha sonra bu bilişsel becerileri ölçecek maddelerin oluşturulmasını önermektedir.

### **Bilişsel Tanı Modelleri**

Son 20 yılda, testi alan bireylerden tanılayıcı bilgi elde etmek amacıyla çok boyutlu sınıflandırma yapan modellere olan ilgi giderek artmıştır (Kış 2007 Journal of Educational Measurement Özel Sayısı). Benzer şekilde Huebner (2010), son

yıllarda BTM'nin tanısal gücünün giderek farkına varılmasının bir sonucu olarak, uygulamasının arttığını belirtmiştir. Bu talep karşısında bilişsel süreçlerin, becerilerin veya niteliklerin öğrencilerin değerlendirmelerdeki cevaplarını nasıl etkilediğine dair farklı varsayım veya teorilere dayanan, çok sayıda BTM geliştirilmiştir (Ma, Iaconangelo ve de la Torre, 2016). Sinharay ve arkadaşları (2011), geliştirilen çeşitli BTM modellerinin ortak noktalarını: a) her bir test maddesinin bir veya daha fazla niteliği gerektirmesi, (b) her bireyin her bir beceriye karşılık gelen ayrık bir gizil yetenek parametresine sahip olması ve (c) bireyin soruya doğru cevap verme olasılığının, maddenin gerektirdiği becerilerin ve bireyin gizil yetenek parametrelerinin matematiksel fonksiyonu olduğunun varsayılması şeklinde belirtmiştir. Farklı BTM modellerinin analiz sürecini Lee ve Sawaki (2009b), 1) test ile ölçülen niteliklerin belirlenmesi, 2) her bir maddeyi doğru cevaplamak için gereken niteliklerin belirlenmesi, 3) her bir bireyin test performansına dayalı olarak nitelik profillerinin kestirilmesi ve 4) bireylere ya da diğer paydaşlara tanılayıcı geri bildirim sunulması şeklinde tanımlamıştır. Ravand ve Robitzsch (2015) ise BTM analiz sürecini; 1) niteliklerin belirlenmesi, 2) maddeleri ölçülecek niteliklerle eşleştirerek Q-matrisinin oluşturulması, 3) uygun BTM modelinin seçilmesi ve 4) bireylerin nitelik profillerinin kestirilmesi şeklinde açıklamıştır.

Rupp ve arkadaşları (2010) farklı BTM'lerin; 1) birey cevaplarının (gözlenen tepkilerin) iki ya da çok kategorili olarak puanlanması, 2) Q-matrisinde gizil niteliklerin iki ya da çok kategorili olarak tanımlanması, 3) modellerin telafi edici ya da telafi edici olmayan olması, 4) modellerin bağlayıcı ya da bağlayıcı olmayan olması ve 5) modellerin tahmin, kaydırma veya her iki parametreyi ele alma biçimi ve bu parametrelere uyguladıkları kısıtlamalar gibi faktörlere göre sınıflandırılabilirliğini belirtmiştir. BTM'lerde tahmin ve kaydırma parametreleri tüm nitelikler için madde düzeyinde, tüm maddeler için nitelik düzeyinde ya da hem madde hem nitelik düzeyinde uygulanabilir (de la Torre, 2011; de la Torre ve Minchen, 2014). Bir diğer sınıflandırma ise bireyin niteliğe sahip olma derecesinin iki ya da çok kategorili olarak tanımlanması ve birey yanıtlarının doğru-yanlış (1-0) şeklinde iki ya da çok kategorili olarak sınıflandırılmasına göre yapılmaktadır.

BTM'ler temelde nitelikler arası ilişkilere göre 1) telafi edici, 2) telafi edici olmayan ve 3) genel modeller olarak sınıflandırılabilir (Ravand ve Robitzsch, 2015). Telafi edici modellerde, soruyu doğru cevaplamak için gerekli olan niteliklerden

birindeki eksiklik, bir başka niteliğe sahip olma ile telafi edilebilir. Telafi edici olmayan modeller ise sorunun doğru cevaplanabilmesi için bireyin soru ile ölçülen tüm niteliklere sahip olmasını gerektirmektedir. Aynı testte hem telafi edici hem de telafi edici olmayan ilişkilere izin veren modeller ise genel BTM modelleri olarak adlandırılır (Ravand, 2016).

BTM'lerin sınıflandırılmasında bir diğer faktör ise modelin bağlayıcı olup olmamasıdır. Telafi edici olmayan modellerde bağlayıcı yoğunlaşma kuralı kullanılır ve bireyin madde ile ölçülen tüm gizil niteliklere sahip olma-olmama durumları birbiriyle çarpılarak, 0 ya da 1 değeri elde edilir (Fay, 2018). Bu doğrultuda, bireyin soruyu doğru cevaplayabilmesi için soru ile ölçülen tüm niteliklere sahip olması gerekmektedir. Bağlayıcı olmayan modellerde ise telafi edici modeller ile benzer şekilde, sorunun doğru cevaplanması için bireyin madde ile ölçülen en az bir niteliğe sahip olması yeterlidir. Bir başka ifadeyle, belli bir nitelikten yoksun olunması, bir başka niteliğe sahip olunması ile dengelenebilir (Ma, Iaconangelo ve de la Torre, 2016). Bağlayıcı olmayan veya telafi edici modeller, sorunun doğru cevaplanması için çeşitli yolların kullanımına izin vermektedir (Maris, 1999).

Bu çalışma kapsamında, madde ve nitelikler arasındaki ilişkilere yönelik ortak varsayımları temsil eden 7 BTM kullanılmıştır. Aşağıdaki tabloda çalışmada kullanılan modeller telafi edici, telafi edici olmayan ve genel model olma durumlarına göre sınıflandırılmış ve sonrasında bu modeller açıklanmıştır.

Tablo 1

*Çalışma Kapsamında Kullanılan Bilişsel Tanı Modelleri*

Tür	Kullanılan model
Telafi Edici	<ul style="list-style-type: none"><li>▪ Deterministic Inputs, Noisy "Or" Gate (DINO) modeli</li><li>▪ Additive Cognitive Diagnostic Model (A-CDM)</li><li>▪ Linear Logistic Model (LLM) ya da Compensatory Reparameterized Unified Model (C-RUM)</li></ul>
Telafi Edici Olmayan	<ul style="list-style-type: none"><li>▪ Deterministic Inputs, Noisy "and" Gate (DINA) modeli</li><li>▪ Reduced Reparameterized Unified Model (R-RUM)</li><li>▪ Higher Order DINA (HO-DINA)</li></ul>
Genel	<ul style="list-style-type: none"><li>▪ Generalized Deterministic Inputs, Noisy "and" Gate (G-DINA) modeli</li></ul>

## DINA Modeli

DINA modeli en basit BTM'lerden biri olmakta birlikte, üzerinde yapılan çeşitli değişiklikler ile çeşitli BTM modellerinin temelini oluşturmaktadır (de la Torre, Hong ve Deng, 2010). İstatistiksel olarak bağlayıcı yoğunlaşma fonksiyonu ile temsil edilen, telafi edici olmayan bir modeldir (Maris, 1995, 1999). Bireyin soruyu doğru cevaplayabilmesi için soruda ölçülen niteliklerin hepsine sahip olmasını gerektirmektedir (Henson, Templin ve Willse, 2009). DINA modeline göre  $i$  bireyinin  $j$  maddesini doğru cevaplama olasılığı aşağıdaki gibi tanımlanmaktadır:

$$\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}$$

Formülde verilen  $i$  bireyi,  $j$  maddeyi,  $k$  niteliği,  $\alpha_i$   $i$  bireyin  $k$ . nitelik kategorisini,  $q_{jk}$   $j$ . madde ve  $k$ . nitelik için Q-matrisi girdisini ve  $\eta_{ij}$   $i$  bireyinin  $j$  maddesine vermesi beklenen yanıtı temsil etmektedir. Bireyin sahip olduğu nitelikler ve madde özellikleri arasındaki etkileşim, ideal cevap olarak bilinen ( $\eta_{ij}$ ), gizil cevap değişkenini tanımlamaktadır (Tatsuoka, 1995).  $\eta_{ij}$ 'nin 1 değerini alması  $i$  bireyinin sorunun doğru cevaplanması için gerekli tüm niteliklere sahip olduğunu, 0 değerini alması ise bireyin niteliklerden en az birine sahip olmadığını göstermektedir. Bu doğrultuda DINA modeli bireyleri, gerekli niteliklerin hepsine sahip olan ve en az birine sahip olmayanlar şeklinde iki gruba ayırmaktadır (Henson, Templin ve Willse, 2009). Bir başka ifadeyle bu iki grup, madde ile ölçülen niteliğe sahip olmayanların oluşturduğu yokluk sınıfı ve tüm niteliklere sahip olma durumunu ifade eden tam sınıf şeklindedir. Maddeyi doğru cevaplamak için gereken niteliklerden herhangi birine sahip olmayan birey ile bu niteliklerden hiçbirine sahip olmayan bireyler yokluk sınıfına atanırlar.

DINA modeli, örtük özellik ve ilgili gözlenen özellik arasındaki ilişkiyi olasılık temeli olarak modellemekte ve her madde için tahmin ( $g_j$ ) ve kaydırma ( $s_j$ ) madde parametrelerini üretmektedir.  $j$  maddesinin tahmin ve kaydırma olasılıkları aşağıdaki gibi tanımlanmaktadır:

$$g_j = P(X_{ij} = 1 | \eta_{ij} = 0)$$

$$s_j = P(X_{ij} = 0 | \eta_{ij} = 1)$$

ve

$$1 - s_j = 1 - P(X_{ij} = 0 | \eta_{ij} = 1) = P(X_{ij} = 1 | \eta_{ij} = 1)$$



$g_j$  parametresi, bireyin  $j$  maddesini doğru yanıtlayabilmek için gerekli nitelik/niteliklere sahip olmamasına rağmen soruyu doğru cevaplama (yanlış pozitif olasılık) olasılığını göstermektedir. Dolayısıyla tahmin parametresi, MTK'daki şans parametresinden farklı bir yapıdadır (de la Torre ve Douglas, 2004).  $g_j$  parametresinin değeri arttıkça, maddeyi doğru yanıtlamak için gerekli niteliğe sahip olmayan bireylerin soruyu doğru cevaplama olasılığı da artmaktadır.  $s_j$  parametresi ise bireyin  $j$  maddesini doğru yanıtlayabilmek için gerekli tüm nitelik/niteliklere sahip olmasına rağmen soruyu yanlış cevaplama olasılığıdır (yanlış negatif olasılık).  $s_j$  parametresinin değeri düştükçe, ilgili niteliklere sahip bireylerin soruyu doğru cevaplama olasılığı artmaktadır. DINA modelinin madde tepki fonksiyonu aşağıdaki gibidir:

$$P(X_{ij} = 1 | \eta_{ij}) = (1 - s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}}$$

DINA modeli çok basit ve sınırlandırıcı bir model olduğu için (de la Torre, 2011; Henson ve Douglas, 2005) uygulama alanı sınırlıdır. Ancak, özellikle bireyin madde ile ölçülen tüm niteliklere sahip olmasının önemli olduğu durumlarda kullanılabilir (de la Torre ve Minchen, 2014).

### **DINO Modeli**

DINO modeli (Templin ve Henson, 2006), bireyin bir nitelikteki eksikliğinin maddenin ilişkili olduğu diğer bir nitelikteki yetkinliği ile telafi edilebileceğini varsaydığı için bağlayıcı olmayan ve telafi edici bir BTM modelidir. DINA modeli ile benzer şekilde, DINO modeli testi alan bireyleri iki sınıfa ayırmaktadır. Bu sınıflama, madde ile ölçülen niteliklerden en az birine sahip olanlar ( $v_{ij}=1$ ) ve gerekli niteliklerden hiçbirine sahip olmayanlar ( $v_{ij}=0$ ) şeklindedir. Bireyin maddenin gerektirdiği niteliklerden sadece birine sahip olması bile maddeyi doğru cevaplama olasılığının yüksek olması için yeterlidir (Templin ve Henson, 2006). DINO modeli için ideal cevap değişkeni aşağıdaki gibi tanımlanmaktadır.

$$v_{ij} = 1 - \prod_{k=1}^K (1 - \alpha_{ik})^{q_{jk}}$$

DINA modelinde olduğu gibi DINO modeli, her bir madde için nitelikler arasında eşit olmak üzere bir tahmin ve bir kaydırma parametresi kestirmektedir.  $j$  maddesinin tahmin ve kaydırma olasılıkları aşağıdaki gibi hesaplanmaktadır.

$$g_j = P(X_{ij} = 1 | v_{ij} = 0)$$

$$s_j = P(X_{ij} = 0 | v_{ij} = 1)$$

DINO modelinin madde tepki fonksiyonu ise aşağıdaki gibidir:

$$P(X_{ij} = 1 | v_{ij}) = (1 - s_j)^{v_{ij}} g_j^{1-v_{ij}}$$

DINO modeli özellikle çoklu strateji problemlerini modellemek (de la Torre ve Minchen, 2014) veya birden fazla nitelik arasından sadece bir niteliğe sahip olmanın soruyu doğru cevaplamak için yeterli olduğu durumlarda (Rupp vd., 2010) kullanılabilir. Q-matrisi, tüm maddeler tek bir niteliği ölçecek şekilde oluşturulduğunda ise DINA ve DINO modelleri arasındaki fark ortadan kalkacaktır.

### G-DINA Modeli

de la Torre (2011) tarafından geliştirilen model, DINA modelinin genelleştirilmiş halidir. DINA modeli kapsamında, madde ile ölçülen bir ya da daha fazla niteliğe sahip olmayan bireyler, sahip oldukları nitelik sayısına bakılmaksızın, eşit başarı olasılığına sahiptir (de la Torre ve Minchen, 2014). G-DINA modeli ile DINA modelinin bu güçlü sınırlandırması ortadan kalkmaktadır.  $K_j^* = \sum_{k=1}^K q_{jk}$ ,  $j$ . maddeyi doğru cevaplamak için gereken nitelik sayısı olmak üzere, G-DINA modeli bireyleri  $2^{K_j^*}$  gizil gruba ayırır ve her bir grubun kendine ait bir başarı olasılığı vardır. Ancak  $K_j^* = 1$  olması durumunda, olası grup sayısı 2 olduğu için G-DINA ve DINA modelleri aynı sonucu vermektedir.

$j$  maddesi için gerekli nitelikler  $l=1, \dots, 2^{K_j^*}$  olmak üzere  $\alpha_{lj}^* = (\alpha_{l1}, \dots, \alpha_{lK_j^*})'$  indirgenmiş vektörü ile temsil edilebilir (Chen ve Chen, 2016). G-DINA modeli için madde tepki fonksiyonu ise aşağıdaki gibi tanımlanmaktadır.

$$P(\alpha_{lj}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{k'-1} \delta_{jkk'} \alpha_{lk} \alpha_{lk'} \dots + \delta_{j12\dots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}$$

Yukarıdaki formülde verilen  $\delta_{j0}$  (keşism), madde ile ölçülen niteliklerin hiçbirine sahip olmayan bireyin soruyu doğru cevaplama olasılığını;  $\delta_{jk}$  (temel etki), tek başına  $\alpha_k$  niteliğine sahip olmanın sorunun doğru cevaplanma olasılığı üzerindeki etkisini;  $\delta_{jkk'}$  ( $\alpha_k$  ve  $\alpha_{k'}$ 'nin çoklu etkileşim etkisi),  $\alpha_k$  ve  $\alpha_{k'}$  niteliklerine sahip olmanın sorunun doğru cevaplama olasılığındaki değişime katkısını ve  $\delta_{j12\dots K_j^*}$

( $\alpha_1, \dots, \alpha_{K_j^*}$ 'nin etkileşim etkisi), madde ile ölçülen tüm niteliklere sahip olmanın sorunun doğru cevaplanma olasılığındaki değişime katkısını temsil etmektedir (de la Torre, 2011). G-DINA modeli olası tüm etkileşim etkilerini içermekte ve bu etkileşimler, maddenin gerektirdiği niteliklerden daha fazlasına sahip bireyin daha azına sahip bireye göre daha düşük doğru cevaplama olasılığına sahip olmasını mümkün kılmaktadır (de la Torre ve Minchen, 2014).

### **A-CDM Modeli**

G-DINA modelindeki tüm etkileşim etkilerinin 0'a sabitlenmesi ile telafi edici BTM modellerinden A-CDM modeli (de la Torre, 2011) elde edilmektedir. A-CDM modeli sadece her bir niteliğin keşişim ve temel etkilerini içermektedir. A-CDM modelinde,  $j$  maddesi için  $K_j^* + 1$  (maddeyi doğru cevaplamak için gerekli nitelik sayısından bir fazla) parametre üretildiği için eklemeli bir modeldir (Sorrel, de la Torre, Abad ve Olea, 2017). A-CDM'nin madde tepki fonksiyonu aşağıdaki gibi ifade edilmektedir.

$$P(\alpha_{ij}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk}$$

Bireyin madde ile ölçülen  $\alpha_{lk}$  niteliğine sahip olması, diğer niteliklerin katkısından bağımsız olarak,  $j$  maddesinin doğru cevaplanma olasılığını arttırmaktadır.

### **R-RUM Modeli**

R-RUM, NIDA modelinin geliştirilmiş (Junker ve Sijtsma, 2001) ve yetenek parametresinin kaldırılmasıyla RUM modelinin (fusion model) daha basit halidir. Yetenek parametresinin kaldırılması ile bireyin soruya verdiği cevapları etkileyebilecek ek bir yetenek bulunmadığı varsayılmaktadır (Lin ve Chang, 2018). R-RUM, farklı nitelik profilleri için maddenin doğru cevaplanma olasılığını modellemedeki esnekliği nedeniyle araştırmacılar tarafından sıklıkla kullanılmaktadır (van der Ark, Bolt, Wang, Douglas ve Wiberg, 2014, s.271). A-CDM ile benzer şekilde, sadece keşişim ve ana etkileri içermekte ve  $j$  maddesi için  $K_j^* + 1$  parametre kestirmektedir. A-CDM modeli ile arasındaki fark, kullanılan link fonksiyonundan kaynaklanmakta: A-CDM modeli identity link fonksiyonunu, R-RUM modeli ise log fonksiyonunu kullanılmaktadır. Böylece parametrelerin

yorumlanması, modeller arasında farklılık göstermektedir. Ayrıca DINA modelinden farklı olarak R-RUM modelinde, tahmin ve kaydırma olasılıkları maddeler arasında değişiklik göstermektedir (Tu, Wang, Cai, Douglas ve Chang, 2018). Bireyin madde tepki fonksiyonu ise aşağıdaki eşitlikteki gibi tanımlanmaktadır.

$$\log[ P(\alpha_{ij}^*) ] = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk}$$

veya

$$P(X_{ij} = 1 | \alpha_i) = \pi_j^* \prod_{k=1}^K r_{jk}^{* q_{jk}(1-\alpha_{ik})}$$

Eşitlikte verilen  $\pi_j^*$ ,  $j$  maddesi ile ölçülen tüm özelliklere sahip bireyin  $j$  maddesini doğru cevaplama olasılığını;  $r_{jk}^*$  parametresi ise  $j$  maddesi için gerekli olan  $k$ . niteliğe sahip olmama durumu için ceza parametresini temsil etmekte ve 0 ile 1 arasında bir değer almaktadır.

### LLM Modeli

LLM modeli, genel tanılayıcı modelin (GDM; von Davier, 2005) özel bir halidir ve telafi edici RUM (C-RUM; Compensatory Reparameterized Unified Model) modeli ile eşdeğerdir (Hartz, 2002). A-CDM ve R-RUM modellerinde olduğu gibi bireyin madde ile ölçülen her bir niteliğe sahip olması, soruyu doğru cevaplama olasılığını arttırmakta ancak bir nitelikte başarılı olunması, diğer nitelikleri etkilememektedir (Ma, Iaconangelo ve de la Torre, 2016). LLM modelinde sorunun cevaplanması, madde ya da nitelikler arasında herhangi bir eşitlik sınırlaması olmadan madde\*nitelik düzeyinde modellendiği için parametre kestirimleri, aynı nitelikleri ölçen maddeler arasında farklılık göstermektedir (Yi, 2017). Tüm etkileşim etkilerinin 0 olduğu logit link G-DINA modeli olan LLM için madde tepki fonksiyonu aşağıdaki gibi tanımlanmaktadır.

$$\text{logit}[ P(\alpha_{ij}^*) ] = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk}$$

veya

$$P(\alpha_{ij}^*) = \frac{\exp(\delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk})}{1 + \exp(\delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk})}$$

## HO-DINA Modeli

de la Torre ve Douglas (2004) HO-DINA modelini, DINA modelinin bir uzantısı olarak önermişlerdir. HO-DINA telafi edici olmayan bir modeldir ve sorunun doğru cevaplanması için tüm alt niteliklere sahip olunmasını gerektirmektedir. HO-DINA modeli, 1) test ile ölçülen alt niteliklere sahip olmanın, üst düzey ve daha kapsamlı bir yetenek parametresi ile ilişkili olduğunu ve 2) alt nitelikler arasında bir ilişki olmadığını varsaymaktadır (de La Torre ve Douglas, 2004). Nitelikler,  $\theta$  olarak adlandırılan, sürekli bir üst düzey faktörü paylaşmakta ve alt düzeyin BTM ve üst düzeyin MTK modeli olduğu, iki düzeyli bir model kullanmaktadır (Liu, Huggins-Manley ve Bulut, 2017). Bireyin soruyu doğru cevaplama olasılığı aşağıdaki gibi hesaplanmaktadır.

$$P(\alpha_{ek}|\theta_e) = \frac{\exp(\tau_k \theta_e - \beta_k)}{1 + \exp(\tau_k \theta_e - \beta_k)}$$

Yukarıdaki eşitlikte verilen  $e$  bireyi,  $\tau_k$  ve  $\beta_k$  ise sırasıyla  $k$  niteliğinin ayırt edicilik ve güçlük parametresini temsil etmektedir.

## Bilişsel Tanı ve Çok Boyutlu Madde Tepki Modellerinin Karşılaştırılması

Literatürde bilişsel tanı modelleri için çeşitli isimler kullanılmakta ve bunlardan biri de yapılandırılmış MTK modelleridir (Rupp ve Mislevy, 2007). Yapılandırılmış MTK modeli terimi, bu modelleri örtük/gizil değişken modelleri familyası, özellikle, MTK modelleri ile ilişkilendirmektedir (Rupp ve Templin, 2008b). Çok boyutlu örtük sınıf modelleri olarak BTM, ÇB-MTK modelleri ile çeşitli özellikleri paylaşmaktadır. Bütün MTK temelli örtük sınıf modellerin arkasında, “bireyin soruyu cevaplamak için gerekli olan becerilerdeki yeterliliği ne kadar yüksek olursa, soruyu doğru cevaplama olasılığı o kadar yüksektir” genel düşüncesi yatmaktadır (Roussos, Templin ve Henson, 2007). Her iki model de birden fazla gizil yordayıcı değişken içerir ve gizil değişkenlerin etkileşimine göre telafi edici ve telafi edici olmayan modeller şeklinde iki kategoriye ayrılabilir. Telafi edici modelde, birey soruda ölçülen bir özellikte yetersiz olsa bile bireyin soruda ölçülen diğer bir özellikte yüksek yeteneğe sahip olması, soruyu doğru cevaplama olasılığını arttırmaktadır (Reckase, 2009, s.87). Telafi edici olmayan modeller ise sorunun doğru cevaplanması için gerekli tüm özelliklere sahip olunması gerektiğini, bir özellikteki eksikliğin bir başka özellik ile

telafi edilemeyeceğini varsaymaktadır (de la Torre, 2009; Maris, 1999; Roussos vd., 2007). Başka bir ifadeyle, telafi edici modeller genellikle, bireylerin problemi çözmek için çeşitli alternatif stratejilerden birini kullandıklarını kabul etmektedir (Embretson ve Yang, 2013). Gözlenen tepki değişkenlerinin doğasına göre de BTM, tetrakorik veya polikorik korelasyon matrislerini kullanan MTK modellerine benzemektedir. Ancak genel olarak, ÇB-MTK modelleri basit yapıya sahipken BTM daha karmaşık yapıya sahiptir.

Bilişsel tanı modellerinin, MTK modelleri ile ortak bir başka önemli özelliği ise modellerin doğrulayıcı yapılarıdır (Jurich, 2014). BTM doğrulayıcı bir model olduğundan, soruyu doğru cevaplamak için gereken özelliklerin analizler yapılmadan önce belirlenmesi ve gerekçelendirilmesi gerekmektedir (Rupp, 2007). BTM sadece kaç tane örtük sınıfın mevcut olması gerektiği konusunda değil, aynı zamanda, maddenin ölçtüğü becerilerin kombinasyonu ve bireyin hakim olduğu becerilere dayalı olarak madde yanıtlarını tahmin etmesi açısından da doğrulayıcı bir yapı sergilemektedir (Templin ve Hoffman, 2013).

ÇB-MTK ve BTM arasındaki temel farklılık, ÇB-MTK modelleri bireyin yetenek kestirimini sürekli bir ölçekte yaparken BTM'nin bireyleri, testte ölçülen özelliklerden hangilerine sahip olup hangilerine sahip olmadıklarını gösteren örtük sınıflara ayırmasından kaynaklanmaktadır (Cui vd., 2012). BTM, öğrenme ve öğretime yardımcı olmak için çok boyutlu tanılayıcı dönüt sağlamaktadır. ÇB-MTK modellerinden tanılayıcı bilgi edinmek mümkün olsada temelde farklı amaçlar için kullanılmaktadır. Ayrıca, BTM'de ölçülen özellikler daha kapsamlı bir şekilde belirlenmektedir. Her bir özellik ve onunla ilişkili madde, öğretim açısından anlamlı ve ölçülebilir olmalıdır. Bu farklılıklara ek olarak, BTM birden fazla ölçüt referanslı yorumlara olanak sağlarken, ÇB-MTK modelleri birden fazla norm referanslı yorumlamaya olanak sağlamaktadır (Rupp ve Templin, 2008b).

### **Bilişsel Tanı Modellerinin Uyarlanması (Retrofitting)**

Eğitimde kullanılan testlere BTM'nin başarılı bir şekilde uygulanması, BTM'ye rehberlik eden bilişsel modellerin belirli gereksinimleri olduğu için kısmen, yeni test geliştirme prosedür ve uygulamaları gerektirmektedir (Gierl ve Cui, 2008). Rupp ve Templin (2008b) BTM'nin kendine özgü özelliklerini 9 başlık altında toplamıştır. Bunlar: 1) çok boyutlu yapıları, 2) doğrulayıcı yapıları, 3) yük yapısının karmaşıklığı,

4) uygun oldukları gözlenen tepki değişkeni türleri, 5) gizil yordayıcı değişken türleri, 6) gizil yordayıcı değişkenlerin etkileşimlerinin doğası, 7) ölçüte dayalı yorumlamalar, 8) yorumların tanılayıcı yapısı ve 9) modelleyebilecekleri hetorejenlik türleri şeklindedir. BTM'nin bu gereksinimlerden dolayı, bireylerin ölçülen niteliklerden hangi/hangilerine sahip olduğuna dair daha detaylı tanılayıcı bilgi edinmenin en iyi yolu, tanılayıcı amaçla bir test tasarlamaktır. Bu nedenle, eğitimde BTM'nin başarılı bir şekilde uygulanması, yukarıda belirtilen gereksinimleri karşılayabilmek adına yeni test geliştirme süreç ve uygulamalarını gerektirecek ve bu da oldukça maliyetli olacaktır (Gierl ve Cui, 2008). Tanılayıcı test geliştirmenin maliyetinin yüksek olması nedeniyle, mevcut madde yanıtlarından tanılayıcı dönüt elde edebilmek için BTM mevcut testlere uyarlanmaktadır (örneğin, Jang vd., 2013; Kim, 2014; Li, Hunter ve Lei, 2015). Uyarlama süreci, geleneksel test geliştirme yöntem ve uygulamalarını kullanan mevcut bir test sistemindeki öğrenci yanıtlarına, yeni bir istatistiksel veya psikometrik modelin uygulanması şeklinde düşünülebilir (Gierl, Roberts, Alves ve Gotzmann, 2009). Uyarlama, testin oluşumunda kullanılan çerçeveye bakılmaksızın, BTM'nin tanılayıcı bilgiler elde etmek için kullanılabileceğini varsaymaktadır (de la Torre ve Karelitz, 2009). Nitekim Lee ve Sawaki (2009, s.174), “.. yeni bir bilişsel tanı testi tasarlamamanın pahalı ve zaman alıcı sürecine başlamadan önce, mevcut değerlendirmelerden ne derece kullanışlı bilişsel tanı bilgisinin elde edileceğinin incelenmeye değer” olduğunu belirtmiştir. Literatürde BTM çalışmaları ise iki şekilde yapılmaktadır: 1) mevcut bilişsel olmayan testlerden daha zengin bilgi elde etmek üzere uyarlama çalışmalarının yapılması ve 2) tanılayıcı bilişsel testlerin hazırlanması. Ancak eğitimsel değerlendirmelerde BTM çalışmalarının birçoğu uyarlama çalışmasıdır (Ravand, 2016).

### **İlgili Araştırmalar**

Bu bölümde, çalışma konusuyla ilgili araştırmalar iki başlık altında ele alınmıştır. İlk başlıkta, bilişsel tanı modellerinin mevcut testlere uyarlanması, çeşitli koşullar altında BTM'lerin model uyumu ve doğru sınıflama oranlarının karşılaştırılması ile ilgili alan yazındaki araştırmalar incelenmiştir. İkinci alt başlıkta ise çalışmanın temel odak noktası olan MTK ve BTM'nin birbirlerinin yerine kullanımı ve karşılaştırılması ile ilgili araştırmalar yer almaktadır.

## Bilişsel Tanı Modellerinin Uyarlanması ile İlgili Araştırmalar

Kunina-Habenicht, Rupp ve Wilhelm (2009), çok boyutlu yetenek profillerinin kestiriminde doğrulayıcı faktör analizi (DFA) ve bilişsel tanı modellerini karşılaştırmıştır. Üç bölümden oluşan çalışmanın ilk kısmında araştırmacılar tarafından geliştirilen tanılayıcı teste ilişkin teorik çerçeve, ikinci kısmında testin veri yapısı, uygun DFA ve BTM'nin seçilmesi ve üçüncü kısımda analiz sonuçları sunulmuştur. Araştırmacılar tarafından geliştirilen tanılayıcı matematik değerlendirme testi (diagnostic mathematics assessment: DMA), 3. ve 4. sınıf öğrencilerinin temel aritmetik ve modelleme becerilerine yönelik bir profil sağlamayı amaçlamaktadır. 52'si hesaplama ve 35'i bağlam içinde aritmetik becerilerinin kullanılmasını gerektiren toplam 87 maddeden oluşan test, 464 ilkokul 3. ve 4. sınıf öğrencisine uygulanmıştır. Öğrencilerin cevapları doğru ve yanlış şeklinde, iki kategorili olarak puanlanmıştır. Çalışmada katılımcı sayısının sınırlı olması nedeniyle, model parametreleri açısından daha basit bir model olan GDM kullanılmıştır. KTK'ya göre hesaplanan madde güçlük parametresi 0.90'dan büyük ve 0.10'dan küçük, madde ayırt edicilik parametresi 0.20'nin altında olan ve açımlayıcı faktör analizi sonuçlarına göre düşük faktör yüküne sahip maddeler, zayıf madde uyumundan kaynaklanacak problemleri önlemek amacıyla, testten çıkarılmıştır. DFA kapsamında kurulan 9 modelin model veri uyumu test edilerek, en iyi modelin tüm boyutlar arasında korelasyon olan model olduğu görülmüş ve GDM-DFA karşılaştırmalarında bu model kullanılmıştır. GDM'ye göre yapılan analizler sonucu maddelerin %2'sinin iyi, %50'sinin orta ve %48'inin zayıf uyum gösterdiği bulunmuştur. DFA ve GDM ile kestirilen birey profillerinin ise benzer olduğu görülmüştür. Örtük sınıfların dağılımları incelendiğinde, her iki sınıf düzeyinde, hiçbir niteliğe sahip olmayan ve tüm niteliklere sahip olan birey profilleri en yüksek oranda gözlenmiştir. DFA kapsamında kestirilen faktör puanları ve GDM birey parametreleri arasındaki ilişkilerin ise her bir boyut için 0.83 ile 0.89 arasında değiştiği bulunmuştur.

Chen ve Chen (2016), PISA 2006 uygulaması okuma becerileri testini G-DINA modeli çerçevesinde uyarlamayı amaçlamıştır. Bu amaçla, 9. kitapçıkta yayınlanan 26 maddeyi ve 2238 katılımcının cevaplarını kullanmışlardır. Kısmi olarak puanlanan soruları, tam doğru cevap 1 ve diğer cevaplar 0 olacak şekilde iki kategorili hale getirmişlerdir. Q-matrisini madde içerik analizlerine göre oluşturmuş



ve doygun (saturated) ve indirgenmiş (reduced) BTM'leri kullanarak, doğrulamasını yapmışlardır. Bu doğrultuda, öncelikle PISA okuma becerisi testinde yer alan becerileri kullanarak Q1-matrisini oluşturmuşlardır. Ardından bu becerileri ve alan yazınında yer alan okuma becerisi taksonomisindeki becerileri kullanarak, 5 alan uzmanından soruları doğru cevaplamak için gereken becerileri kodlamalarını istemişlerdir (Q2-matrisi). Madde içeriklerinin analiz edilmesiyle son hali verilen Q3-matrisinin veriye kabul edilebilir düzeyde uyum sağladığı bulunmuştur. Örtük sınıfların dağılımı incelendiğinde, tüm özelliklere sahip katılımcıların çoğunlukta olduğu görülmüştür. Sorular hakkında daha fazla bilgi sahibi olmak için G-DINA modeline dayalı olarak oluşturulan Q-matrisini kullanarak DINA, DINO, A-CDM, LLM ve R-RUM modellerini test etmişlerdir. LLM'in en iyi model uyumunu göstermesine rağmen testteki soruların hepsinin uyum gösterdiği bir modelin olmadığını tespit etmişlerdir. Benzer şekilde madde uyum istatistikleri de LLM'nin en fazla maddenin uyum sağladığı model olduğunu göstermiştir. Birden fazla özelliği ölçen maddelerin hiçbiri DINA ve DINO modeline uyum göstermemiştir. Sonuç olarak, Q-matrisinin doğru oluşturulması ve uygun BTM'nin belirlenmesi koşulunda, uyarlamanın BTM'ler için alternatif bir yöntem olabileceğini belirtmişlerdir.

Ma ve de la Torre (2016) çalışmalarında, madde kategorilerinin sıralı olduğu çok kategorili veriler için ardışık G-DINA modelini geliştirmiş ve önerilen modelin uygulanabilirliğini incelemek üzere iki simülasyon çalışması yapmıştır. Birinci simülasyon çalışmasında, örneklem büyüklüğü (500, 1000, 2000 ve 4000) ve madde kalitesi (düşük, orta ve yüksek) değişkenlerinin etkisi incelenmiştir. Çalışmanın verileri sınırlandırılmış ardışık G-DINA (RS-GDINA) modeline göre üretilmiştir. RS-GDINA ve sınırlandırılmamış ardışık G-DINA (US-GDINA) modelleri ile kestirilen madde parametrelerine ilişkin RMSE değerleri ve birey parametrelerine ilişkin nitelik vektörü doğru sınıflama oranları (PCV) karşılaştırılmıştır. Kestirilen RMSE değerlerinin 0.012 ile 0.090 arasında değiştiği ve en yüksek değerini örneklem büyüklüğünün 500 ve test kalitesinin düşük olduğu koşullarda elde ettiği görülmüştür. RS-GDINA modelinin tüm koşullar için US-GDINA modelinden daha küçük RMSE değerine sahip olduğu ve iki model arasındaki RMSE değerleri arasındaki farkın, örneklem büyüklüğü ve madde kalitesi düştükçe arttığı gözlenmiştir. Ardışık G-DINA ve G-DINA modellerine göre incelenen PCV değerleri karşılaştırıldığında, ardışık G-DINA modellerinin daha yüksek doğru sınıflama

oranına sahip olduđu, tüm modeller için örneklem büyüklüğü ve madde kalitesinin artması ile PCV değerlerinin arttığı tespit edilmiştir. İkinci simülasyon çalışmasında ise LRT, AIC ve BIC değerlerinin doğru modeli seçme oranları karşılaştırılmış ve tüm simülasyon koşulları için AIC'a göre seçilen modelin daha yüksek PCV değerine sahip olduğu bulunmuştur. Çalışmanın gerçek veriye dayalı analizlerinde ise TIMSS 2007 uygulaması 4. sınıf matematik testinde yer alan 6'sı çok kategorili ve 5'i iki kategorili olmak üzere 11 maddeye verilen cevaplar kullanılmıştır. İki kategorili maddeler için G-DINA ve ardışık G-DINA modellerinin aynı sonuçları verdiğini, ancak çok kategorili veriler için kategoriler ile nitelikler arasındaki ilişkilerin belirlenmesi ile daha tanılayıcı bilgilerin elde edilebileceğini ve bireylerin daha doğru sınıflanabileceğini belirtmişlerdir.

Gao, Miller ve Liu (2017), Q-matrisinin yanlış tanımlanmasının ve yanlış bilişsel tanı modeli kullanımının sınıflama doğruluğu üzerine etkisini incelemiştir. Çalışmanın verileri G-DINA modeline göre üretilmiştir. Simülasyon çalışması olarak yürütülen çalışmanın koşulları; katılımcı sayısı (500, 1000 ve 5000), nitelikler arası korelasyon (0.4 ve 0.8) ve madde sayısı (14 ve 28) şeklindedir. Nitelik sayısı tüm koşullar için 4'e sabitlenmiş ve her bir koşul için 1000 tekrar yapılmıştır. Üretilen veri setleri G-DINA, A-CDM ve DINA modelleri ile analiz edilmiştir. Doğru Q-matrisindeki 1 değerlerinin 0 (altında), 0 değerlerinin 1 (üstünde) ve 1 ve 0 değerlerinin yer değiştirmesi (dengeli) gibi çeşitli şekillerde oluşturulan yanlış Q-matrislerinin etkisinin incelendiği çalışmada, biri doğru ve beşi yanlış tanımlanan toplam 6 Q-matrisi kullanılmıştır. Sınıflama doğruluğu, tüm birey ve replikasyonlar için genel doğru sınıflama oranları (OCA) ve her bir örtük sınıf için doğru sınıflama oranları (CCA) bazında incelenmiştir. OCA sonuçları, tüm koşullar için test uzunluğu arttıkça genel doğru sınıflama oranının da arttığını, örneklem büyüklüğünün azalmasıyla ise doğru sınıflama oranının biraz düştüğünü göstermiştir. Nitelikler arasındaki korelasyonların artması ise birkaç istisnai durum dışında, doğru sınıflama oranını biraz arttırmıştır. Kestirim yapılan modellerin etkisi incelendiğinde, kullanılan Q-matrisinden bağımsız olarak, tüm koşullar için G-DINA modelinin genel doğru sınıflama oranının daha yüksek olduğu görülmüştür. Bu durum, verilerin G-DINA modeline göre üretilmesinden kaynaklanmaktadır. Diğer iki model karşılaştırıldığında ise A-CDM'nin DINA modeline göre daha yüksek doğru sınıflama oranına sahip olduğu bulunmuştur. Yanlış belirlenen Q-matrisinin sınıflama

doğruluğu üzerindeki etkisini incelemek üzere modellerin doğru ve yanlış tanımlanan Q-matrisi kestirimleri karşılaştırılmıştır. G-DINA ve A-CDM'nin yanlış tanımlanan Q-matrisleri için benzer sonuçlar ürettiği, DINA modelinin genel doğru sınıflama oranlarının daha düşük olduğu ve Q-matrisinin yanlış tanımlanma oranı arttıkça, tüm modeller için doğru sınıflama oranının düştüğü gözlenmiştir. Her bir örtük sınıf düzeyindeki doğru sınıflama oranları incelendiğinde ise daha fazla niteliğin olduğu örtük sınıfların daha yüksek doğru sınıflama oranına sahip olduğu görülmektedir. (1,1,1,1) örtük sınıfı için doğru sınıflanma oranlarının, kullanılan BTM ve Q-matrisinden bağımsız olarak, oldukça yüksek olduğu bulunmuştur. Özellikle DINA modelinde, (1,1,1,1) örtük sınıfı için yanlış sınıflanan birey olmamış, doğru sınıflama oranı tüm koşullar için 1 olarak tespit edilmiştir. Genel olarak, G-DINA modelinin neredeyse tüm örtük sınıflar için en yüksek doğru sınıflama oranına sahip olduğu gözlenmiştir.

Liu, Huggins-Manley ve Bradshaw (2017) çalışmalarında, Q-matrisi tasarımı için 3 yaklaşım sunup, bu tasarımların farklı nitelik yapıları altında sınıflandırma doğruluğu üzerine etkisini incelemeyi amaçlamıştır. Çalışmada kullanılan hiyerarşik yapılar doğrusal, ıraksak, yakınsak ve yapılandırılmamış; kullanılan Q-matrisi tasarımları ise bağımsız (maddeler nitelikleri birbirinden izole bir şekilde ölçmektedir), bitişik (bitişik nitelikler ile birlikte) ve ulaşılabilir (ulaşılabilir nitelikler ile birlikte) şeklindedir. Nitelik sayısının 5'e, nitelikler arası korelasyonun 0.70'e ve örneklem büyüklüğünün 2000'e sabitlendiği çalışmada madde niteliği (düşük ve yüksek) olacak şekilde düzenlenmiş ve iki çeşit simülasyon koşul kümesi kullanılmıştır. Birinci simülasyon çalışmasında, test uzunluğu 30 olarak alınmış ve her bir nitelik en az 6 kere ölçülmüştür. Test uzunluğunun sabit olması durumunda, farklı nitelik hiyerarşileri altında Q-matrisi tasarımındaki 3 farklı yaklaşımın sınıflama sonuçları üzerindeki etkisi incelenmiştir. İkinci simülasyon çalışmasında ise her niteliğin ölçülme sayısı 6'ya sabitlenmiştir. Bireylerin nitelik profilleri MAP'a göre kestirilmiş ve her bir koşul 1000 tekrar ile yapılmıştır. Çalışmanın verileri ise HDCM'ye göre üretilmiştir. Yapılan simülasyon çalışmalarının sınıflama doğruluğu ve güvenilirliği, nitelik (ACA) ve nitelik profili (PCA) bazında değerlendirilmiştir. Bitişik ve bağımsız Q-matrisi tasarımı yaklaşımlarının, aynı uzunluktaki testler için benzer sınıflama sonuçları verdiği gözlenmiştir. Test uzunluğu sınırlandırıldığı ya da madde nitelikleri düşük olduğunda ise bitişik Q-matrisi yaklaşımının diğer iki yaklaşımdan

daha doğru sınıflama yaptığı görülmüştür. Niteliklerin ölçülme sayısı sabit alındığında ise niteliklerin birbirinden izole olarak ölçüldüğü bağımsız yaklaşım, diğer iki yaklaşımdan daha yüksek doğru sınıflama oranına sahiptir. Yapılan simülasyon çalışması madde niteliğinin, nitelik ve nitelik profillerinin doğru sınıflama oranı üzerinde önemli bir etkisi olduğunu göstermiştir. Yüksek nitelikli maddeler kullanıldığında, çalışmaya dahil edilen diğer tüm faktörlerden bağımsız olarak, daha yüksek sınıflama doğruluğu ve güvenilirliği elde edilmiştir. Q-matrisi tasarımındaki üç yaklaşımdan ulaşılabilir Q-matrisi tasarımı, düşük nitelikli maddeler kullanıldığında modeldeki üst düzey etkileşimlerin kestirilmesi zor olduğu için madde havuzunun niteliğinden daha fazla etkilenmiştir. Çalışmanın sonuçları, test uzunluğu sınırlandırıldığı ya da madde niteliği düşük olduğunda, bitişik Q-matrisi tasarımının daha başarılı sınıflama yaptığını göstermiştir.

Liu, Huggins-Manley ve Bulut (2017), BTM'yi mevcut MTK temelli bir teste uyarlayarak uyarlama sürecinde izlenecek adımlardan, uyarlamanın avantaj ve zorluklarından bahsetmiş ve BTM'nin uyarlanması ile elde edilebilecek bilgileri incelemiştir. Uyarlama sürecinin aşamalarını; 1) bilginin toplanması, 2) nitelik ve maddeler arası ilişkilerin belirlenmesi, 3) madde tepkilerinin modellenmesi ve 4) sonuçların yorumlanması şeklinde açıklamışlardır. Okuyucunun uyarlama sürecini daha iyi anlamasına yardımcı olabilmek için 422 katılımcının 51 maddelik TOEFL mock dinleme testine verdikleri cevaplar üzerinden bir uyarlama çalışması örneği sunmuşlardır. Mock test, Çin'deki bir test geliştirme firması tarafından içerik, form ve test uzunluğu açısından TOEFL testine benzeyecek şekilde geliştirilmiştir. Test, 17'şer soruluk 3 bölümden oluşmakta ve her bölüm aynı yapıyı farklı güçlük seviyelerinde (kolay, orta ve zor) ölçmektedir. BTM'lerin model-veri uyumlarını değerlendirmek üzere G-DINA, A-CDM, DINA, DINO ve HO-DINA modelleri kullanılmıştır. Analiz sonuçları, bütün modeller için SRMSR değerlerinin 0.05'ten küçük olduğunu ancak BTM'lerin mutlak model uyumunun, orta güçlük seviyesindeki madde havuzunda, kolay ve zor seviyelerdeki madde havuzlarına göre daha iyi olduğunu göstermiştir. Modellerin göreceli uyum indeksleri de birbirine oldukça benzer olmakla birlikte, 3 madde havuzunda da AIC değerlerine göre DINA modeli en iyi uyumu gösterirken, BIC ve CAIC değerlerine göre en iyi uyumu HO-DINA göstermiştir. Madde uyumunu değerlendirmek için RMSEA kullanılmış ve G-DINA modelinin orta güçlükteki madde havuzundaki kestirimleri dışında, tüm

modellerin RMSEA deęerlerinin 0.05'ten küçük olduęunu bulunmuştur. Modellerin RMSEA deęerleri zor güçlükteki madde havuzunda en küçük ve kolay güçlükteki madde havuzlarında en yüksek deęerini alırken, BTM'ler arasından G-DINA modeli, en kötü madde uyumunu sergilemiştir. Nitelikler arası kestirilen tetrakorik korelasyonlar incelendięinde ise tüm madde havuzları için bu deęerlerin 0.70 ve üzeri olduęu görölmektedir. Araştırmacılar bu durumun, BTM'ye uyarlaması yapılan verinin aslında tek boyutlu olmasından kaynaklanabileceęini belirtmişlerdir. Örtük sınıfların dağılımı incelendięinde ise üç madde havuzu için de katılımcıların büyük çoęunluęunun tüm niteliklere sahip olan ve hiçbir nitelięe sahip olmayan örtük sınıflarında olduęu görölmüştür.

Yi (2017) çalışmasında, İkinci Dil Olarak İngilizce (English as a Second Language: ESL) gramer testi verisi için en uygun bilişsel tanı modelini incelemiştir. Çalışmanın verilerini, Henson ve Templin'nin 2007 yılında yaptıkları çalışmada kullandıkları, 2922 katılımcının İngilizce Yeterlik Sertifika Sınavı (Examination for the Certificate of Proficiency in English: ECPE) gramer testindeki 30 soruya verdikleri cevaplar oluşturmaktadır. Q-matrisi, Liao'nun (2007) gramer bilgisinin teorik modeline yönelik yaptıęı açımlayıcı faktör analizi ve yapısal eşitlik modeli çalışmalarına ve ECPE gramer testinde başarılı olmak için gereken gizil yeteneklere göre oluşturulmuştur. İlk olarak, genel model çerçevesinde LCDM'nin uyumu incelenmiş, ardından C-RUM, NIDO, DINO ve DINA modellerinin hangisinin daha fazla maddeye uyumu sağladıęı araştırılmıştır. Madde uyumu analizleri C-RUM'un testteki en fazla maddeye uyum sağladıęını ve bunu sırasıyla DINA, DINO ve NIDO modellerinin takip ettięini göstermiştir. Mutlak uyum indekslerine göre bütün BTM'ler veriye kabul edilebilir uyum gösterdięi için BTM'nin çok boyutlu analizlerinin yapılması uygun görölmüş ve analizlere devam edilmiştir. Çalışmada tek boyutlu 2PL MTK'nın, çok boyutlu BTM'lerden daha iyi göreceli model uyumu gösterdięi bulunmuştur. Araştırmacı bu durumun, BTM'lerin mevcut MTK verisine uyarlanmasından kaynaklandıęını belirtmiştir. BTM'lerin göreceli model uyumları incelendięinde, LCDM ve C-RUM'un en iyi, telafi edici modellerden NIDO'nun ise en kötü model uyumu gösterdięi tespit edilmiştir. Modele en iyi uyum sağlayan LCDM ve C-RUM'a göre yapılan analizler sonucunda, bireylerin çocuęun hiçbir nitelięe sahip olmayan (0,0,0) ve niteliklerin hepsine sahip olan (1,1,1) iki uç nitelik profilinde sınıflandıęı görölmektedir. Bireylerin %13'ünün (0,0,1) ve %18'inin (0,1,1)

sınıflarında olması, nitelik 1'in (biçim-sözdizimsel bilgi) ECPE testini alan bireyler için en zor nitelik olduğunu göstermektedir. LCDM ve C-RUM'a göre her bir nitelik için ortalama geçen kişi yüzdesinin ise neredeyse aynı olduğu görülmektedir. Tek boyutlu 2PL MTK ve C-RUM'un parametre kestirimleri arasındaki korelasyonların da oldukça yüksek olduğu bulunmuştur (0.82 ile 0.88 arasında değişmektedir).

Aryadoust (2018) çalışmasında, Singapur-Cambridge Genel Eğitim Sertifikası dinleme testinin altında yatan yapıyı incelemiştir. Çalışmanın katılımcılarını Singapur'daki 6 ortaokulda öğrenim gören 205 öğrenci oluşturmaktadır. Test, 3 aylık eğitim boyunca katılımcılara çoklu oturumlarda uygulanmış ve her öğrenciye madde ve nitelik düzeyinde bireysel tanılayıcı rapor, okullara da öğrencilerin genel performanslarına ilişkin bir rapor verilmiştir. Q-matrisinin geliştirilmesinde, çalışmanın kuramsal çerçevesi, 20 katılımcı ile yapılan sesli düşünme protokolleri ve göz takibinden (eye-tracking) yararlanılmıştır. Bilişsel tanı modellerinden DINA, G-DINA, DINO, HO-DINA ve R-RUM'un model veri uyumları göreceli ve mutlak uyum indeksleri ile karşılaştırılmış, R-RUM'un en iyi ve DINO'nun en kötü model uyumu gösterdiği bulunmuştur. R-RUM'un örneklem düzeyinde, sınıflama doğruluğu ve tutarlılığı değerleri sırasıyla 0.765 ve 0.860 olarak bulunmuştur. Nitelik bazında sınıflama doğruluğu değerlerinin ise 0.073-0.889 arasında değiştiği ve 6 nitelik için oldukça düşük olduğu görülmüştür. Sınıflama tutarlılıkları ise 0.542 ile 0.988 arasında değişmektedir. Örtük sınıfların dağılımı incelendiğinde, katılımcıların %67.3'ünün (137 kişi) 9 niteliğin hepsine sahip olduğu, hiçbir niteliğe sahip olmayan katılımcı sayısının ise 0 olduğu tespit edilmiştir. Araştırma sonucunda, katılımcıların dinleme testi performansında etkili olabilecek faktörlerden bahsedilmiştir.

Tu, Wang, Cai, Douglas ve Chang (2018), nitelikler arasında hiyerarşik ilişkilerin tanımlanmasının model uyumu, madde parametrelerinin tahmini ve sınıflama doğruluğu üzerine etkilerini simülasyon çalışması ile incelemiştir. Çalışmanın verileri DINA, R-RUM, A-CDM ve G-DINA modelleri ile üretilmiştir. Nitelikler arası hiyerarşik ilişkiler ve bu ilişkilerin dikkate alınmadığı, iki kestirim yapılmıştır. Örneklem büyüklüğünün 1000'e, madde sayısının 30'a ve nitelik sayısının 5'e sabitlendiği çalışmada doğrusal, yakınsak, ıraksak ve karma hiyerarşik yapıları kullanılmıştır. Sınırlandırılmış ve sınırlandırılmamış kestirimlerin model uyumu AIC ve BIC değerleri ile, madde parametre kestirimleri RMSE değeri ile,

sınıflama doğruluğu ise nitelik profillerinin doğru sınıflanma oranları ve izin verilmeyen nitelik örüntüsü oranı (IPR=izin verilmeyen nitelik profili kestirimleri/örneklem büyüklüğü) ile karşılaştırılmıştır. IPR'nin küçük değer alması, kestirim yönteminin doğru sınıflama oranının yüksek olduğunu göstermektedir. Yapılan simülasyon çalışmaları sonucunda, tüm koşullar için sınırlandırılmış modellerin daha iyi model uyumu gösterdiği bulunmuştur. Madde parametreleri için RMSE değerleri incelendiğinde, DINA modelinin her iki kestiriminin birbirine benzer ve düşük RMSE'ye sahip olduğu; R-RUM, A-CDM ve G-DINA modellerinde ise sınırlandırılmış ve sınırlandırılmamış madde parametresi kestirimlerin RMSE değerleri arasında sınırlandırılmış modellerin lehine bir fark olduğu görülmektedir. Farklı nitelik yapıları altında bireylerin doğru sınıflanma oranları incelendiğinde, modellerin sınırlandırılmış kestirimlerinin daha başarılı olduğu, sınırlandırılmış ve sınırlandırılmamış kestirimlerin doğru sınıflama oranları arasındaki farkın DINA modeli için %35-%50, R-RUM için %30-%42, A-CDM için %34-%51 ve G-DINA için %40-%61 arasında değiştiği gözlenmiştir. Modellerin sınırlandırılmamış kestirimleri için IPR değerleri ise %42.6 ile %47.8 arasında değişmektedir. Modellerin sınırlandırılmış (16) ve sınırlandırılmamış nitelik profili (2<sup>7</sup>) kestirimlerini gerçek veri uygulaması üzerinde karşılaştırmak üzere 769 lisans öğrencisinin 15 maddeye verdikleri cevaplar kullanılmıştır. BIC değerlerine göre tüm modellerin sınırlandırılmış kestirimleri özellikle genel modellerde daha iyi model uyumu gösterirken, AIC değerlerine göre model uyumları değişkenlik göstermiştir. Modeller için hesaplanan IPR değerleri ise DINA, R-RUM, A-CDM, G-DINA, LCDM ve HDCM için sırasıyla %40.7, %61.1, %56.2, %76.9, %66.1 ve %89.5'tir.

Tu, Zheng, Cai, Gao ve Wang (2018) çalışmalarında, DINA modelini temel alarak çok kategorili veriler için DINA-GD modelini geliştirmişlerdir. Önerdikleri modelin birey ve madde parametrelerini kestirmede uygulanabilirliğini incelemek üzere iki simülasyon çalışması ve gerçek veri uygulaması yapmışlardır. DINA-GD'nin birey ve madde parametrelerini doğru kestirmedeki başarısını değerlendirmek için doğru nitelik profili sınıflama oranı (CPCR), ortalama nitelik uyum oranı (AAMR) ve ortalama mutlak yanlılık değerlerini (MAB) kullanmışlardır. Yaptıkları birinci simülasyon çalışmasında nitelik sayısı (5, 7 ve 9), örneklem büyüklüğü (500, 1000 ve 1500) ve test uzunluğu (20 ve 40) olmak üzere 18 koşul incelenmiş ve tüm nitelikler birbirinden bağımsız olarak ele alınmıştır. Birinci

simülasyon çalışmasının sonuçlarına göre tüm koşullar için DINA-GD'ye göre kestirilen AAMR değerlerinin 0.90'dan büyük olduğu gözlenmiştir. Test uzunluğunun 40 madde olduğu koşullar için CPCR değerlerinin 0.94'ten büyük olduğu, test uzunluğunun 20 madde olduğu koşullar için ise ortalama CPCR değerlerinin 0.78 olmakla birlikte en düşük CPCR değerinin 0.62 olduğu bulunmuştur. Madde parametreleri için kestirilen ortalama MAB değerlerinin, kaydırma ve tahmin parametreleri için sırasıyla 0.034 ve 0.038 olduğu gözlenmiştir. Bu bulgular, çalışmada önerilen DINA-GD modelinin yanlılık değerlerinin makul olduğunu ve yüksek bir doğru sınıflama oranına sahip olduğunu göstermektedir. Yaptıkları ikinci simülasyon çalışmasında ise nitelik sayısını 5'e, örneklem büyüklüğünü 1000'e ve test uzunluğunu 20'ye sabitleyip, iki kategorili veriler için DINA modeli ile çok kategorili veriler için DINA-GD modelini karşılaştırmışlardır. Çok kategorili verilerin iki kategorili veriye dönüştürülmesi ile DINA modelinden elde edilen AAMR ve CPCR değerlerinin, DINA-GD'ye göre oldukça düşük olduğu bulunmuştur. Son olarak, PISA 2006 fen okuryazarlığı verisi GDM ve DINA-GD ile analiz edilmiş ve DINA-GD'nin daha iyi model ve madde uyumu gösterdiği, iki modelin sınıflama tutarlılıkları arasında ise fark olduğu tespit edilmiştir.

### **Çok Boyutlu MTK ve Bilişsel Tanı Modellerinin Karşılaştırılması ile İlgili Araştırmalar**

de la Torre ve Karelitz (2009) yaptıkları simülasyon çalışmalarında, hiyerarşik nitelik yapısı altında DINA ve 2PL tek boyutlu MTK'nın birbirine karşılıklı uyarlanmasını incelemiştir. Çalışmada 5 nitelik kullanılmış ve nitelikler arasında doğrusal bir ilişki tanımlanmıştır. MTK kapsamında kestirilen sürekli yetenek puanlarını kesikli hale dönüştürmek için madde güçlük parametresini kullanmışlardır. Çalışmada madde ayırt edicilik parametresi (düşük, orta ve yüksek), gerçek model (2PL MTK ve DINA) ve kestirilen model (2PL MTK ve DINA) değişkenleri manipüle edilerek 12 koşul oluşturulmuş ve her bir koşul 100 tekrar ile gerçekleştirilmiştir. 2PL MTK madde parametrelerini DINA modelinin  $g$  ve  $s$  parametrelerine dönüştürmek için lojistik adım düşünümü (logistic-to-step transformation: LST) metodunu geliştirmişlerdir. Analiz sonuçları maddeler düşük ayırt ediciliğe sahip olduğunda, gerçek ve uyarlanan modelin model uyumları arasındaki farkın diğer koşullara göre daha küçük olduğunu ancak tüm koşullarda



verinin üretildiği modelin daha iyi uyum sağladığını göstermiştir. MTK ve BTM verilerinin DINA modeli ile analizlerinde madde parametrelerinin kestirim doğruluğu ayırt edicilik parametresine göre monotonik olarak artmamış, her iki veri seti için orta ayırt edicilik düzeyinde en yüksek hata ile kestirim yaptığı gözlenmiştir. DINA modelinin nitelik ve nitelik vektörü bazında yanlış sınıflama oranları ise maddelerin ayırt edicilik parametresi arttıkça düşmektedir. Ancak madde ayırt ediciliği düşük olduğunda, DINA modelinin uç nitelik sınıflarındaki nitelik vektörü bazında yanlış sınıflama oranının daha yüksek olduğu bulunmuştur. 2PL MTK modelinin BTM ve MTK verisi ile analizlerinde ise madde ayırt ediciliği arttıkça güçlük parametresi için RMSE'nin küçüldüğü, ayırt edicilik parametresi için ise arttığı görülmüştür. 2PL MTK parametreleri DINA parametrelerine dönüştürüldüğünde ise kestirimlerin doğruluğunun nitelik sınıflarına göre değişkenlik gösterdiği, genellikle orta nitelik sınıfı için madde parametrelerinin ( $g$  ve  $s$  parametresi) en doğru şekilde kestirildiği tespit edilmiştir. Bireylerin doğru sınıflanma oranları incelendiğinde ise madde ayırt ediciliği arttıkça RMSE değerlerinin ve nitelik sınıflarının yanlış sınıflanma oranlarının düştüğü görülmüştür. Çalışmanın sonucunda, BTM'nin MTK verisine uyarlanmasının yüksek nitelikli maddeler kullanıldığında daha uygun olduğu belirtilmiştir.

Wang (2009), telafi edici çok boyutlu MTK ve R-RUM modellerini birbirine uyarladığı çalışmasında, modelleri sağladığı geri bildirimler açısından karşılaştırmıştır. Test niteliği (düşük, orta ve yüksek ayırt edicilik), Q-matrisi yapısı (basit ve karmaşık yapı) ve nitelikler arası korelasyon (0.20, 0.50 ve 0.90) değişkenlerinin manipüle edildiği Monte Carlo Markov Chain (MCMC) simülasyon çalışmasında tüm koşullar için örneklem büyüklüğü 2000, madde sayısı 40 ve nitelik sayısı 4 alınmış ve 10'ar tekrar ile iki simülasyon çalışması yapılmıştır. Birinci çalışmada R-RUM ve 2PL ÇB-MTK modellerini test niteliği ve madde parametreleri açısından karşılaştırmak üzere BTM ve MTK'ya göre üretilen veriler her iki model ile analiz edilmiştir. Analiz sonuçları, iki modelin test niteliğini farklı şekilde tanımladığını ve madde parametrelerinin birbirleriyle düşük düzeyde ilişkili olduğunu göstermiştir. İkinci çalışmada ise modeller sağladıkları geri bildirimler ve doğru sınıflama oranları açısından ham uyuma oranları ve Cohen'in kappa katsayısı ile karşılaştırılmıştır. Kappa istatistiği sonuçları, R-RUM'a göre üretilen veriler için iki model arasındaki uyumun daha yüksek olduğunu göstermiştir. Ayrıca test niteliği

düştükçe, modellerin bireyleri sınıflamada (başarılı-başarısız) daha az ayırt edici olduğu ve modeller arası uyumun düştüğü bulunmuştur. Nitelikler arası korelasyon arttığında ise test daha fazla tek boyutlu gibi olmuş ve modeller arasındaki uyum düşmüştür. Düşük korelasyon koşulunda modeller arası uyumun orta ve yüksek korelasyon koşullarına göre daha yüksek olduğu görülmüştür. Son olarak, modeller arası uyumun Q-matrisi yapısından etkilendiği, basit yapıdaki Q-matrisi için modeller arası uyumun karmaşık yapıdaki Q-matrisine göre daha yüksek olduğu bulunmuştur.

Bradshaw ve Templin (2014) adlandırılmalı tepki tek boyutlu MTK ve bilişsel tanı modellerini birleştirerek, bireylerin yeteneklerini ölçmek ve onları kavram yanlışlarına göre sınıflandırmak üzere SCIM (Scaling Individuals and Classifying Misconceptions) modelini önermiştir. SCIM modeline göre kestirimlerinin doğruluğu ve güvenilirliğini değerlendirmek için yaptıkları simülasyon çalışmasında kavram yanlışlığı sayısı (3 ve 6), örneklem büyüklüğü (3000 ve 10000), test uzunluğu (30 ve 60 madde), yetenek ve kavram yanlışlığı için temel etkilerin büyüklüğü (düşük ve yüksek) olmak üzere 4 faktör manipüle edilmiştir. Düşük temel etkiler için  $U(0.3, 0.6)$  ve kavram yanlışlığı için  $U(0.75, 1.25)$  dağılımından, yüksek temel etkiler için  $U(0.6, 0.8)$  ve kavram yanlışlığı için  $U(1.75, 2.25)$  dağılımından üretilmiştir. Düşük ve yüksek temel etki koşullarını çaprazlayarak, bu gizil değişkenler için temel etkilerin büyüklüğünü ve göreceli büyüklüklerini değiştirmişlerdir. Madde kesişimleri  $U(-1, 1)$ , iki yönlü etkileşimler  $U(0.5, 1)$  ve bireylerin yetenek kestirimi standart normal dağılımdan üretilmiştir. Dört seçenekli testte, doğru cevabın dışındaki her bir yanlış seçenek bir ya da iki kavram yanlışlığını ölçecek şekilde Q-matrisi oluşturulmuştur. Kavram yanlışlığı arasındaki tetrakorik korelasyonlar ise 0.5'e sabitlenmiştir. Oluşturulan 32 koşul, 50'şer tekrar ile gerçekleştirilmiştir. Test uzunluğunun artması, kavram yanlışlığı sayısının azalması ve örneklem büyüklüğünün artması kestirimlerin doğruluğunu arttırmıştır. Yapısal parametre kestirimlerinin doğruluğunun ise en fazla kavram yanlışlığı sayısından etkilendiği bulunmuştur. Birey parametre kestirimlerinin doğruluğu ve güvenilirliği testi alan birey sayısından daha az etkilenirken; testin uzunluğu, kavram yanlışlığı sayısı ve maddelerin niteliğinden daha fazla etkilenmiştir. Birçok koşul için kavram yanlışlıklarının doğru sınıflanma oranının (CCR) 0.90 ve üzeri olduğu gözlenmiştir. SCIM'in gerçek veri üzerinde uygulamaları, Newton Kuvveti konusundaki kavram

yanılgılarını belirlemek amacıyla geliştirilen FCI testine 10039 öğrencinin verdikleri cevaplar kullanılarak yapılmıştır. Model veri uyumu SCIM ve adlandırılmalı tepki MTK modeline göre incelenmiştir. Göreceli model uyumu sonuçları, yordayıcı olarak kavram yanılgılarını içeren modelin veriye tek bir sürekli niteliğin kestirildiği MTK'dan daha iyi uyum sağladığını ve testin sürekli bir özellikten daha fazlasını ölçtüğünü göstermiştir.

Sen ve Bradshaw (2017) yaptıkları simülasyon çalışması ile çeşitli koşullar altında BTM ve MTK modellerinin model uyumlarını göreceli uyum indeksleri ile karşılaştırmıştır. Bu doğrultuda yaptıkları iki simülasyon çalışmasının ilkinde çok boyutlu modelleri (3 boyutlu model), ikincisinde ise tek boyutlu modelleri incelemişlerdir. Çalışmanın simülasyon koşulları; örneklem büyüklüğü (1000), test uzunluğu (12 ve 24 madde), test maddelerinin niteliği (orta ve yüksek), nitelik sayısı (1 ve 3), niteliğe sahip kişi (geçen kişi) oranı (0.25 ve 0.50), tetrakorik korelasyon (0.7), verilerin üretilmesinde kullanılan model (doygun LCDM, indirgenmiş LCDM, Rasch modeli, 2PL ve 3PL MTK modelleri) ve kestirim modelleri (doygun LCDM, indirgenmiş LCDM, DINA, DINO, C-RUM, ÇB-MTK, rasch model, 2PL ve 3PL MTK) şeklindedir. Her bir koşul 100 tekrarla yapılmıştır. Birinci simülasyon çalışması için Q-matrisi (1,0,0), (0,1,0), (0,0,1), (1,1,0), (1,0,1) ve (0,1,1) nitelik örüntülerinin her biri 4'er maddeyle ölçülecek ve nitelikler arasında 0.70 korelasyon olacak şekilde oluşturulmuştur. Çalışmanın verileri ise doygun ve indirgenmiş LCDM'ye göre üretilmiş, doygun ve indirgenmiş LCDM, DINA, DINO, C-RUM ve ÇB-MTK'ya göre kestirilmiştir. İkinci simülasyon çalışmasının verileri ise tek nitelikli LCDM, rasch model, 2PL ve 3PL MTK modellerine göre üretilmiş ve aynı modeller ile kestirilmiştir. Kestirilen madde parametrelerinin doğruluğunu RMSE, birey parametrelerinin doğruluğunu ise gerçek ve kestirilen birey profillerinin eşit olma yüzdesine göre hesaplamışlardır. Modellerin göreceli uyumları ise AIC, BIC ve SABIC değerlerine göre karşılaştırılmıştır. Çalışmada, madde ayırt ediciliği arttıkça kestirimlerin doğruluğunun da arttığı ve daha basit yapıdaki indirgenmiş LCDM ile üretilen verilerin, doygun LCDM'e göre daha doğru kestirildiği gözlenmiştir. Ayrıca, MTK modellerinde olduğu gibi tek niteliğin kestirildiği bilişsel tanı modellerinin 3 niteliğin kestirildiği modellere göre daha doğru kestirim yaptığı bulunmuştur. Madde güçlük parametresi için ortalama RMSE değerleri Rasch, 2PL ve 3PL MTK modelleri için sırasıyla 0.103, 0.188 ve 0.249 şeklindedir. Madde ayırt edicilik parametresi için

ortalama RMSE deęerleri ise 2PL MTK için 0.136 ve 3PL MTK için 0.145'tir. Şans parametresinin ortalama RMSE deęeri ise 0.009'dur. Bireylerin doęru sınıflanma oranları, madde kalitesinin artması ve daha basit yapıda model kullanılmasıyla artış göstermiştir. Genel olarak, AIC, BIC ve SABIC deęerlerinin madde nitelięinin yüksek olduęu koşullarda daha iyi performans gösterdięi, madde nitelięi düştüğünde doęru modeli kestirmede başarısız olduęu gözlenmiştir.

Yakar (2017) doktora tez çalışmasında, bilişsel tanı ve çok boyutlu MTK modellerinin birbiri yerine kullanımını incelemiştir. Bu doęrultuda, BTM ve MTK'ya göre üretilen veriler, çok kategorili BTM modellerinden pG-DINA, fA-M ve 2PL ÇB-MTK modeli ile analiz edilmiştir. Araştırma kapsamında madde ayırt edicilięi (düşük, orta ve yüksek), madde yapılarının testteki oranı (basit ve karmaşık madde oranları: %20-%80, %50-%50 ve %80-%20), madde sayısı (15, 30 ve 60) ve yetenekler arası korelasyon (0.0 ve 0.6) deęişkenlerinin etkisi araştırılmıştır. Her bir koşulun 25 tekrarla gerçekleştirildięi çalışmada, 5000 ve 100000 kişilik iki örneklem büyüklüęü kullanılmıştır. ÇB-MTK kapsamında 2 boyutlu olarak tasarlanan yapı, BTM çerçevesinde 4'er düzeye sahip iki nitelik olarak tanımlanmıştır. Bireylerin doęru sınıflanma oranları (DSO) vektör, nitelik ve yeterlik düzeyleri için hesaplanmıştır. Analiz sonuçlarına göre, modellerin DSO'ları üzerinde en etkili deęişkenin madde sayısı olduęu ve bu koşulu sırasıyla madde ayırt edicilięi, madde yapısı ve yetenekler arası korelasyon deęişkenlerinin izledięi bulunmuştur. BTM'lerin ÇB-MTK verisine uyarlanmasında, her koşulda fA-M'ların ÇB-MTK ile yakın DSO'ya sahip olduęu, pG-DINA modelinin DSO deęerlerinin ise daha düşük olduęu görülmüştür. Telafi edici ve telafi edici olmayan verilerden elde edilen DSO deęerleri arasındaki farkın çok büyük olmamakla birlikte, telafi edici veri lehine olduęu gözlenmiştir. pG-DINA ve fA-M verilerinin ÇB-MTK, pG-DINA ve fA-M ile analizleri incelendięinde, pG-DINA modelinin kendi verisini yüksek bir DSO ile kestirdięi görülmüştür. pG-DINA verisi, fA-M'lar tarafından başarılı bir şekilde kestirilirken, ÇB-MTK ise kötü kestirimde bulunmuştur. Araştırma kapsamında önerilen fA-M'lar ise pG-DINA ve ÇB-MTK verilerine başarılı bir şekilde uyarlanabilmiştir.

## İlgili Araştırmalar Özet

Yapılan çalışmalar incelendiğinde, BTM'nin büyük ölçekli mevcut testlere uyarlanmasının yapıldığı çalışmaların ağırlıklı olduğu ve uyarlama yapılan alanların genel olarak matematik, okuma becerisi ve dil eğitimi ile ilgili olduğu görülmektedir. BTM ve MTK'nın karşılıklı uyarlanmasının yapıldığı çalışmaların sayısı ise oldukça sınırlıdır. Çünkü MTK modelleri tanılayıcı dönüt üretmek için kullanılmasına rağmen daha kapsamlı çok boyutlu dönüt verme konusunda yetersiz kalabilmektedir. Yapılan literatür çalışmaları da verinin üretildiği modele daha iyi uyum gösterdiğini raporlamıştır. Bu doğrultuda, BTM verisinin MTK'ya uyarlanması çok anlamlı olmadığı için BTM'nin mevcut madde yanıtlarına uygulanması ile tanılayıcı dönüt elde edilmeye çalışılmaktadır. Q-matrisinin oluşturulmasında ise uzman yargıları, ilgili literatür ve faktör analitik yöntemler kullanılmıştır. BTM'lerin gerçek veri uygulaması üzerine yapılan çalışmalarda, hangi modelin veriye daha iyi uyum sağladığı mutlak ve göreceli uyum indeksleri ile incelenmiş, bireylerin nitelik ve nitelik vektörü bazında profilleri değerlendirilmiştir. Bireylerin nitelik profillerine göre dağılımının incelenmesi ile nitelikler arasındaki ilişkiler ve bireylerin güçlü ve zayıf yönlerine yönelik çıkarımlar yapılmıştır. Ancak çok az sayıda çalışmada öğretmen, öğrenci ya da kurumlara geri bildirimde bulunulmuştur.

BTM ve MTK'nın birbirine karşılıklı uyarlandığı çalışmaların neredeyse hepsi çeşitli faktörlerin manipüle edildiği simülasyon çalışmalarıdır. Etkisi incelenen değişkenler, genel olarak, örneklem büyüklüğü, nitelikler arası korelasyon, nitelik sayısı, test uzunluğu, Q-matrisi yapısı, madde niteliği, verilerin üretilmesi ve analizinde kullanılan modeller şeklindedir. Modeller arası karşılaştırmalar ise mutlak ve göreceli model uyumu indekslerine, nitelik ve nitelik vektörü bazında doğru sınıflama oranlarına göre yapılmıştır. Çalışmaların büyük bir çoğunluğu, modeller arası karşılaştırmaların yapılabilmesi adına MTK çerçevesinde elde edilen sürekli yetenek puanlarını kesikli hale getirmek için madde güçlük parametresini kullanmış, BTM sınıflamalarına ise herhangi bir müdahalede bulunmamıştır. Bazı çalışmalar ise kesme puanının belirlenmesi sürecinden bahsetmemiştir. Ancak, yapılan karşılaştırmaların geçerli olabilmesi için en önemli hususlardan biri BTM ve ÇB-MTK modelleri ile yapılan kestirimlerin uyumunu maksimum düzeye çıkaracak kesme puanının belirlenmesidir. Aksi halde analiz sonuçlarına dayalı yapılan çıkarımlar, gerçek durumu yansıtmayacaktır. BTM ve MTK'nın birbirine uyarlandığı

alıřmalarının sonuları incelendiėinde ise verinin retildiėi modellerin diėer modellere gre daha iyi uyum saėladıėı ancak bazı kořullar altında ok benzer kestirimler yaptıkları grlmřtr.



## **Bölüm 3**

### **Yöntem**

Bu bölümde araştırmanın türü, gerçek verilerin elde edilmesi ve analizler için istenilen koşullara uygunluğunun test edilmesi, Q-matrisinin oluşturulması, bilişsel tanı modellerinin model-veri uyumunun test edilmesi, simülasyon koşulları, verilerin üretilmesi, analizlerde kullanılacak programlar ve değerlendirme kriterleri ile ilgili bilgilere yer verilmiştir.

#### **Araştırmanın Türü**

Bu araştırma kapsamında, 3PL ÇB-MTK'ya göre üretilen verilerin farklı test koşulları altında 3PL ÇB-MTK ve BTM modelleri ile analizinden kestirilen birey parametrelerinin hata ve yanlışlık değerleri, nitelik ve nitelik profili bazında doğru sınıflama oranları karşılaştırılmıştır. İlgili modellerin farklı koşullar altında test edilmesi ile hangi koşullar altında birbiriyle tutarlı ya da tutarsız kestirim yaptıkları incelenmiştir. Bu çalışma, ÇB-MTK ve BTM modellerinin birbiri yerine kullanılma imkanları hakkında mevcut bilgiye katkıda bulunacağı ve BTM'lerin tanılayıcı bilgi sağlama amacı taşımayan mevcut veriye uyarlanmasına yönelik çalışmaların gelişmesine katkı sağlanacağı düşünüldüğü için temel araştırma niteliği taşımaktadır (Karasar, 2009; Neuman, 2014). Araştırmanın TIMSS 2015 uygulaması matematik testinin sağladığı tanılayıcı bilgilerin verildiği ve 3PL ÇB-MTK ve BTM modellerine göre kestirilen birey parametresi kestirimlerinin nasıl değiştiğinin incelendiği son bölüm ise betimsel araştırma özelliği taşımaktadır.

#### **Verilerin Elde Edilmesi**

Çalışmada 3PL ÇB-MTK ve BTM'lere göre kestirilen birey parametrelerini karşılaştırmak amacıyla Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS-Trends in International Mathematics and Science Study) 2015 uygulaması matematik testi verilerinin madde parametrelerine göre oluşturulan veri setlerinin kullanıldığı simülasyon çalışması ve gerçek veri uygulaması yapılmıştır. Bu bölümde, öncelikle çalışma kapsamında kullanılan gerçek veri tanıtılmış, ardından verinin BTM uygulamaları için hazırlanmasından ve simülasyon koşullarından bahsedilmiştir.

Araştırmada kullanılan veri setleri, Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu'nun (IEA-International Association for the Evaluation of Educational Achievement) 2015 yılında düzenlemiş olduğu TIMSS kapsamında 8. sınıf öğrencilerine uygulanan 14 test kitapçığından 8 numaralı kitapçıktan elde edilen gerçek veri parametrelerine dayalı olarak üretilmiştir. İncelenen matematik kitapçıkları arasından kitapçık 8'in seçilme nedeni, bilişsel tanı ve çok boyutlu MTK modellerinin temel gereksinimlerinden, çok boyutluluk varsayımını sağlamasıdır. TIMSS 2015 uygulaması 8. kitapçıkta yer alan matematik sorularını 20396 öğrenci cevaplamıştır. Ancak veri setinin eksik, kayıp ve aykırı değerler açısından incelenmesi sonrasında bu araştırmada 9208 öğrencinin cevapları üzerinden yürütülmüştür. Araştırmada kullanılan veriler TIMSS & PIRLS International Study Center internet sitesinden (<https://timssandpirls.bc.edu>) elde edilmiştir.

TIMSS 2015 uygulamasında yer alan 225 matematik sorusu, konu alanları ve bilişsel düzeyleri dikkate alınarak 14 bloğa dağıtılmıştır. Her kitapçıkta 2 blok matematik sorusu olacak şekilde, 14 matematik test kitapçığı oluşturulmuştur. TIMSS 8. sınıf matematik testi için konu alanları; sayılar, cebir, geometri, veri ve olasılık olarak belirlenmiştir. Çalışma kapsamında ele alınan kitapçık 8'deki soruların konu alanlarına göre dağılımı incelendiğinde soruların %33.33'ünün (10 soru) sayılar, %26.67'sinin (8 soru) cebir, %20'sinin (6 soru) geometri ve %20'sinin (6 soru) veri ve olasılık konu alanlarında olduğu görülmektedir. 8. sınıf matematik testinin bilişsel düzeyleri ise bilme, uygulama ve akıl yürütme olmak üzere üç kategoride ele alınmaktadır. Soruların bilişsel düzeylere göre dağılımı %40'ı (12 soru) bilme, %46.67'si (14 soru) uygulama ve %13.33'ü (4 soru) akıl yürütme şeklindedir.

TIMSS 2015 uygulaması kitapçık 8'de, 16 çoktan seçmeli ve 14 açık uçlu (çok kategorili olarak puanlanan) toplam 30 madde bulunmaktadır. Çok kategorili olarak puanlanan maddeler, tam doğru cevap 1 (doğru cevap) ve diğer cevaplar 0 (yanlış cevap) olacak şekilde kodlanarak iki kategorili hale getirilmiştir. Testte yer alan 30 maddeye ilişkin betimsel istatistikler aşağıdaki tabloda yer almaktadır.



Tablo 2

*TIMSS Matematik Testi 8. Kitapçığın Maddelerine ait Betimsel İstatistikler*

No	Madde	Tür	p	No	Madde	Tür	p
1	M062005	ÇS	0.637	16	M052413	ÇS	0.741
2	M062139	AU	0.419	17	M052134	ÇS	0.705
3	M062164	ÇS	0.655	18	M052078	ÇS	0.460
4	M062142	AU	0.636	19	M052034	ÇS	0.608
5	M062084	ÇS	0.264	20	M052174A	AU	0.489
6	M062351	ÇS	0.398	21	M052174B	AU	0.298
7	M062223	ÇS	0.693	22	M052130	ÇS	0.397
8	M062027	AU	0.470	23	M052073	ÇS	0.552
9	M062174	ÇS	0.530	24	M052110	AU	0.414
10	M062244	AU	0.462	25	M052105	AU	0.193
11	M062261	ÇS	0.240	26	M052407	ÇS	0.662
12	M062300	AU	0.421	27	M052036	AU	0.480
13	M062254	AU	0.252	28	M052502	AU	0.627
14	M062132A	AU	0.658	29	M052117	AU	0.175
15	M062132B	ÇS	0.518	30	M052426	ÇS	0.750

ÇS: çoktan seçmeli, AU: açık uçlu

Tablo 2 incelendiğinde, madde güçlük indekslerinin 0.175 ile 0.750 arasında değiştiği ve testin ortalama güçlüğü'nün 0.493 olduğu görülmektedir. Henning'e (1987) göre soruların %20'sinin zor, %70'inin orta güçlükte ve %10'unun kolay olduğu söylenebilir.

### Matematik Testinin Faktör Yapısı

TIMSS matematik testinin genel amacı, öğrencilerin matematik alanındaki başarılarını ölçmek, eğitim ve öğretimin okullarda nasıl gerçekleştiğini, eğitim sisteminin etkinlik ve verimliliğini, ülkelerin eğitim sistemleri arasındaki farklılıkları belirlemek ve değerlendirmektir (MEB-Milli Eğitim Bakanlığı, 2016). Bu doğrultuda öğrencilere; sayılar, cebir, geometri, veri ve olasılık konu alanları ve genel matematik başarısına yönelik puan verilmektedir. Öğrencilere her bir konu alanındaki performanslarına yönelik bir puan verilmesi toplam puanlara göre daha kapsamlı ve tanılayıcı bilgi sağlmasına rağmen TIMSS'in temel amacı öğrencilerin güçlü ve zayıf yönlerine yönelik tanılayıcı bilgiler sunmak değildir. Bu yüzden, sorular ve soruları doğru cevaplamak için gereken niteliklerin eşleştirildiği Q-matrisi, test geliştirme sürecinde oluşturulmamıştır.

Eğitimdeki testlerde Q-matrisi, genel olarak, öğrenme ya da biliş teorisi, konu alanına hakim uzman yargıları, ampirik araştırma ya da katılımcıların sesli düşünme protokollerine dayalı olarak oluşturulabilir (Buck ve Tatsuoka, 1998). Q-matrisinin oluşturulmasında bu yöntemlerin dışında, faktör analitik yöntemler de kullanılabilir

(örneğin, Gierl, Tan ve Wang, 2005). Nitekim Q-matrisindeki girdiler, doğrulayıcı faktör analizinde faktör yapısını belirten yük göstergeleri olarak düşünülebilir (Rupp ve Templin, 2008a). Henson ve Templin (2007) de Q-matrisini geliştirirken faktör analizinden faydalanmış, açıklayıcı faktör analizi ile altta yatan yapıyı belirleyip ardından maddeleri tekrar gözden geçirerek, küçük değişikliklerle, Q-matrisine son halini vermişlerdir.

Yukarıda yapılan açıklamalar doğrultusunda, TIMSS uygulamalarında testteki tüm sorular paylaşılmadığı için Q-matrisini oluşturmak üzere matematik testinin faktör yapısı açıklayıcı faktör analizi (AFA) ile incelenmiştir. İlgili veriler, R programında yazılan psych (Revelle, 2019) paketinde Ağırlıklandırılmamış En Küçük Kareler (ULS) ve varimax döndürme yöntemine göre tetrakorik korelasyon matrisine dayalı olarak analiz edilmiştir.

AFA sonucu elde edilen Kaiser-Meyer-Olkin (KMO) değerinin 0.97 olması, örneklem büyüklüğünün yeterli olduğunu göstermektedir (Kaiser, 1974). Barlett küresellik testi sonucunun anlamlı olması ( $X^2_{(406)} = 67212.63$ ;  $p=0.00$ ) ise korelasyon matrisinin birim matristen farklı olduğunu göstermekte ve verinin faktör analizi için uygun olduğuna işaret etmektedir (Dziuban ve Shirkey, 1974).

Testin nitelik sayısına karar vermek için 1, 2, 3 ve 4 boyutlu modeller test edilmiş ve model uyum indeksleri Tablo 3'te sunulmuştur.

Tablo 3

*Testin Farklı Sayıda Boyutlardan Oluşan Formlarına İlişkin Uyum İndeksleri*

Boyut Sayısı	Madde Sayısı	RMSR	TLI Faktör Güvenirliği	RMSEA	BIC	Açıklanan Varyans
1	30	0.04	0.903	0.060	9955.91	0.42
2	29	0.03	0.940	0.048	4591.98	0.46
3	29	0.02	0.958	0.040	2193.18	0.48
4	29	0.02	0.963	0.038	1555.47	0.49

Tablo 3 incelendiğinde birinci faktörün toplam varyansın %42'sini açıkladığı ve model uyumlarının kabul edilebilir sınırlar içerisinde olduğu görülmektedir. Bu bulgular doğrultusunda yapının tek boyutlu olduğu öne sürülebilir ancak verinin faktör yapısına göre Q-matrisinin oluşturulabilmesi ve matematik testinin alt konu alanlarına bölünebilmesi için 2, 3 ve 4 boyutlu yapılar için tanımlanan AFA sonuçları yorumlanmıştır.

Matematik testinin 30 maddelik formuna 2, 3 ve 4 boyutlu yapılar için AFA yapılması sonucunda faktör yük değeri 0.30'un altında olan M062084 kodlu maddenin çıkarılmasına karar verilmiştir (Floyd ve Widaman, 1995). Çünkü bu durum, söz konusu maddenin ilgili boyutla düşük düzeyde ilişkili olduğuna işaret etmektedir. Tablo 3'te verilen uyum indeksleri incelendiğinde, 1 faktörlü yapıdan 4 faktörlü yapıya doğru, model-veri uyumunun iyileştiği ve en iyi model-veri uyumunu 3 ve 4 faktörlü modellerin sağladığı görülmektedir.

Matematik testinin 29 maddelik formu, AFA sonuçlarını doğrulamak üzere mirt (Chalmers vd., 2019) paketi ile analiz edilmiştir. mirt paketi ile açıklayıcı modellerin kestirilmesinde EM, MHRM ve QMCEM yöntemleri kullanılabilir. Bunlardan EM yöntemi 1-3 faktörlü yapılarda etkili olmasına rağmen boyut sayısının 3 ve üzeri olduğu durumlarda QMCEM veya MHRM yöntemlerinin kullanılması önerilmektedir (Chalmers vd., 2019). Bu doğrultuda 2, 3 ve 4 boyutlu olarak test edilen ÇB-MTK modelleri EM, QMCEM ve MHRM yöntemlerine göre analiz edilmiş ve model uyumlarını incelemek üzere mutlak uyum indeksleri Tablo 4'te özetlenmiştir.

Tablo 4

*Testin Farklı Sayıda Boyutlardan Oluşan Formlarına İlişkin Mutlak Model Uyumu İndeksleri*

Boyut Sayısı	Kestirim Yöntemi	M2	df	RMSEA	SRMSR	TLI	CFI
2	EM	1656.871	349	0.0202	0.0204	0.9943	0.9951
	QMCEM	1656.832	349	0.0202	0.0204	0.9943	0.9951
	MHRM	1659.312	349	0.0202	0.0204	0.9943	0.9951
3	EM	990.7818	322	0.0150	0.0187	0.9969	0.9975
	QMCEM	985.4691	322	0.0150	0.0178	0.9969	0.9975
	MHRM	986.6266	322	0.0150	0.0181	0.9969	0.9975
4	EM	822.7944	296	0.0139	0.0229	0.9973	0.9980
	QMCEM	858.1029	296	0.0144	0.0298	0.9971	0.9979
	MHRM	959.2871	296	0.0156	0.0326	0.9966	0.9975

Tablo 4'te verilen mutlak model uyum indeksleri, hata miktarlarını ve ilgili veri için modellerin doğruluğunu göstermektedir. Tabloya göre tüm modeller, model uyumu için gerekli olan "RMSEA<0.05, SRMSR<0.05, TLI ≥ 0.95, CFI ≥ 0.90" koşullarının hepsini karşılamaktadır. AFA sonuçları ile benzer şekilde, boyut sayısının 2'den 4'e doğru çıkması ile tabloda yer alan uyum istatistiklerinden M2,

RMSEA, SRMSR'in daha küçük, TLI ve CFI'nin ise daha büyük değer alması, model uyumunun arttığına işaret etmektedir.

Tablo 3 ve Tablo 4'e göre en iyi model-veri uyumunu 3 ve 4 faktörlü modellerin sağladığı görülmektedir. Bu nedenle, faktör sayısına karar vermek için 3 ve 4 faktörlü modellerin model veri uyumu  $\chi^2$ 'ler için ölçeklendirilmiş fark testi kullanılarak karşılaştırılmış ve Tablo 5'te sunulmuştur.

Tablo 5

*Testin 3 ve 4 Boyutlu Formlarına İlişkin  $\chi^2$ 'ler İçin Ölçeklendirilmiş Fark Testi Sonuçları*

Model1	Model2	$\chi^2$	p
3F-EM	4F-EM	27.4	
	4F-QMCEM	120.3	<0.05
	4F-MHRM	91.4	<0.05
3F-QMCEM	4F-EM	9.3	
	4F-QMCEM	102.2	<0.05
	4F-MHRM	73.3	<0.05
3F-MHRM	4F-EM	27.7	
	4F-QMCEM	120.6	<0.05
	4F-MHRM	91.7	<0.05

3F: 3 faktörlü model, 4F: 4 faktörlü model

Tablo 5 incelendiğinde, 3 boyutlu modeller ile QMCEM ve MHRM yöntemlerine göre kestirilen 4 boyutlu modeller arasında anlamlı bir fark olduğu görülmektedir. Bu bulgular doğrultusunda, 4 boyutlu yapının en iyi model veri uyumunu sergilediği söylenebilir. Ancak 4. boyutun tek başına toplam varyansın sadece %1'ini açıklaması ve tüm faktör yüklerinin 0.30'un altında olması (EK-A) göz önünde bulundurulduğunda, veri setinin 3 boyutlu olduğu kabul edilmiştir. Teste ilişkin faktör deseni ve faktör yük değerleri Tablo 6'da sunulmaktadır.

Tablo 6

*Testin 29 Maddelik Formunda Yer Alan Maddelere İlişkin Faktör Yük Değerleri*

Madde	1. boyut	2. boyut	3. boyut
M1	0.30		0.47
M2	0.62	0.31	
M3	0.54		0.54
M4	0.60	0.30	0.33
M5		0.49	
M6	0.63		0.34
M7	0.53		
M8	0.31	0.46	
M9	0.57	0.33	0.33
M10		0.63	
M11	0.62	0.40	0.37
M12	0.47		
M13	0.77		
M14	0.49		
M15			0.60
M16	0.66		
M17	0.46	0.33	
M18	0.45	0.32	
M19	0.72		
M20	0.67	0.37	
M21		0.53	0.36
M22		0.43	0.52
M23	0.52	0.44	0.54
M24	0.47	0.49	
M25	0.38	0.33	
M26	0.49	0.32	
M27	0.71		0.39
M28	0.51		
M29	0.57		

*RMSR= 0.02, Tucker Lewis Index of factoring reliability = 0.958, RMSEA= 0.04*

Tablo 6'da hesaplanan faktör yük değerlerinin; a) 1. alt ölçek için 0.30 ile 0.77 arasında, b) 2. alt ölçek için 0.30 ile 0.63 arasında ve c) 3. alt ölçek için 0.33 ile 0.60 arasında değiştiği görülmektedir. Ayrıca birinci faktör toplam varyansın %26'sını, ikinci faktör %12'sini ve üçüncü faktör %10'unu olmak üzere, 3 faktörlü yapı toplam varyansın %48'ini açıklamaktadır.

AFA sonucu elde edilen 3 faktörlü yapının doğruluğunu test etmek amacıyla doğrulayıcı faktör analizi (DFA) yapılmıştır. Bu doğrultuda R programında yazılan lavaan (Rosseel vd., 2019) paketi kullanılmıştır. Kategorik veriler yaklaşık olarak normal dağılım gösteriyormuş gibi gözükse de doğaları gereği kesiklidir ve normal dağılım sergilemezler (Muthén ve Kaplan, 1985). Özellikle sıralı kategori sayısının az olduğu durumlarda, çok değişkenli normallik varsayımı karşılanmamaktadır (Mîndrilă, 2010). Yapısal eşitlik modellerinde kullanılan kestirim yöntemlerinin seçimi ise önemli bir karardır. Nitekim, sayılıtların ihlal edilmesi model-veri uyumu,

parametre kestirimleri ve kestirimlerin standart hataları açısından yanlı sonuçlar üretebilir. Bu doğrultuda, çalışmada kullanılan veri yapısı dikkate alındığında kestirim yöntemi olarak, sıralı kategorik veriler için tasarlanan diyagonal ağırlıklandırılmış en küçük kareler (diagonally weighted least squares estimation: DWLS) yöntemi kullanılmıştır (Li, 2016). DFA sonucunda elde edilen uyum indeksleri  $\chi^2_{(352)} = 1618.437$  ( $p=0.00$ ), CFI= 0.981, TLI= 0.978, RMSEA= 0.020, SRMR= 0.015 şeklindedir. Bu değerlere bakıldığında tanımlanan ölçme modelinin, model uyumu için gerekli ölçütleri karşıladığı görülmektedir.

AFA ve DFA sonuçlarına göre oluşturulan faktör yapısı kullanılarak, madde ve nitelikler arasındaki ilişkiler belirlenmiştir (Q-matrisi oluşturulmuştur). Maddenin ölçülen boyutla ilişkili olması durumunda, ilgili niteliğin madde ile ölçüldüğünü belirtmek üzere Q-matrisi girdisi olarak 1, maddenin boyutla ilişkisi olmaması durumunda ise 0 ile temsil edilmiştir.

Q-matrisindeki nitelikler, simülasyon ve gerçek veri uygulamaları kapsamında, bir testin alt konu alanları olarak ele alınmıştır (Gerçek veri uygulaması için TIMSS matematik testinin alt konu alanları olarak düşünülmüştür). Araştırma kapsamındaki ÇB-MTK ve BTM temelli modellerin karşılaştırılabilmesi adına, BTM kapsamında nitelik olarak ele alınan yapılar, ÇB-MTK kapsamında boyut olarak düzenlenmiştir (Örneğin, Bradshaw ve Templin, 2014; de la Torre ve Karelitz, 2009; Sen ve Bradshaw 2017; Wang, 2009; Yakar, 2017). Bu doğrultuda, boyut ve nitelik ifadeleri birbirinin yerine kullanılacaktır.

### **Model Veri Uyumunun İncelenmesi**

TIMSS 2015 matematik testi 8. kitapçığının AFA sonucu elde edilen faktör desenine göre oluşturulan üç boyutlu yapının 1PL, 2PL ve 3PL ÇB-MTK modellerinden hangisine daha iyi uyum sağladığına dair kanıt sağlamak için mutlak ve göreceli model uyumları EM, QMCEM ve MHRM yöntemlerine göre hesaplanmış ve EK-B'de sunulmuştur. En yüksek hata QMCEM yöntemine göre kestirilen modellerde gözlenirken, EM ve MHRM yöntemlerine göre kestirilen model-veri uyumlarının benzer olduğu tespit edilmiştir. Ancak modellerin kestirim süresi, daha iyi optimizasyon yapması ve simülasyon koşullarında 3'ten fazla niteliğin kullanılması gibi gerekçelerden ötürü, gerçek veri parametrelerine dayalı olarak yapılan simülasyon çalışmalarında MHRM kestirim yöntemi kullanılmıştır.

Farklı modellerinin model veri uyumunun karşılaştırılması ve model seçiminde -2LL değerlerinin yanı sıra en yaygın olarak kullanılan iki ölçüt Akaike bilgi kriteri (AIC) ve Bayesian bilgi kriteridir (BIC). AIC (Akaike, 1974), veri seti için istatistiksel modellerin göreceli uyumlarının bir kestiricisidir. BIC (Schwarz, 1976) ise model uyumu ile model karmaşıklığı arasındaki dengeyi ölçmektedir (Stone, 1979). Daha küçük AIC ve BIC değerleri, daha iyi model uyumuna işaret etmektedir (Böckenholt ve Tsai, 2006; Dayton ve Macready, 2006; Huang ve Wang, 2014; Jansen ve Van der Mass, 1997; Li, Cohen, Kim ve Cho, 2009). Bu doğrultuda, veri setinin hangi modele daha iyi uyum sağladığına dair kanıt sağlamak için -2LL, AIC ve BIC değerleri kullanılmıştır.

Tablo 7’de MHRM yöntemine göre kestirilen mutlak ve göreceli model uyumu indeksleri yer almaktadır.

Tablo 7

*MHRM Yöntemine göre Kestirilen 1PL, 2PL ve 3PL Modellere İlişkin Mutlak ve Göreceli Model Uyumu İndeksleri*

Uyum İndeksleri	Modeller			
		1PL	2 PL	3 PL
Mutlak	M2	7541.618	1190.176	1020.222
	Sd	400	352	323
	RMSEA	0.0440	0.0161	0.0153
	SRMSR	0.0775	0.0191	0.0159
	TLI	0.9731	0.9964	0.9967
	CFI	0.9735	0.9969	0.9974
Göreceli	-2LL	141311.2	138589.3	138024.7
	AIC	282692.3	277344.6	276273.5
	BIC	282941.8	277936.2	277071.8
	Parametre Sayısı	35	83	112

Tablo 7’de MHRM yöntemine göre kestirilen mutlak model uyumu indekslerine göre 2PL ve 3PL modellerin 1PL modele göre daha iyi uyum sergilediği görülmektedir. Ki-kare fark testi sonucu ise 3PL modelin veriye daha iyi uyum sağladığını göstermektedir ( $\chi^2_{(29,0.01)} = 49.588$ ). Bu durumda verilerin 3PL model ile analiz edilmesinin önemli bir fark yaratacağı söylenebilir. Nitekim karşılaştırılan üç model arasında 3PL model, AIC ve BIC değerlerine göre de veriye diğer iki modelden daha iyi uyum göstermiştir. Bu doğrultuda simülasyon çalışması kapsamında, MHRM yöntemi ile kestirilen 3PL ÇB-MTK modelinin madde parametreleri kullanılacaktır.

## Q-Matrisinin Oluşturulması

Bilişsel tanı modellerinin en önemli bileşenlerinden Q-matrisinin doğru yapılandırılması büyük önem taşımaktadır. Q-matrisinin yanlış oluşturulması, model veri uyumsuzluğu, yanlış parametre kestirimleri ve bireylerin yanlış sınıflandırılması ile sonuçlanabilir (Kunina-Habenicht, Rupp ve Wilhelm, 2012; Rupp ve Templin, 2008a, 2008b). Bu doğrultuda başlangıç noktası olarak, AFA sonucu elde edilen faktör desenine göre maddenin ölçülen boyutla ilişkili olması durumunda ilgili madde ile ölçülen nitelikler arasında ilişki olduğunu göstermek üzere Q-matrisi girdisi olarak 1, maddenin boyutla ilişkili olmaması durumunda ise 0 ile temsil edilerek Q-matrisi oluşturulmuştur. Tablo 8'de Q-matrisi ve 3PL ÇB-MTK modeli madde parametreleri sunulmaktadır.

Tablo 8

*Q-Matrisi ve 3PL ÇB-MTK Modeli Madde Parametreleri*

Madde	Q-matrisi			3PLÇB-MTK Madde Parametreleri				
	$\alpha_1$	$\alpha_2$	$\alpha_3$	a1	a2	a3	b	c
M1	1	0	1	0.436	0.000	1.195	0.232	0.213
M2	1	1	0	1.679	0.239	0.000	-0.529	0.000
M3	1	0	1	1.327	0.000	1.438	0.972	0.088
M4	1	1	1	1.541	0.525	0.308	0.690	0.097
M5	0	1	0	0.000	2.384	0.000	-3.060	0.135
M6	1	0	1	2.006	0.000	0.627	1.105	0.147
M7	1	0	0	1.452	0.000	0.000	-0.270	0.034
M8	1	1	0	1.051	1.680	0.000	-1.499	0.312
M9	1	1	1	1.255	0.517	0.182	-0.242	0.000
M10	0	1	0	0.000	3.386	0.000	-3.854	0.100
M11	1	1	1	1.758	0.513	0.473	-0.705	0.013
M12	1	0	0	1.014	0.000	0.000	-1.298	0.000
M13	1	0	0	2.590	0.000	0.000	1.291	0.000
M14	1	0	0	1.982	0.000	0.000	-0.806	0.220
M15	0	0	1	0.000	0.000	2.483	1.153	0.270
M16	1	0	0	2.298	0.000	0.000	1.087	0.168
M17	1	1	0	1.491	0.358	0.000	-0.919	0.162
M18	1	1	0	1.601	0.590	0.000	-0.468	0.309
M19	1	0	0	2.265	0.000	0.000	-0.102	0.000
M20	1	1	0	2.264	0.056	0.000	-1.584	0.005
M21	0	1	1	0.000	1.454	1.148	-1.586	0.144
M22	0	1	1	0.000	1.289	1.685	-0.248	0.155
M23	1	1	1	1.063	0.890	1.104	-0.730	0.000
M24	1	1	0	1.074	0.936	0.000	-2.201	0.000
M25	1	1	0	1.413	1.090	0.000	-0.526	0.406
M26	1	1	0	0.873	0.531	0.000	-0.102	0.000
M27	1	0	1	2.052	0.000	0.452	0.969	0.000
M28	1	0	0	1.177	0.000	0.000	-1.921	0.000
M29	1	0	0	1.692	0.000	0.000	1.171	0.210

Madde ile yoklanan niteliğe sahip olan ve olmayan bireylerin ne derecede ayırt edildiğinin bir göstergesi olan madde ayırt edicilik parametresi, madde niteliği



olarak da adlandırılabilir (de la Torre ve Karelitz, 2009; Minchen, de la Torre ve Liu, 2017). Madde ayırt ediciliği, bireyleri performanslarına göre sınıflara ayırmada etkilidir. Bu doğrultuda, Tablo 9'a göre  $\alpha$  parametresi 0.50'nin altında olan maddelerin o boyutu ölçmediği varsayılarak, Q-matrisinde revizyonlara gidilmiş ve son hali Tablo 9'da sunulmuştur.

Tablo 9

*Q-Matrisinin Son Hali*

Madde	Nitelik			Madde	Nitelik		
	$\alpha_1$	$\alpha_2$	$\alpha_3$		$\alpha_1$	$\alpha_2$	$\alpha_3$
1	0	0	1	16	1	0	0
2	1	0	0	17	1	0	0
3	1	0	1	18	1	1	0
4	1	1	0	19	1	0	0
5	0	1	0	20	1	0	0
6	1	0	1	21	0	1	1
7	1	0	0	22	0	1	1
8	1	1	0	23	1	1	1
9	1	1	0	24	1	1	0
10	0	1	0	25	1	1	0
11	1	1	0	26	1	1	0
12	1	0	0	27	1	0	0
13	1	0	0	28	1	0	0
14	1	0	0	29	1	0	0
15	0	0	1				

Tablo 9 incelendiğinde karmaşık yapıdaki Q-matrisinde 1.niteliğin 23 (%79.31), 2. niteliğin 13 (%44.83) ve 3. niteliğin 7 (%24.14) madde ile ilişkilendirildiği görülmektedir. Sadece bir niteliği ölçen madde sayısı 16 (%55.17), iki niteliği ölçen madde sayısı 12 (%41.38) ve üç niteliği ölçen madde sayısı 1 (%3.45)'dir. Her bir maddenin doğru cevaplanması için ortalama 1.48 nitelik gerektiği söylenebilir. Maddelerin örtük sınıflara göre dağılımı ise Tablo 10'da görülmektedir.

Tablo 10

*Maddelerin Örtük Sınıflara göre Dağılımı*

Örtük Sınıf	Madde Sayısı	Yüzdesi
(1,0,0)	12	41.38
(1,1,0)	8	27.59
(1,0,1)	2	6.90
(0,1,0)	2	6.90
(0,1,1)	2	6.90
(0,0,1)	2	6.90
(1,1,1)	1	3.45

## Bilişsel Tanı Modellerinin Model Veri Uyumunun İncelenmesi

Model veri uyumu sağlanmadan yapılan analiz sonuçları anlamsız olacağı için model uyumunun incelenmesi, modelleme süreci için önem taşımaktadır (Rupp, Templin ve Henson, 2010). Nitekim uyarılma çalışmaları model veri uyumunu önemli ölçüde etkileyebilir (Gierl, Alves ve Majeau, 2010). Benzer şekilde, Q-matrisinin yanlış tanımlanması da zayıf model uyumuna neden olacaktır (Fay, 2018). Dolayısıyla model veri uyumu test edilerek, model veri uyumunun yanı sıra Q-matrisinin doğru tanımlanıp tanımlanmadığı da belirlenebilir. Bu doğrultuda, her bir modelin uyumunu değerlendirmek üzere mutlak uyum indeksleri ve iki modelin uyumunun birbirinden önemli ölçüde farklı olup olmadığını belirlemek üzere göreceli uyum indeksleri incelenmiştir.

Tablo 10 bilişsel tanı modellerinin mutlak model uyumu sonuçlarını göstermektedir. Tabloda yer alan uyum istatistiklerinden  $\max X^2$ , MADcor, SRMSR,  $100 \cdot \text{Res}$ , MADQ<sub>3</sub> ve MADaQ<sub>3</sub>'ün daha küçük değerler alması (0'a yakın değerler), daha iyi model uyumuna işaret etmektedir (Brown, 2015; Robitzsch vd., 2019). MADcor değeri kestirilen ve gözlenen ikili madde korelasyonları arasındaki mutlak sapmaların ortalamasını, MADres değeri ise ikili madde kovaryans artıklarının mutlak değerlerinin ortalamasını vermektedir (Li, Hunter ve Lei, 2016). SRMSR, 0 ile 1 arasında bir değer alabilir ve iyi model uyumu için  $\text{SRMSR} \leq 0.05$  olması gerekmektedir (Byrne, 1998; Diamantopoulos ve Siguaw, 2000; George ve Robitzsch, 2015; Maydeu-Olivares, 2013). Ancak 0.08'den küçük SRMSR değerleri de modelin veriye kabul edilebilir düzeyde uyum sağladığını göstermektedir (Hu ve Bentler, 1999).

Tablo 11

### *Bilişsel Tanı Modellerine İlişkin Mutlak Model Uyum İndeksleri*

Model	$\max X^2$	MADcor	SRMSR	$100 \cdot \text{Res}$	MADQ <sub>3</sub>	MADaQ <sub>3</sub>
DINA	198.272	0.049	0.056	1.090	0.033	0.024
DINO	199.581	0.047	0.053	1.032	0.036	0.025
A-CDM	213.121	0.039	0.046	0.864	0.026	0.024
LLM	205.321	0.037	0.044	0.823	0.025	0.024
R-RUM	280.200	0.050	0.059	1.092	0.030	0.023
G-DINA	202.481	0.038	0.045	0.836	0.025	0.024
HO-DINA	209.491	0.051	0.059	1.129	0.031	0.023

Tablo 11'e göre tüm BTM'lerin veriye kabul edilebilir düzeyde uyum sağladığı ancak A-CDM, LLM ve G-DINA modellerinin diğer modellere göre daha iyi model

uyumu gösterdiği söylenebilir. Bütün modeller mutlak uyum indekslerine göre veriye iyi uyum sergilediği için göreceli uyum indeksleri ile en iyi uyumu sergileyen bilişsel tanı modeli daha kesin olarak belirlenebilir (Yi, 2017).

Modellerin göreceli model uyumu sonuçlarının sunulduğu Tablo 12 incelendiğinde, yedi model arasından AIC, BIC ve CAIC değerlerine göre en iyi uyumu A-CDM, LLM ve G-DINA modellerinin sergilediği görülmektedir. HO-DINA modeli ise mutlak model uyumu sonuçlarıyla benzer şekilde, en kötü uyumu göstermiştir.

Tablo 12

*Bilişsel Tanı Modellerine İlişkin Göreceli Model Uyumu İndeksleri*

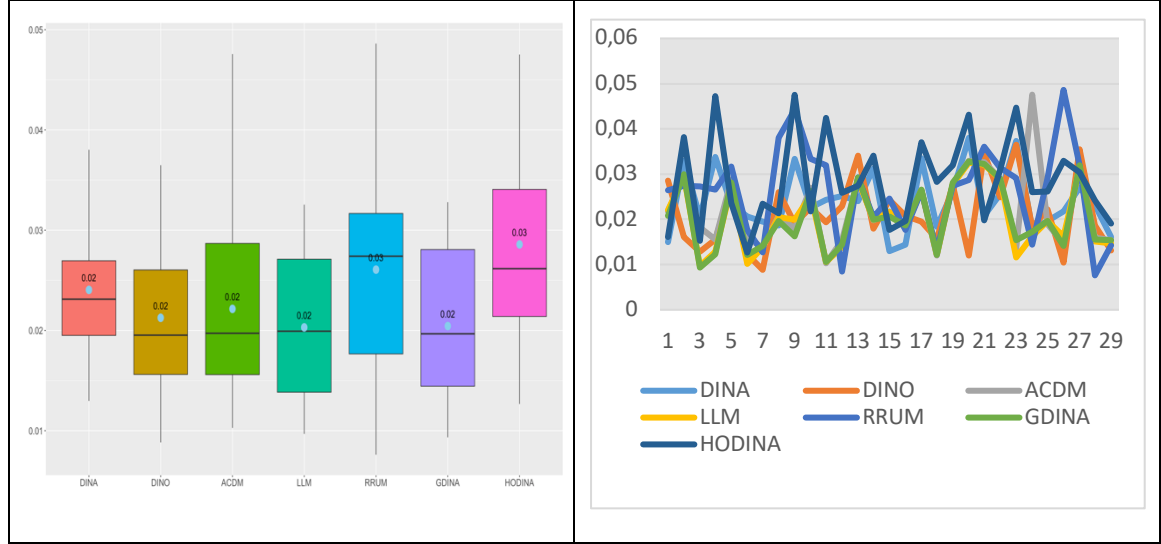
Model	-2LL	Parametre Sayısı	AIC	BIC	CAIC
DINA	142568.79	65	285267.579	285730.888	285795.888
DINO	142772.18	65	285674.357	286137.666	286202.666
A-CDM	141071.42	79	282300.844	282863.943	282942.943
LLM	140887.26	79	281932.514	282495.613	282574.613
R-RUM	142096.13	79	284350.261	284913.359	284992.359
G-DINA	140833.07	95	281856.136	282533.280	282628.280
HO-DINA	142779.60	64	285687.200	286143.381	286207.381

Tablo 12'ye göre modellerin -2LL değerleri A-CDM'den LLM ve G-DINA modeline doğru düşmektedir. Hangi modelin daha iyi uyum gösterdiğine karar verebilmek için LLM ve G-DINA modelleri karşılaştırılmıştır.

$$\chi^2 = -2LL_{LLM} - (-2LL_{G-DINA}) = 140887.26 - 140833.07 = 54.19$$

Hesaplanan ki-kare değerinin, 16 serbestlik derecesindeki tablo  $\chi^2$  değerinden (26.296) büyük olması, verilerin G-DINA modeline daha iyi uyum gösterdiğine işaret etmektedir.

Madde uyumunu değerlendirmek için RMSEA değerleri kullanılmıştır. Modellerin madde uyum istatistiği için hesaplanan RMSEA değerleri (EK-C) Şekil 1'de ve Tablo 13'te görülmektedir.



Şekil 1. Madde uyum istatistikleri.

Tablo 13

Bilişsel Tanı Modellerine İlişkin Madde Uyum İstatistikleri

MADDE	DINA	DINO	A-CDM	LLM	R-RUM	G-DINA	HO-DINA
1	0.0149	0.0286	0.0218	0.0220	0.0264	0.0208	0.0160
2	0.0329	0.0161	0.0298	0.0296	0.0274	0.0299	0.0382
3	0.0205	0.0128	0.0183	0.0097	0.0273	0.0093	0.0152
4	0.0337	0.0155	0.0156	0.0129	0.0265	0.0123	0.0473
5	0.0234	0.0268	0.0287	0.0271	0.0317	0.0281	0.0237
6	0.0206	0.0123	0.0159	0.0102	0.0176	0.0122	0.0127
7	0.0195	0.0088	0.0139	0.0138	0.0127	0.0141	0.0235
8	0.0186	0.0260	0.0208	0.0205	0.0381	0.0197	0.0214
9	0.0334	0.0190	0.0180	0.0199	0.0443	0.0163	0.0475
10	0.0226	0.0225	0.0271	0.0253	0.0334	0.0263	0.0218
11	0.0243	0.0194	0.0103	0.0106	0.0319	0.0106	0.0425
12	0.0251	0.0229	0.0153	0.0138	0.0085	0.0144	0.0258
13	0.0240	0.0341	0.0290	0.0287	0.0293	0.0292	0.0275
14	0.0312	0.0179	0.0197	0.0199	0.0205	0.0200	0.0341
15	0.0130	0.0244	0.0218	0.0216	0.0246	0.0206	0.0177
16	0.0144	0.0208	0.0184	0.0181	0.0177	0.0187	0.0196
17	0.0335	0.0195	0.0262	0.0265	0.0262	0.0266	0.0370
18	0.0184	0.0156	0.0120	0.0123	0.0132	0.0122	0.0283
19	0.0262	0.0276	0.0277	0.0277	0.0274	0.0282	0.0319
20	0.0380	0.0120	0.0329	0.0326	0.0287	0.0328	0.0432
21	0.0208	0.0346	0.0326	0.0322	0.0361	0.0322	0.0198
22	0.0255	0.0248	0.0290	0.0288	0.0314	0.0289	0.0309
23	0.0373	0.0365	0.0135	0.0116	0.0292	0.0154	0.0447
24	0.0179	0.0183	0.0476	0.0163	0.0143	0.0172	0.0260
25	0.0197	0.0221	0.0189	0.0197	0.0291	0.0197	0.0262
26	0.0218	0.0105	0.0147	0.0163	0.0486	0.0142	0.0330
27	0.0270	0.0355	0.0317	0.0316	0.0320	0.0320	0.0303
28	0.0231	0.0186	0.0159	0.0151	0.0076	0.0155	0.0242
29	0.0161	0.0131	0.0153	0.0148	0.0143	0.0154	0.0191

Madde uyumu istatistiği olarak RMSEA, farklı örtük sınıflardaki katılımcıların doğru cevap için model tarafından kestirilen madde cevap olasılıklarını her bir örtük sınıftaki katılımcıların oranı ile ağırlıklandırılan cevaplarla, gözlenen doğru cevapları

karşılaştırmaktadır (Kunina-Habenicht, Rupp ve Wilhelm, 2009). RMSEA değerinin 0.05'ten küçük olması iyi, 0.05 ile 0.10 arasında bir değer alması orta, 0.10'dan büyük olması ise zayıf uyuma işaret etmektedir (George vd., 2016).

Madde/maddelerin yeterli uyumu sağlamaması durumunda, ilgili maddenin uyarılma çalışması kapsamında çıkarılması düşünülebilir (Liu, Huggins-Manley ve Bulut, 2017). Ancak çalışmada kullanılan veri seti için tüm maddelerin RMSEA değeri 0.05'in altında olduğu için maddelerin hepsi analize dahil edilmiştir. Model veri uyumu analizi sonuçlarıyla benzer şekilde, LLM ve G-DINA modelleri en iyi madde uyumunu göstermiştir. Bu doğrultuda, ÇB-MTK kapsamında kestirilen sürekli yetenek puanlarını lojistik regresyon yöntemi ile kesikli hale getirirken, G-DINA modeline göre kestirilen sınıflamaların (başarılı-başarısız) kullanılmasına karar verilmiştir.

### **Sınıflama Doğruluğu ve Tutarlığının Değerlendirilmesi**

Cui, Gierl ve Chang (2012) sınıflama doğruluğunu, gözlenen madde cevap örüntülerine göre bireyin ait olduğu örtük sınıfının, gerçek örtük sınıfı ile uyuma derecesi olarak açıklamıştır. Sınıflama tutarlığı ise iki bağımsız uygulama ya da testin paralel formları arasındaki sınıflamaların, aynı gizil sınıfa ait olma durumudur (Cui, Gierl ve Chang, 2012). Örneğin, iki ayrı ya da paralel uygulamada,  $\alpha$  nitelik örüntüsü için bireyin cevap örüntüleri  $X_1$  ve  $X_2$  olsun.  $X_1$  ve  $X_2$ 'nin aynı örtük sınıfa düşmesi, tutarlı sınıflamalar yapıldığının bir göstergesidir. Çalışmada kullanılacak BTM'lerin sınıflama tutarlılığı ve doğruluğu, Maksimum Olabilirlik Kestirimi (MOK; Maximum Likelihood Estimation: MLE) ve Bayesci yaklaşımlardan Maksimum Sonsal Dağılım (MSD; Maximum a Posteriori: MAP) yöntemlerine göre CDM (Robitzsch vd., 2019) R paketi ile incelenmiş ve Tablo 14'te sunulmuştur.

Tablo 14

*Sınıflama Doğruluğu ve Tutarlılığının Bilişsel Tanı Modellerine göre Dağılımı*

MODEL		Nitelik	SD	ST	MODEL		Nitelik	SD	ST
DINA	MLE	$\alpha_1$	0.960	0.923	DINO	MLE	$\alpha_1$	0.960	0.924
		$\alpha_2$	0.856	0.771			$\alpha_2$	0.882	0.807
		$\alpha_3$	0.857	0.765			$\alpha_3$	0.870	0.779
	MAP	$\alpha_1$	0.970	0.945		MAP	$\alpha_1$	0.976	0.955
		$\alpha_2$	0.914	0.892			$\alpha_2$	0.904	0.847
		$\alpha_3$	0.890	0.892			$\alpha_3$	0.925	0.882
A-CDM	MLE		SD	ST	LLM	MLE		SD	ST
		$\alpha_1$	0.963	0.930			$\alpha_1$	0.961	0.926
		$\alpha_2$	0.895	0.815			$\alpha_2$	0.895	0.817
	MAP	$\alpha_3$	0.871	0.778		$\alpha_3$	0.864	0.767	
		$\alpha_1$	0.972	0.948		MAP	$\alpha_1$	0.972	0.948
		$\alpha_2$	0.928	0.880			$\alpha_2$	0.939	0.898
$\alpha_3$	0.919	0.872	$\alpha_3$	0.920	0.869				
R-RUM	MLE		SD	ST	G-DINA	MLE		SD	ST
		$\alpha_1$	0.967	0.937			$\alpha_1$	0.963	0.929
		$\alpha_2$	0.906	0.831			$\alpha_2$	0.888	0.805
	MAP	$\alpha_3$	0.879	0.789		$\alpha_3$	0.882	0.795	
		$\alpha_1$	0.977	0.956		MAP	$\alpha_1$	0.973	0.949
		$\alpha_2$	0.984	0.971			$\alpha_2$	0.933	0.891
$\alpha_3$	0.925	0.882	$\alpha_3$	0.920	0.875				
HO-DINA	MLE		SD	ST	HO-DINA	MAP		SD	ST
		$\alpha_1$	0.958	0.937			$\alpha_1$	0.970	0.961
		$\alpha_2$	0.844	0.769			$\alpha_2$	0.952	0.942
	MAP	$\alpha_3$	0.825	0.744		$\alpha_3$	0.922	0.938	

SD: sınıflama doğruluğu; ST: sınıflama tutarlığı

Tablo 14'e göre;

- a)  $\alpha_1$  niteliği için MLE kestirim yöntemine göre sınıflama doğruluğunun 0.958-0.967, sınıflama tutarlığının ise 0.923-0.937; MAP kestirim yöntemine göre sınıflama doğruluğunun 0.970-0.977, sınıflama tutarlığının ise 0.945-0.961;
- b)  $\alpha_2$  niteliği için MLE kestirim yöntemine göre sınıflama doğruluğunun 0.844-0.906, sınıflama tutarlığının ise 0.769-0.831; MAP kestirim yöntemine göre sınıflama doğruluğunun 0.904-0.984, sınıflama tutarlığının ise 0.847-0.971;
- c)  $\alpha_3$  niteliği için MLE kestirim yöntemine göre sınıflama doğruluğunun 0.825-0.882, sınıflama tutarlığının ise 0.744-0.795; MAP kestirim yöntemine göre sınıflama doğruluğunun 0.890-0.925, sınıflama tutarlığının ise 0.869-0.938 arasında değiştiği görülmektedir.

Bu sonuçlar, tüm nitelikler göz önünde bulundurulduğunda, rastgele seçilen bir bireyin doğru örtük sınıfına yerleştirilme olasılığının MLE'ye göre yaklaşık %83

ile %97 arasında ve MAP'a göre %89 ile %98 arasında deęiřtięini, iki uygulama ya da paralel uygulamalara gre aynı rtk sınıfa dřme olasılıęının ise MLE'ye gre %74 ile %94 arasında, MAP'a gre ise %85 ile %97 arasında deęiřtięini gstermektedir.

Yukarıda sunulan bu bulgular doęrultusunda, TIMSS 2015 uygulaması matematik testi kitapık 8'den elde edilen verilere ynelik bir uyarlama alıřmasının (ok boyutlu MTK'dan BTM'ye uyarlama) yapılabilceęi sylenebilir.

### **Simlasyon Kořulları**

alıřmanın verileri 3PL B-MTK modeline gre retilmiř ve verilerin analizinde B-MTK ve BTM temelli yntemler kullanılmıřtır. Arařtırma kapsamındaki modellere gre kestirilen birey parametrelerinin sınıflama doęruluęunun karřılařtırılması zerine daha ayrıntılı ve destekleyici kanıtlar sunmak zere farklı kořullarda simlasyon alıřmaları yapılmıřtır.

Simlasyon kořulları: a) testteki madde sayısı, b) nitelikler arası korelasyon, 3) boyut/nitelik sayısı ve 4) Q-matrisinin yapısı řeklinindedir. rnekleme byklę tm kořullar iin 10000'e sabitlenmiř ve her bir kořul 100 tekrar ile yapılmıřtır.

**Test Uzunluęu:** TIMSS 2015 uygulaması matematik kitapıkları incelendięinde, matematik testlerinin ikiřer bloktan oluřtuęu ve her bir kitapıktaki soru sayılarının 28 ile 35 arasında deęiřtięi, ortalama soru sayısının ise 30.71 olduęu grlmektedir. Alan yazındaki alıřmalar incelendięinde ise, genel olarak, kısa testlerin 15 ve uzun testlerin 30 maddeden oluřtuęu grlmektedir (Chen, de la Torre ve Zhang, 2013; Minchen, de la Torre ve Liu, 2017; Minchen ve de la Torre, 2018; Terzi ve de la Torre, 2018).

Testteki madde sayısı ve llen nitelik sayısı birbiriyle iliřkilidir. Tm madde-nitelik sınıflarının istatistiksel olarak iyi tanımlanabilmesi iin  $k$  nitelik sayısı olmak zere, en az  $2^k-1$  maddeye ihtiya duyulmaktadır (Fay, 2018). TIMSS matematik testlerindeki soru sayısı, ilgili alan yazını ve analiz sresinin uzunluęu dikkate alındıęında bu arařtırmada madde sayısı, kısa testler iin 15 ve uzun testler iin 30 olarak belirlenmiřtir.

**Nitelikler Arası Korelasyon:** Nitelikler arasında 3 korelasyon dzeyi kullanılmıřtır: 0.1, 0.4 ve 0.7 sırasıyla dřk, orta ve yksek korelasyonu temsil etmektedir.

Bulgular üzerindeki farklı korelasyon düzeylerinin karıştırıcı etkisini kontrol etmek için nitelikler arasındaki korelasyonlar aynı alınmıştır. Yapılan çalışmalar, nitelikler arasındaki korelasyonların parametre kestirimi üzerinde etkisinin olduğunu göstermiştir (de la Torre ve Patz, 2005; Shin, 2007; Shin, Ansley, Tsai ve Mao, 2005). Ancak niteliklerin birbirinden yeterince ayrık olabilmesi için nitelikler arası korelasyonun 0.85'ten daha düşük olması gerekmektedir (Sinharay, 2010). Nitelikler arasındaki yüksek korelasyon, aslında tek niteliği ölçen bir testin birden fazla niteliği ölçüyormuş gibi ele alınmasına neden olacaktır. Bu nedenle, nitelikler arası korelasyonun 1 ya da 1'e yakın bir değer alması, testi alan kişilerin ağırlıklı olarak başarılı ya da başarısız gruplarından birine düşmesine neden olacak ve bireyler hakkında yanıltıcı bilgiler sağlayacaktır (Liu, Huggins-Manley ve Bulut, 2017). Eğitim alanında yapılan BTM araştırmaları incelendiğinde ise nitelikler arası korelasyonların genel olarak 0.20 ile 0.90 arasında değiştiği görülmektedir (Bradshaw ve Templin, 2014; Cui, Gierl ve Chang, 2012; Henson, Templin ve Douglas, 2007; Henson, Roussos, Douglas ve He, 2008; Kunina-Habenicht, Rupp ve Wilhelm, 2012; Liu, 2017; Liu, Huggins-Manley ve Bradshaw, 2017; Wang, 2009). Nitelikler arasında daima orta ya da yüksek düzeyde korelasyon olduğu için (Sinharay vd., 2010), simülasyon koşullarında en düşük korelasyon düzeyi 0.0 yerine 0.1 olarak seçilmiştir. Yüksek korelasyon düzeyi olan 0.70 değeri ise TIMSS matematik testindeki alt konu alanları arasındaki korelasyonların 0.70 civarında olması dikkate alınarak belirlenmiştir. Nitekim Sinharay, Puhan ve Haberman (2011) eğitim alanında yapılan araştırmalarda, 0.70 korelasyon düzeyinin makul olduğunu belirtmişlerdir. Böylece, araştırmadan elde edilen sonuçların uygulayıcılar için bilgilendirici olabileceğini söylenebilir (Liu, Huggins-Manley ve Bradshaw, 2017).

**Nitelik Sayısı:** BTM ile gerçekleştirilen simülasyon çalışmaları incelendiğinde, nitelik sayılarının genel olarak 3 ile 8 arasında değiştiği görülmektedir (Cheng, 2009; Chen, de la Torre ve Zhang, 2013; Chen ve de la Torre, 2018; Choi, 2010; de la Torre, 2009; de la Torre ve Douglas, 2004, 2008; de la Torre ve Lee, 2013; Gao, Miller ve Liu, 2017; Huebner ve Wang, 2011; Kunina-Habenicht, Rupp ve Wilhelm, 2012). Sessoms ve Henson (2018) yaptıkları derleme çalışmalarında, BTM alanında yapılan 36 çalışmayı incelemiş ve ortalama nitelik sayısının 8 olduğunu belirtmiştir. Ancak çok sayıda niteliğin kestirilmesi ve yorumlanması zor olduğu için toplam nitelik sayısının oldukça küçük tutulması gerekmektedir (de la Torre ve



Douglas, 2004; Embretson ve Yang, 2013; Liu, Huggins-Manley ve Bulut, 2017; Xu ve Zhang, 2016). Bu doğrultuda, çalışma kapsamında manipüle edilen değişkenler ve analiz süreleri dikkate alındığında nitelik sayısı 3 ve 6 olarak belirlenmiştir. ÇB-MTK ve BTM kapsamında kestirilen birey parametrelerini karşılaştırabilmek adına BTM'ler için 3 nitelik olarak tasarlanan yapı, ÇB-MTK çerçevesinde 3 boyut olarak ele alınmıştır. 6 nitelikli yapı için de benzer işlem yapılmıştır.

**Q-Matrisinin Yapısı:** Bu çalışmada 2 farklı Q-matrisi yapısı kullanılmıştır. Bunlardan ilki, AFA sonucu elde edilen faktör desenine göre oluşturulan ve doğrulayıcı ÇB-MTK modeli madde ayırt edicilik parametrelerine göre revize edilerek son hali Tablo 9'da verilen Q-matrisindeki nitelik sayı ve profillerinin dağılımı dikkate alınarak oluşturulmuştur. İkinci Q-matrisi ise her bir nitelik ve nitelik profili eşit sayıda ölçülecek şekilde tanımlanmıştır. Gerçek veriye dayalı olarak geliştirilen Q-matrisinde niteliklerin madde ile ölçülme oranları değişkenlik göstermektedir. Dolayısıyla,  $2^k-1$  nitelik profilinden her birinin yeterli sayıda ölçülmesi adına, niteliklerin ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi yapısı tanımlanmıştır. Her iki Q-matrisi yapısında (EK-Ç), bir maddeyi doğru cevaplamak için gereken maksimum nitelik sayısı 2 olarak belirlenmiştir. Bu doğrultuda,  $(K_j^*)_{\max}=1$  ve 2 değerleri sırasıyla, basit ve daha karmaşık yapıdaki Q-matrislerini temsil etmektedir. 6 niteliğin ölçüldüğü Q-matrisleri ise 3 niteliğin ölçüldüğü Q-matrislerinin iki defa tekrarlanması ile elde edilmiştir.

**Örneklem Büyüklüğü:** Sessoms ve Henson (2018) yaptıkları derleme çalışmasında, inceledikleri 36 BTM çalışmasının yaklaşık %61'inde 1000 ve üzeri, %31'inde 1000 ile 2000 arasında ve %30'unda 2000'den büyük örneklem büyüklüklerinin kullanıldığını belirtmiştir. Henson (2009) da BTM uygulamalarının büyük örneklem gerektirdiğini belirtmiştir. Bu doğrultuda, çalışmada örneklem büyüklüğü 10000 olarak belirlenmiş ve sabitlenmiştir.

**Kestirim Yöntemi:** ÇB-MTK kapsamında bireylerin sürekli yetenek puanlarını elde etmek üzere yapılan kestirimlerde MAP yöntemi kullanılırken, BTM'lerde marjinal olasılıklar kullanılmış ve 0.001 yakınsama değerine ulaşıldığında analizler sonlandırılmıştır.

**Kesme Puanının Belirlenmesi:** Tanılayıcı ve tanılayıcı olmayan modeller arasındaki temel farklılık; tanılayıcı modellerin birden çok kategorik özellik belirleyip

bireyleri her bir kategori için sınıflandırması, tanılayıcı olmayan modellerin ise bir ya da daha fazla sürekli özellik belirleyip, bireylere sürekli bir ölçekte puan vermesinden kaynaklanmaktadır (Liu, Huggins-Manley ve Bulut, 2017). Bu nedenle, ÇB-MTK çerçevesinde elde edilen sürekli yetenek puanlarına göre bireyleri kategorik sınıflara ayırabilmek için bir kesme noktasının belirlenmesine ihtiyaç vardır. BTM'nin ÇB-MTK verisine uyarıldığı çalışmalar incelendiğinde, bireylerin sürekli yetenek puanlarını kesikli hale getirmek için madde güçlük parametrelerini kesme puanı olarak kullandıkları görülmektedir (de la Torre ve Karelitz, 2009; Yakar, 2017). Ancak bu kesme noktalarını sadece ÇB-MTK kapsamında yapılan sınıflamalarda kullanmış ve BTM çerçevesinde yapılan sınıflamalara herhangi bir müdahalede bulunmamışlardır. Wang, 2009 yılında yaptığı doktora tez çalışmasında ise lojistik regresyon yöntemini kullanarak, BTM ve ÇB-MTK modelleri ile yapılan kestirimlerin uyumunu maksimum yapmaya çalışmış ve 3PL ÇB-MTK çerçevesinde kestirilen sürekli yetenek puanlarını iki kategorili (1-0) hale getirmek üzere, lojistik regresyon yöntemini kullanmıştır. Bu çalışma kapsamında da kesme puanının belirlenmesinde lojistik regresyon yöntemi kullanılmıştır.

BTM çalışmalarında nitelik profilleri, bireylerin her bir nitelik için kestirilen sonsal olasılıklarına bir kesme puanının uygulanması ile elde edilmektedir (de la Torre, 2009). Bu olasılık 0 ile 1 arasında bir değer almaktadır ve genel olarak, kestirilen marjinal olasılık değerinin 0.50 ve üzerinde olması durumunda birey 1 (başarılı) grubuna, aksi halde 0 (başarısız) grubuna yerleştirilmektedir. Bu çalışmada ise ülkemizdeki büyük ölçekli sınavlar (ALES, KPSS), üniversite akademik not ortalamaları, üniversitelerin yabancı dil muafiyet şartları, yüksek lisans ve doktora başvuru koşulları gibi faktörler göz önüne alındığında kesme puanı 0.70 olarak belirlenmiştir.

## **Verilerin Üretilmesi ve Analizi**

Araştırmada kullanılan simülasyon verileri TIMSS 2015 matematik testinin 8. kitapçığının parametreleri temel alınarak üretilmiştir.

TIMSS 2015 matematik testi kitapçık 8'in madde parametreleri mirt (Chalmers vd., 2019) paketi ile MHRM kestirim yöntemine göre 3PL ÇB-MTK model ile analiz edilmiş ve Tablo 15'te sunulmuştur.

Tablo 15

*TIMSS 2015 Matematik Sınavının Alt Testlerine ait Madde Parametreleri*

Alt test	Parametre					
	a		b		c	
	En Küçük	En Büyük	En Küçük	En Büyük	En Küçük	En Büyük
1. alt test	0.901	2.544	-2.186	1.284	0.000	0.407
2. alt test	0.440	3.430	-3.885	0.710	0.000	0.407
3. alt test	0.530	2.606	-1.523	1.109	0.000	0.327

TIMSS 2015 matematik verisine ait a parametreleri incelendiğinde ranjının oldukça geniş olduğu görülmektedir (0.440-3.430). BTM çerçevesinde madde ayırt ediciliği, her bir niteliğin sınıflandırılmasında maddenin sağladığı bilgi miktarını ölçmektedir (Wang, 2009). Yapılan çalışmalar madde ayırt ediciliğinin, doğru sınıflama oranı ve hata üzerinde etkili olduğunu göstermektedir (de la Torre ve Karelitz, 2009; Ma ve de la Torre, 2016; Sen ve Bradshaw; 2017; Yakar, 2017). Bu nedenle, Tablo 14'te yer alan madde parametreleri ve alan yazında yer alan simülasyon çalışmaları göz önünde bulundurularak bu çalışmada kullanılacak verilerin a parametresi ranjı [0.6,2.5] arasında olacak şekilde log-normal dağılımdan,  $a \sim \ln N(1.5, 0.8)$ ; b parametresi ranjı [-2,1.5] arasında olacak şekilde normal dağılımdan,  $b \sim N(0, 1)$ ; c parametresi ise ranjı [0,0.30] arasında olacak şekilde beta dağılımdan,  $c \sim \text{Beta}(20, 100)$  üretilmiştir. Gerçek yetenek parametreleri ise ortalaması 0, varyansı araştırma koşullarında belirlenen varyans-kovaryans matrisine göre çok değişkenli normal dağılımdan üretilmiştir,  $\theta_i \sim \text{MVN}(0, \Sigma)$ .

ÇB-MTK ve BTM çerçevesinde birey parametrelerinin kestirilmesi için sırasıyla mirt (Chalmers vd., 2019) ve GDINA (Ma vd., 2019) R paketleri kullanılmıştır. İki kategorili madde yanıtlarının üretilmesinde mirt paketi kullanılırken, birey parametrelerinin gerçek değeri MASS (Ripley vd., 2019) R paketi ile üretilmiştir. Gerçek verinin bilişsel tanı modelleri ile uyumu GDINA paketi ile analiz edilmiş ve analiz sonuçları CDM (Robitzsch, Kiefer, George ve Uenlue, 2019) R paketi ile doğrulanmıştır (EK-C ve EK-D). Belirlenen simülasyon koşullarının her biri 100 tekrar ile gerçekleştirilmiş ve her bir koşulun analizi 28 ile 40 saat arasında bir zaman almıştır.

### Değerlendirme Kriteri

Tanılayıcı testler bireyleri her bir içerik alanındaki yeterliliklerine göre başarılı-başarısız şeklinde sınıflayarak, yeterlilikleri hakkında daha detaylı bilgi

sağlamaktadır. Kestirilen birey parametrelerinin sınıflama doğruluğu, bireyin gerçek örtük sınıfına doğru bir şekilde yerleştirilmesinin bir ölçüsüdür (George vd., 2016). Bir başka ifadeyle, gözlenen ve gerçek sınıflamalar arasındaki uyum derecesidir. Bu doğrultuda, çalışma kapsamındaki yöntemlerin bireyleri sınıflamadaki başarılarını değerlendirmek için ortalama hata kareler kökü (root mean square error: RMSE) ve yanlışlık değerleri hesaplanmıştır. RMSE değeri kestirilen ve gerçek sınıflamalar arasındaki hatalarının toplam varyansını, yanlışlık ise kestirilen ve gerçek sınıflamalar arasındaki ortalama farkını ifade etmektedir (Steiger, 1990). Kestirilen ve gerçek sınıflamalar arasındaki tutarsızlığı temsil eden bu iki istatistiğe ait matematiksel formüller aşağıdaki gibidir:

$$RMSE = \sqrt{\frac{\sum_{r=1}^R \sum_{i=1}^N \sum_{k=1}^K (\hat{\alpha}_{ik}^{(r)} - \alpha_{ik}^{(r)})^2}{N \times R \times K}}$$

ve

$$Yanlışlık = \frac{\sum_{r=1}^R \sum_{i=1}^N \sum_{k=1}^K (\hat{\alpha}_{ik}^{(r)} - \alpha_{ik}^{(r)})}{N \times R \times K}$$

Yukarıdaki eşitlikte verilen  $R$  replikasyon sayısını,  $N$  örneklem büyüklüğünü,  $K$  nitelik sayısını,  $\hat{\alpha}_{ik}^{(r)}$   $r$ . replikasyon için  $i$  bireyinin  $k$ . nitelik için kestirilen geçme durumu ve  $\alpha_{ik}^{(r)}$   $r$ . replikasyon için  $i$  bireyinin  $k$ . nitelik için gerçek geçme durumu temsil etmektedir.

RMSE ve yanlışlık değerlerinin yanı sıra kestirilen birey parametrelerinin sınıflama doğruluğunu değerlendirmek üzere her bir nitelik (NDSO) ve nitelik profili (NPDSO) için doğru sınıflama oranları aşağıdaki gibi hesaplanmıştır:

$$NDSO = \frac{\sum_{r=1}^R \sum_{i=1}^N \sum_{k=1}^K I[\alpha_{ik}^{(r)} = \hat{\alpha}_{ik}^{(r)}]}{R \times N \times K}$$

ve

$$NPDSO = \frac{\sum_{r=1}^R \sum_{i=1}^N \prod_{k=1}^K I[\alpha_{ik}^{(r)} = \hat{\alpha}_{ik}^{(r)}]}{R \times N}$$

Formülde yer alan  $I(.)$  gösterge fonksiyonudur ve  $r$ . replikasyon için  $\alpha_{ik}^{(r)}$  ile  $\hat{\alpha}_{ik}^{(r)}$ 'nin birbirine eşit olması durumunda 1'e, aksi halde 0'a eşittir.

## **Bölüm 4**

### **Bulgular ve Yorumlar**

Bu bölümde, çalışmanın her bir alt problemlerine ilişkin bulgulara ve bu bulgulara yönelik yorumlara yer verilmiştir.

#### **Birinci Alt Probleme İlişkin Bulgular ve Yorumlar**

Araştırmanın birinci alt problemi kapsamında, ÇB-MTK verisinin 3PL ÇB-MTK ve BTM modelleri ile analizinden elde edilen birey parametrelerine ilişkin hata (RMSE) ve yanlılık (bias) değerleri testteki madde sayısı (15 ve 30), nitelikler arası korelasyon (0.1, 0.4 ve 0.7) ve Q-matrisi yapısına (gerçek verinin faktör yapısına göre oluşturulan Q-matrisi ve nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi) göre 3 nitelikten (alt konu alanı) oluşan tek boyutlu veri setleri için incelenmiştir.

#### **Üç Nitelikten Oluşan Tek Boyutlu Veri Setleri için Birey Parametrelerine İlişkin RMSE Değerleri**

Testteki madde sayısı, nitelikler arasındaki korelasyon ve Q-matrisi yapısı koşullarının birey parametre kestirimlerine ilişkin RMSE değerleri üzerindeki etkisi Tablo 16'da sunulmaktadır.

Birey parametrelerine ilişkin hata değerlerindeki değişim, her bir model için Şekil 2-Şekil 9 üzerinde gösterilmiştir. Hata değerlerindeki değişimler 0.35 ile 0.55 değerleri arasında ölçeklenmiş olmak üzere x-eksenine, boyutlar arası korelasyon değerleri (0.1, 0.4, 0.7) ise y-eksenine yerleştirilmiştir. Grafiklerde  $\alpha_1$ ,  $\alpha_2$  ve  $\alpha_3$  nitelikleri aşağıdan yukarı doğru sıralanmış ve her niteliğin altında Q-matrisi yapısı (E ve G) ile test uzunluğu (15 ve 30) değişkenleri yer almıştır. Her bir nitelikteki değişim madde sayısı, nitelikler arası korelasyon ve Q-matrisi yapısı koşullarına göre ayrı ayrı ele alınmıştır.

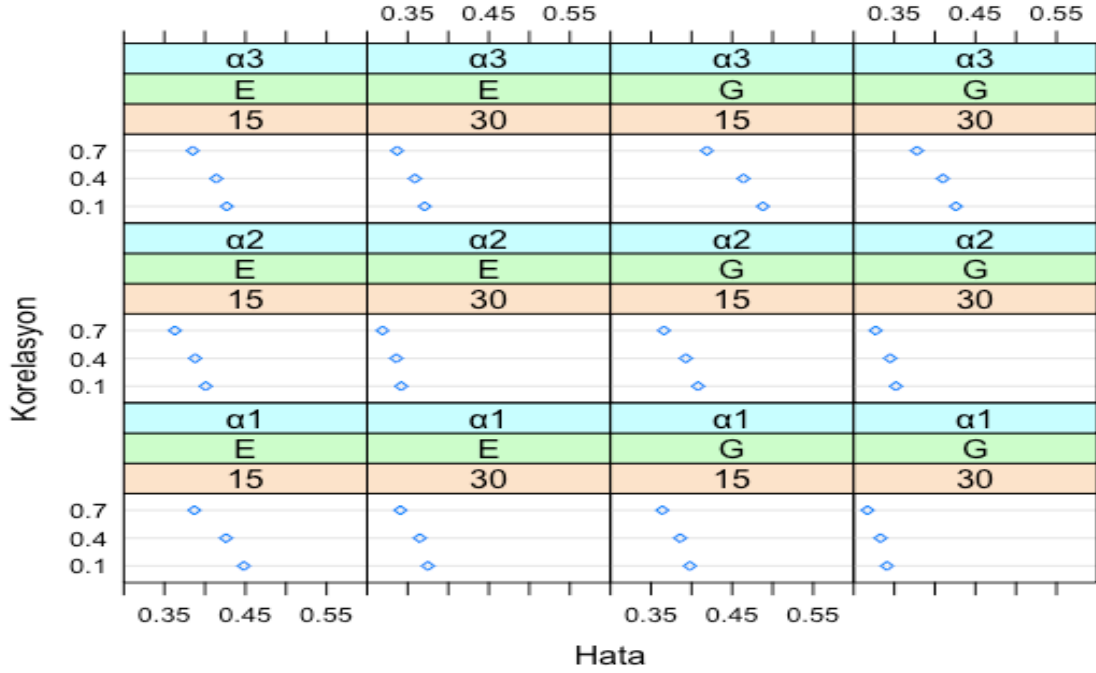
Tablo 16

Üç Nitelikten Oluşan Tek Boyutlu Veri Setleri için Birey Parametrelerine İlişkin Hata Değerleri

MS	Q-Matrisi	Korelasyon	ÇB-MTK			DINA			DINO			A-CDM		
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$
15	G	0.1	0.398	0.408	0.488	0.413	0.590	0.535	0.474	0.433	0.568	0.423	0.476	0.497
		0.4	0.386	0.393	0.464	0.401	0.527	0.492	0.454	0.421	0.511	0.411	0.458	0.470
		0.7	0.364	0.366	0.419	0.379	0.478	0.432	0.419	0.417	0.447	0.393	0.441	0.422
	E	0.1	0.448	0.401	0.427	0.490	0.530	0.447	0.577	0.423	0.470	0.469	0.460	0.434
		0.4	0.426	0.388	0.414	0.453	0.503	0.430	0.527	0.424	0.446	0.444	0.449	0.420
		0.7	0.387	0.363	0.385	0.404	0.464	0.395	0.448	0.418	0.404	0.411	0.434	0.392
30	G	0.1	0.341	0.352	0.426	0.362	0.561	0.486	0.437	0.391	0.504	0.366	0.455	0.437
		0.4	0.333	0.345	0.410	0.354	0.525	0.454	0.423	0.393	0.471	0.359	0.448	0.418
		0.7	0.317	0.327	0.378	0.335	0.486	0.404	0.390	0.400	0.409	0.345	0.438	0.383
	E	0.1	0.375	0.342	0.371	0.411	0.548	0.426	0.492	0.379	0.446	0.400	0.444	0.380
		0.4	0.365	0.336	0.359	0.396	0.522	0.404	0.464	0.381	0.420	0.387	0.440	0.367
		0.7	0.341	0.319	0.337	0.366	0.483	0.366	0.414	0.390	0.375	0.365	0.431	0.344
MS	Q-Matrisi	Korelasyon	LLM			R-RUM			G-DINA			HODINA		
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$
15	G	0.1	0.431	0.452	0.498	0.418	0.510	0.498	0.427	0.463	0.494	0.413	0.592	0.534
		0.4	0.419	0.433	0.474	0.407	0.487	0.468	0.414	0.444	0.469	0.401	0.528	0.492
		0.7	0.399	0.416	0.425	0.391	0.465	0.421	0.395	0.427	0.421	0.378	0.481	0.431
	E	0.1	0.491	0.438	0.438	0.459	0.487	0.434	0.476	0.448	0.435	0.491	0.531	0.447
		0.4	0.464	0.432	0.423	0.434	0.467	0.419	0.451	0.441	0.421	0.454	0.504	0.430
		0.7	0.433	0.423	0.394	0.399	0.440	0.392	0.420	0.430	0.393	0.402	0.466	0.395
30	G	0.1	0.375	0.433	0.438	0.359	0.485	0.441	0.369	0.444	0.435	0.362	0.563	0.486
		0.4	0.369	0.424	0.418	0.352	0.477	0.422	0.363	0.436	0.415	0.354	0.526	0.455
		0.7	0.353	0.415	0.384	0.341	0.463	0.386	0.348	0.427	0.381	0.335	0.489	0.402
	E	0.1	0.410	0.423	0.382	0.391	0.473	0.382	0.403	0.432	0.379	0.410	0.550	0.426
		0.4	0.398	0.417	0.369	0.379	0.464	0.369	0.391	0.428	0.366	0.395	0.524	0.405
		0.7	0.376	0.411	0.345	0.358	0.448	0.345	0.369	0.421	0.343	0.365	0.484	0.365

MS: Madde sayısı, G: Gerçek verinin faktör yapısına göre oluşturulan Q-matrisi, E: Nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi

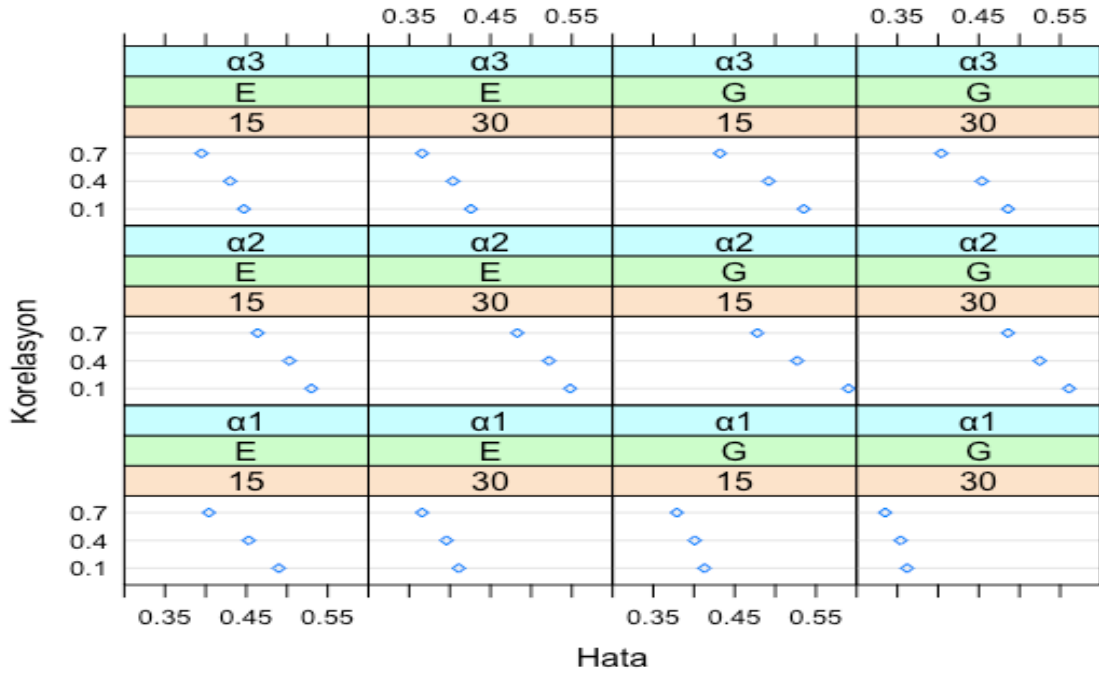
Üç nitelikten oluşan tek boyutlu veri setleri için 3PL ÇB-MTK'ya göre kestirilen hata değerleri Şekil 2'de görülmektedir.



Şekil 2. Üç nitelikten oluşan tek boyutlu veri setleri için ÇB-MTK'ya göre kestirilen birey parametrelerine ilişkin hata değerleri.

Şekil 2 incelendiğinde, 3PL ÇB-MTK' ya göre kestirilen hata değerlerinin 1. nitelik için 0.317-0.448, 2. nitelik için 0.319-0.408 ve 3. nitelik için 0.337-0.488 arasında değiştiği görülmektedir. Madde sayısı 15'ten 30'a çıkarıldığında, tüm koşullar için hata değerleri azalmaktadır. Test uzunluğunun artması ile hata değerlerinde gözlenen azalma miktarı, nitelikler arası korelasyon düştükçe artmaktadır. Nitelikler arasındaki korelasyonun artması da hata değerlerini düşürmekle birlikte en büyük değişim, korelasyonun 0.4'ten 0.7'ye çıkarıldığı koşullarda gözlenmiştir. Nitelikler arası korelasyonun artmasının hata değerleri üzerindeki etkisi ise test uzunluğunun 15 madde olduğu koşullarda, test uzunluğunun 30 madde olduğu koşullara göre daha fazladır. Gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılması 1. niteliğin her düzeyi için hata miktarını azaltırken, 2. ve 3. niteliklerin tüm düzeyleri için hatanın artmasına neden olmuştur. 1. nitelik için hata değerleri 0.023-0.050 arasında azalırken, 2. ve 3. nitelikler için sırasıyla 0.003- 0.010 ve 0.034-0.061 arasında artmıştır. Q-matrisi yapısının 1. ve 3. nitelikler için kestirilen hata değerleri üzerindeki etkisi ise 2. nitelik için kestirilen hata değerleri üzerindeki etkisinden daha fazladır. Bu durum,

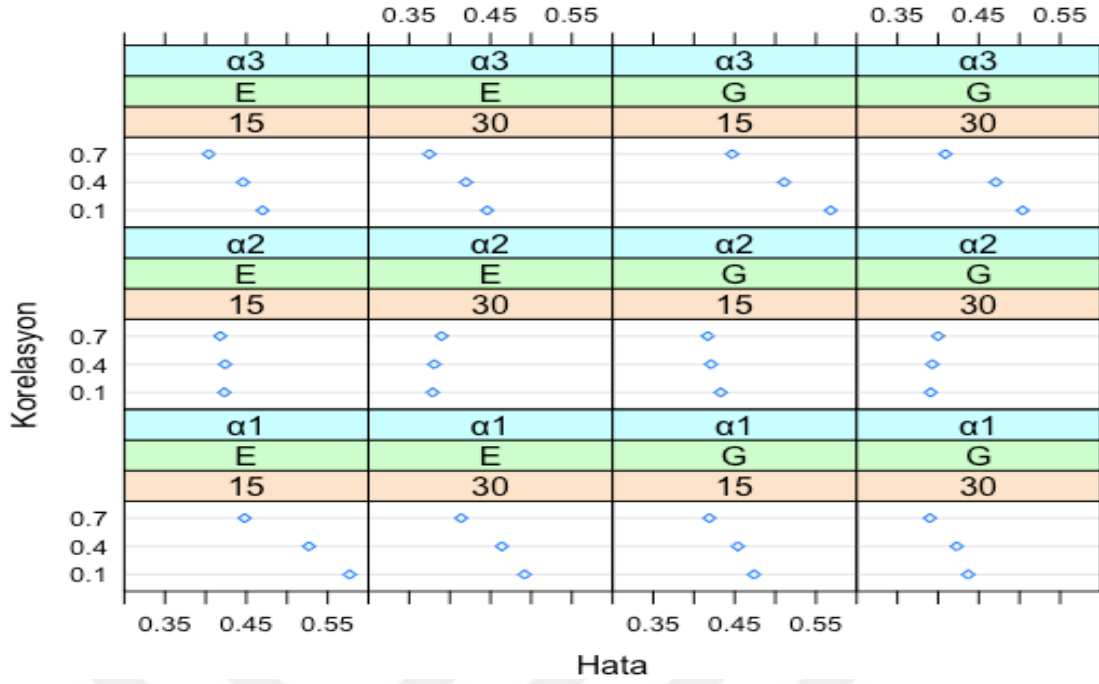
kullanılan farklı Q-matrisi yapılarına göre 2. niteliği ölçen madde sayısının, 1. ve 3. nitelikleri ölçen madde sayısına göre daha az değiştirmesinden kaynaklanabilir.



Şekil 3. Üç nitelikten oluşan tek boyutlu veri setleri için DINA'ya göre kestirilen birey parametrelerine ilişkin hata değerleri.

3 nitelikten oluşan tek boyutlu veri setlerinin tüm koşulları için DINA modeline göre kestirilen RMSE değerleri incelendiğinde, en fazla hatalı kestirimin 2. nitelik için yapıldığı (0.464 ile 0.590 arasında) ve bunu sırasıyla 3. ve 1. niteliklerin takip ettiği görülmektedir. Testteki madde sayısı arttıkça, 1. ve 3. niteliğin tüm düzeyleri için kestirilen hata miktarları sırasıyla 0.039-0.080 ve 0.021-0.049 arasında azalmıştır. 2. nitelik için niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarsa ise madde sayısının artması, hata miktarını 0.018-0.019 arasında arttırmıştır. Nitelikler arasındaki korelasyonun artması ile tüm koşul ve nitelikler için hata miktarının azaldığı gözlenmiştir. 0.4-0.7 korelasyon düzeyleri arasındaki değişiminin RMSE değerleri üzerindeki etkisinin ise 0.1-0.4 korelasyon düzeyleri arasındaki değişimden daha etkili olduğu görülmüştür. Niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması, 2. ve 3. niteliklerin tüm düzeyleri için hata miktarını azaltırken, 1. niteliğin tüm düzeyleri için daha hatalı kestirimler yapılmasına neden olmuştur. Q-matrisi yapısına göre hata değerlerindeki değişim miktarı ise test uzunluğu ve nitelikler arası korelasyonun artması ile azalmaktadır.

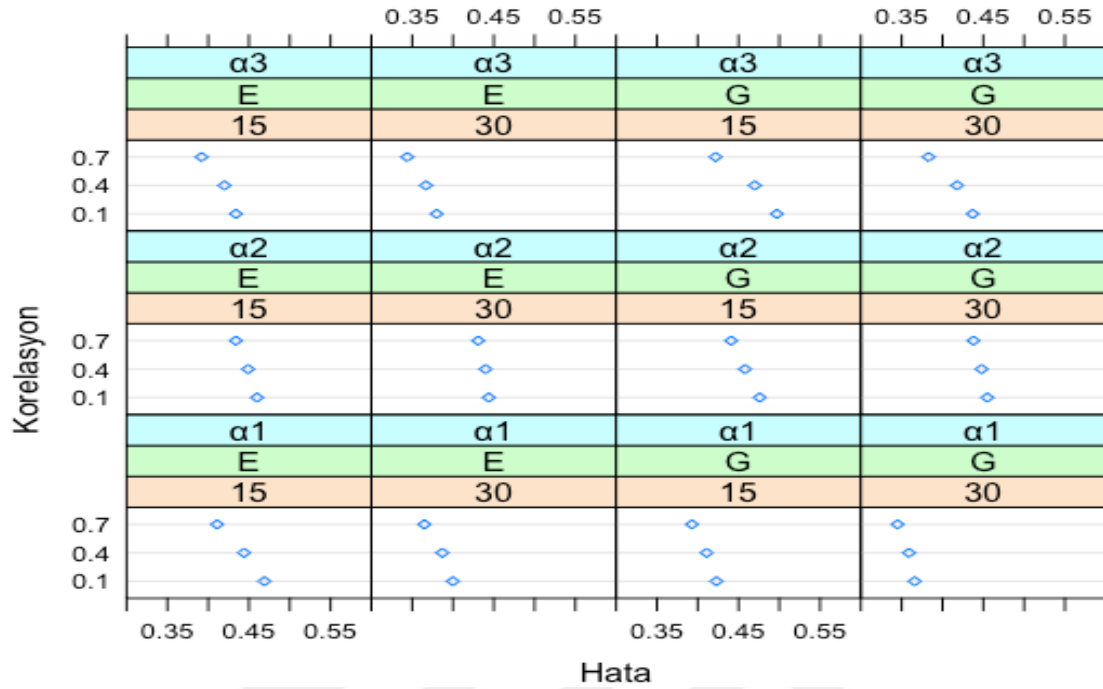




Şekil 4. Üç nitelikten oluşan tek boyutlu veri setleri için DINO'ya göre kestirilen birey parametrelerine ilişkin hata değerleri.

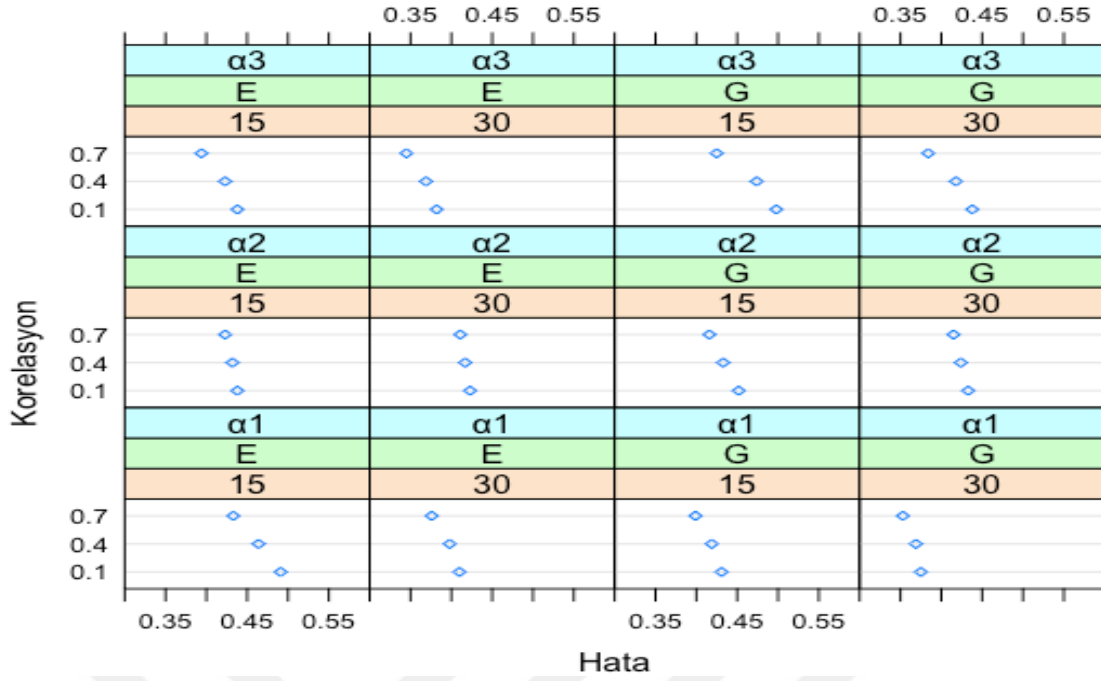
Telafi edici BTM'lerden DINO modeli ile kestirilen hata değerlerinin 1. nitelik için 0.390-0.577, 2. nitelik için 0.379-0.433 ve 3. nitelik için 0.375-0.568 arasında değiştiği görülmektedir. DINO modeli yüksek hata kestirimini ise 0.577 değeri ile 15 maddelik test uzunluğu, 0.1 korelasyon düzeyi ve niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşulda elde etmiştir. Test uzunluğunun artması, tüm nitelik düzeyleri için birey parametre kestirimlerinin hata miktarını azaltmıştır. Hata miktarındaki en büyük düşüşler 1. ve 2. nitelikler için niteliklerin eşit sayıda ölçüldüğü Q-matrisi yapısında görülürken, 3. nitelik için gerçek verinin faktör yapısına göre oluşturulan Q-matrisi koşulunda elde edilmiştir. Madde sayısı ve korelasyon düzeylerindeki değişimin etkileri birlikte değerlendirildiğinde ise hata miktarındaki azalmanın en fazla olduğu korelasyon düzeyinin 0.1 olduğu bulunmuştur. Nitelikler arası korelasyonun artması ise 2. niteliğinin 30 maddelik test koşulu altındaki hata kestirimlerini artırırken, diğer tüm koşullar için hataları azaltmıştır. Orta-yüksek korelasyon düzeyleri arasındaki değişimin hata üzerindeki etkisi ise düşük-orta korelasyon düzeyleri arasındaki değişimden daha fazladır. Q-matrisi yapısının RMSE değerleri üzerindeki etkisi incelendiğinde, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile 1. nitelik için hata miktarının 0.024-0.103 arasında

arttığı, 2. nitelik için hata miktarının 0.009-0.012 ve 3. nitelik için hata miktarının 0.034-0.097 arasında düştüğü gözlenmiştir.



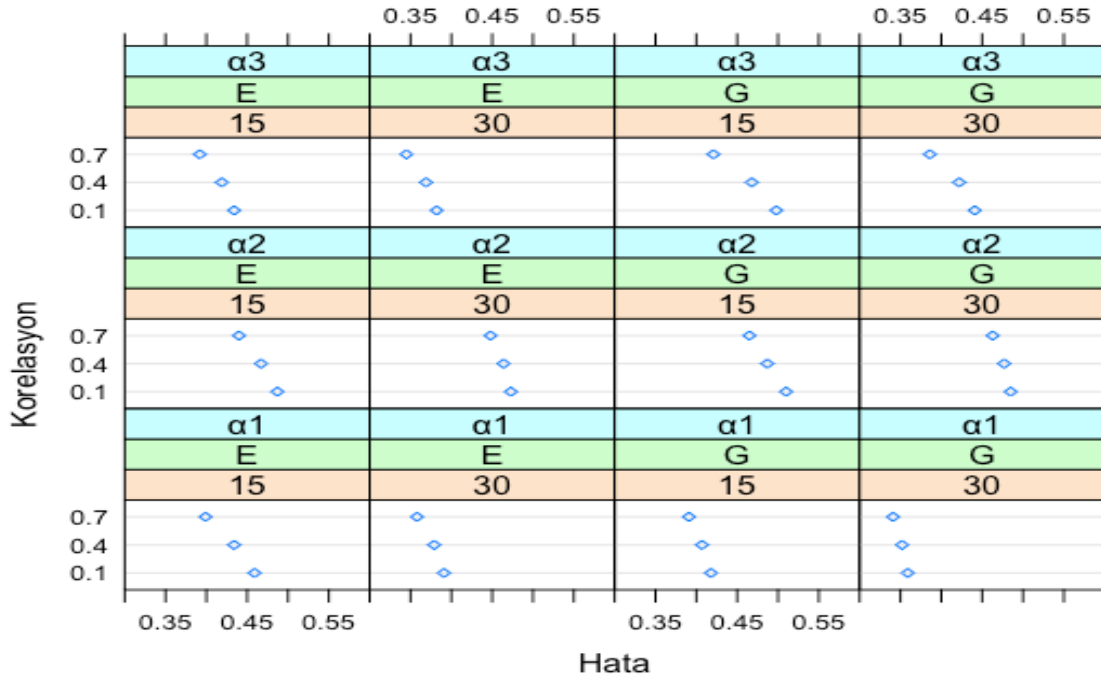
Şekil 5. Üç nitelikten oluşan tek boyutlu veri setleri için A-CDM'ye göre kestirilen birey parametrelerine ilişkin hata değerleri.

A-CDM ile kestirilen hata değerleri 1. nitelik için 0.345-0.469, 2. nitelik için 0.431-0.476 ve 3. nitelik için 0.344-0.497 arasında değişmektedir. Model her üç nitelik için en yüksek hata değerlerini 15 maddelik test uzunluğu ve 0.1 korelasyon koşulunda, en düşük hata değerlerini ise test uzunluğunun 30 madde ve nitelikler arası korelasyonun 0.7 olduğu koşullarda elde etmiştir. Hata değerlerindeki değişim madde sayısına göre incelendiğinde, test uzunluğu arttıkça her üç nitelik için kestirilen hata değerlerinin düştüğü ve hata değerlerindeki bu değişimin 1. ve 3. nitelik için 2. niteliğe göre daha fazla olduğu görülmektedir. Benzer şekilde, nitelikler arası korelasyonun artması ile üç niteliğin her düzeyinde hata miktarı azalmıştır. 0.1-0.4 korelasyon düzeylerindeki değişiminin hata üzerindeki etkisi ise 0.4-0.7 korelasyon düzeyindeki değişime göre daha küçüktür. Niteliklerin eşit sayıda ölçüldüğü Q-matrisi kullanıldığında 1. nitelik için hata değerleri 0.019-0.046 arasında artarken, 2 nitelik için hata değerleri 0.007-0.016 ve 3. nitelik için 0.030-0.062 arasında düşmüştür.



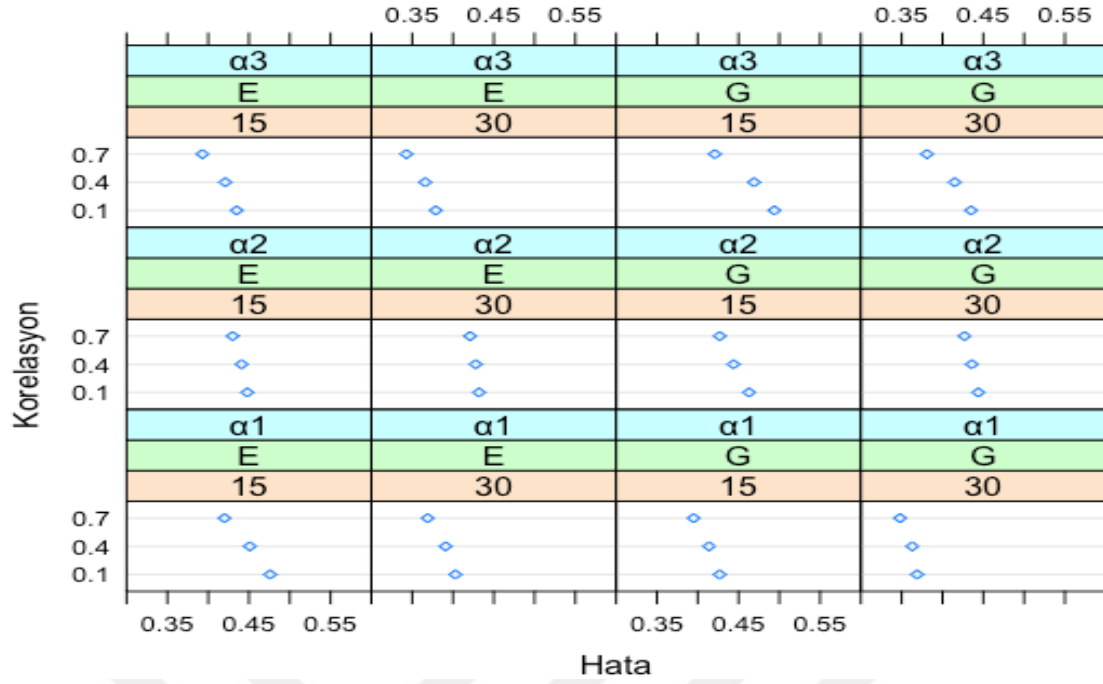
Şekil 6. Üç nitelikten oluşan tek boyutlu veri setleri için LLM'ye göre kestirilen birey parametrelerine ilişkin hata değerleri.

LLM modeline göre kestirilen birey parametrelerinin hata değerleri 1. nitelik için 0.353-0.491, 2. nitelik için 0.411-0.452 ve 3. nitelik için 0.345-0.498 arasında değişmektedir. Tüm koşullar için hata değerleri incelendiğinde, her üç nitelik için en yüksek hata değerlerinin test uzunluğunun 15 madde ve nitelikler arası korelasyonun 0.1 olduğu koşullarda elde ettiği görülmektedir. En küçük hata kestirimleri ise test uzunluğunun 30 madde ve nitelikler arası korelasyonun 0.7 olduğu koşullarda görülmektedir. Diğer koşullar sabit tutulduğunda madde sayısının 15'ten 30'a çıkarılması, üç niteliğin her düzeyinde hata miktarını düşürmektedir. Madde sayısının artmasının hatanın düşmesi üzerindeki etkisi, nitelikler arasındaki korelasyonun 0.1 olduğu düzeylerde daha büyük iken korelasyon düzeyi arttıkça, hatalardaki düşüş miktarı azalmaktadır. Benzer şekilde, nitelikler arasındaki korelasyonun artması, kestirilen hata değerlerini düşürmüştür. Ancak test uzunluğunun hata değerlerinin düşmesi üzerindeki etkisi, nitelikler arasındaki korelasyonun etkisinden daha fazladır. Kestirilen hata değerleri Q-matrisinin yapısı açısından karşılaştırıldığında, niteliklerin eşit sayıda ölçüldüğü Q-matrisi kullanıldığında 1. nitelik için hataların 0.022-0.060 arasında arttığı, 2. ve 3. nitelikler için ise hataların sırasıyla 0.001-0.014 ve 0.031-0.060 arasında düştüğü gözlenmiştir.



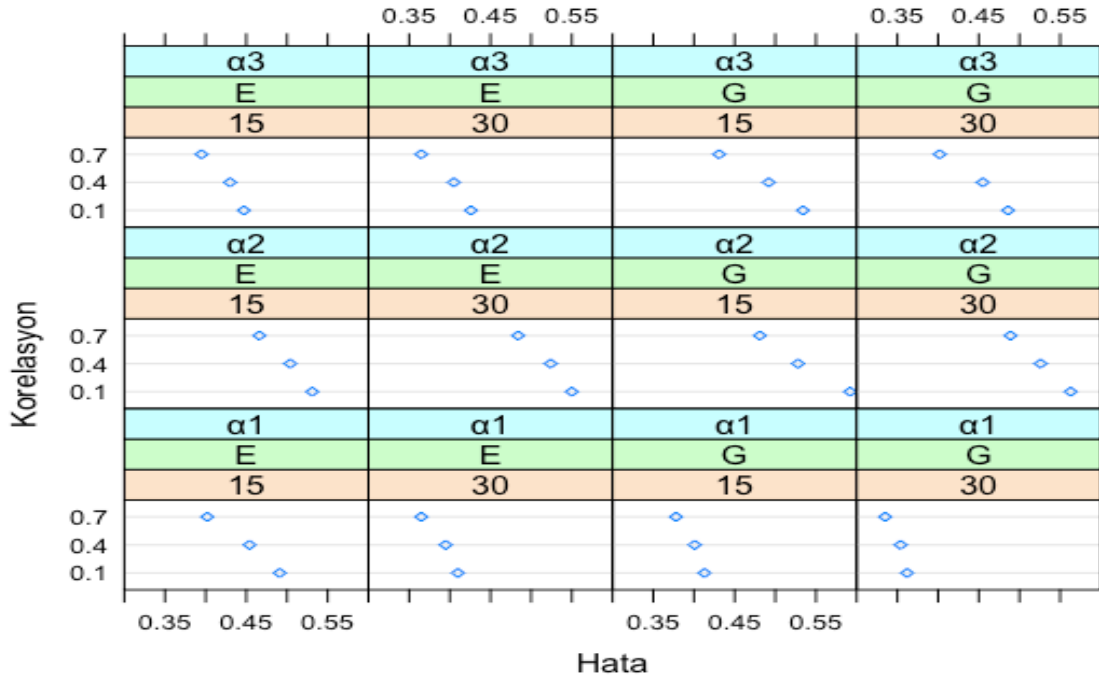
Şekil 7. Üç nitelikten oluşan tek boyutlu veri setleri için R-RUM'a göre kestirilen birey parametrelerine ilişkin hata değerleri.

Üç nitelikten oluşan tek boyutlu veri setleri için R-RUM modeline göre üretilen hata değerlerinin 1. nitelik için 0.341-0.459, 2. nitelik için 0.440-0.510 ve 3. nitelik için 0.345-0.498 arasında değiştiği görülmektedir. Test uzunluğu arttıkça, nitelikler arası korelasyonun 0.7 olduğu ve niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşul için 2. niteliğin hata miktarı 0.007 artarken, diğer tüm nitelik düzeylerinde hata miktarı düşüş göstermektedir. Madde sayısının artması ile 1. ve 3. nitelikler kapsamında kestirilen hata değerlerindeki düşüş ise 2. nitelik için kestirilen hata değerlerindeki değişimden daha fazladır. Nitelikler arasındaki korelasyonun artması ile her üç nitelik için kestirilen birey parametrelerine ilişkin hatalarının azaldığı gözlenmiştir. Orta-yüksek korelasyon düzeyleri için kestirilen hata değerleri arasındaki fark ise düşük-orta korelasyon düzeyleri arasındaki farktan daha fazladır. Q-matrisi yapısının kestirilen hata değerleri üzerindeki etkisi değerlendirildiğinde, niteliklerin eşit sayıda ölçüldüğü Q-matrisi koşulu altında 1. nitelik için hata değerlerinin arttığı, 2. ve 3. nitelikler için ise azaldığı görülmektedir. Q-matrisi yapısına göre kestirilen hata miktarlarındaki değişim ise nitelikler arasındaki korelasyon arttıkça, azalmaktadır.



Şekil 8. Üç nitelikten oluşan tek boyutlu veri setleri için G-DINA'ya göre kestirilen birey parametrelerine ilişkin hata değerleri.

G-DINA modeline göre nitelik sınıflamaları için kestirilen hata değerlerinin 1. nitelik için 0.348-0.476, 2. nitelik için 0.421-0.463 ve 3. nitelik için 0.343-0.494 arasında değiştiği görülmektedir. Her üç nitelik için en yüksek hata kestirimleri test uzunluğunun 15 ve nitelikler arası korelasyonun 0.1 olduğu koşullarda, en düşük hata değerleri ise madde sayısının 30 ve nitelikler arası korelasyonun 0.7 olduğu koşullarda elde edilmiştir. Test uzunluğunun artması, tüm nitelik düzeylerinde kestirilen hata miktarını düşürmüştür. Madde sayısındaki artışa göre 1. ve 3. nitelikler kapsamında kestirilen hata değerleri arasındaki fark sırasıyla 0.047-0.073 ve 0.040-0.060 arasında değişirken, 2. nitelik için 0.000 ile 0.019 arasında değişmektedir. Hata miktarının azalmasındaki bir diğer faktör ise nitelikler arasındaki korelasyonun artmasıdır. Çalışma kapsamındaki diğer modellerin yaptığı kestirimler ile benzer şekilde, nitelikler arasındaki korelasyon düşükten yükseğe doğru çıktıkça, G-DINA modeline göre kestirilen hata miktarı azalmaktadır. Niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanımı 2. ve 3. nitelikler kapsamında kestirilen hata değerlerini sırasıyla 0.003-0.015 ve 0.028-0.059 arasında düşürürken, 1. nitelik için 0.021-0.050 arasında arttırmıştır. Q-matrisi yapısına göre hata miktarlarındaki değişim ise nitelikler arasındaki korelasyon 0.1'den 0.7'ye doğru çıktıkça, azalma göstermektedir.



Şekil 9. Üç nitelikten oluşan tek boyutlu veri setleri için HO-DINA'ya göre kestirilen birey parametrelerine ilişkin hata değerleri.

HO-DINA'ya göre kestirilen hata değerleri incelendiğinde en yüksek değerlerin 2. niteliğe ait olduğu ve 0.466 ile 0.592 arasında değiştiği görülmektedir. 1. ve 3. nitelikler kapsamında kestirilen hata değerleri ise sırasıyla 0.335-0.491 ve 0.365-0.534 arasında değişmektedir. Test uzunluğunun artması, 1. ve 3. niteliklerin tüm düzeyleri için kestirilen hata miktarının düşmesine neden olmakla birlikte hata miktarındaki azalma oranı 1. nitelik için daha fazladır. 2. nitelik kapsamında niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda ise test uzunluğunun artması, hata miktarını 0.019-0.020 arasında arttırmıştır. Nitelikler arası korelasyonun artması ise tüm nitelik düzeyleri için hata miktarını düşürmüştür. 1. nitelik için nitelikler eşit sayıda ölçülecek şekilde oluşturulan Q-matrisinin kullanıldığı koşullarda korelasyonun artmasının hatanın düşmesi üzerindeki etkisi, gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullara göre daha fazla iken 2. ve 3. nitelikler için tam tersi söz konusudur. Diğer koşullar sabit tutulduğunda niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması kestirilen hata değerlerinin 2. nitelik için 0.002-0.061 arasında ve 3. nitelik için 0.037-0.087 arasında düşmesine, 1. nitelik için ise 0.025-0.078 arasında artmasına neden olmuştur. Tüm nitelikler için Q-matrisi yapısına göre hatalardaki değişim ise test

uzunluğunun 15 madde olduğu koşullarda ve nitelikler arası korelasyon düştükçe daha fazladır.

### **Üç Nitelikten Oluşan Tek Boyutlu Veri Setlerinin Hata Değerleri: Özet**

Çalışma kapsamındaki tüm modellerin tek boyutlu veri yapısı altında test uzunluğu, nitelikler arası korelasyon ve Q-matrisi yapısı koşullarının tüm düzeylerindeki RMSE değerleri karşılaştırıldığında, verilerin üretildiği 3PL ÇB-MTK modelinin üç nitelik için de daha hatasız kestirimler yaptığı görülmektedir. 1. ve 3. nitelikler için en hatalı kestirimleri telafi edici BTM'lerden DINO modelinin, 2. nitelik için ise DINA ve HO-DINA modellerinin yaptığı bulunmuştur.

Üç nitelikten oluşan tek boyutlu veri setlerine ilişkin hata değerleri incelendiğinde test uzunluğu ve nitelikler arası korelasyon düzeyi arttıkça, hata değerlerinin genel olarak düştüğü gözlenmiştir. Ayrıca nitelikler arası 0.4-0.7 korelasyon düzeyleri arasındaki değişimin hatanın azalması üzerindeki etkisinin, 0.1-0.4 korelasyon düzeyindeki değişimden daha fazla olduğu bulunmuştur. Q-matrisi yapısının kestirilen hata değerleri üzerindeki etkisi ise niteliklere göre farklılık göstermektedir. Genel olarak, 1. nitelik kapsamında nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi kullanıldığında hata miktarının arttığı, 2. ve 3. nitelikler için ise hataların azaldığı görülmüştür.

### **Üç Nitelikten Oluşan Tek Boyutlu Veri Setleri için Birey Parametrelerine İlişkin Yanlılık Değerleri**

Tek boyutlu veri setleri için testteki madde sayısı, nitelikler arası korelasyon ve Q-matrisi yapısı koşullarına göre 3PL ÇB-MTK ve BTM'ler kapsamında kestirilen birey parametre ilişkin yanlılık değerleri Tablo 17'de sunulmaktadır. Analiz sonuçlarının daha ayrıntılı olarak açıklanabilmesi için öncelikle her bir modeldeki değişimler tek tek incelenmiş ve ardından tüm koşullar için modeller karşılaştırılmıştır.

Tablo 17

Üç Nitelikten Oluşan Tek Boyutlu Veri Setleri için Birey Parametrelerine İlişkin Yanlılık Değerleri

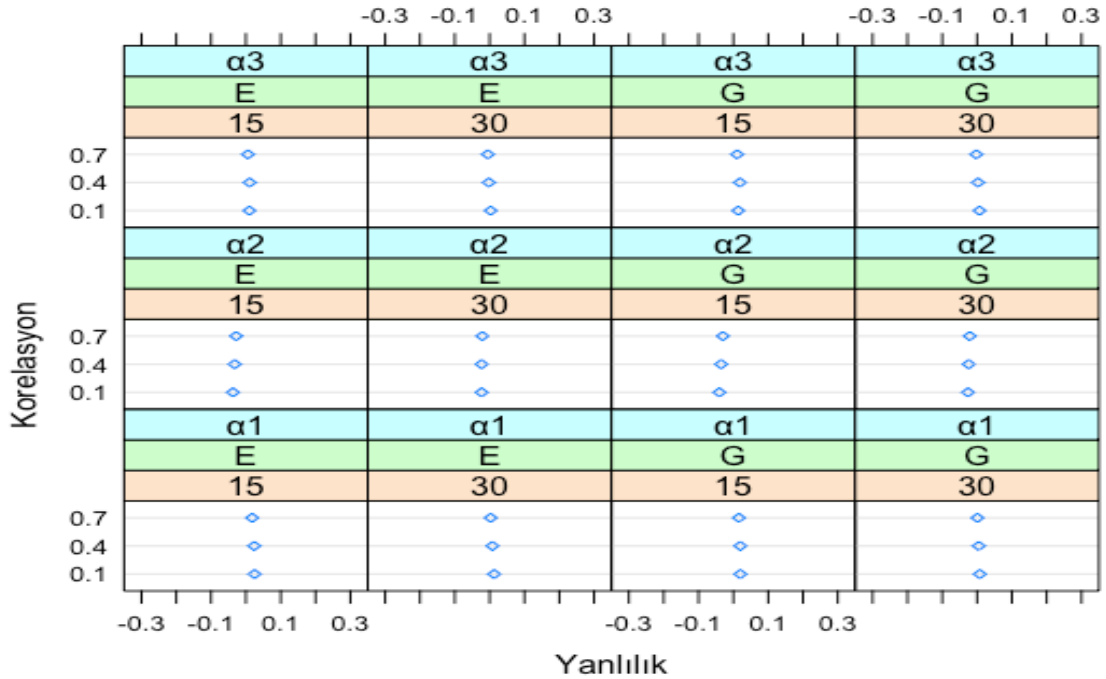
MS	Q-Matrisi	Korelasyon	ÇB-MTK			DINA			DINO			A-CDM		
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$
15	G	0.1	0.020	-0.040	0.014	-0.027	0.313	0.130	-0.186	-0.035	-0.267	-0.105	0.153	-0.051
		0.4	0.020	-0.035	0.019	-0.036	0.233	0.057	-0.165	0.028	-0.176	-0.099	0.154	-0.041
		0.7	0.016	-0.030	0.011	-0.055	0.198	0.011	-0.136	0.107	-0.099	-0.098	0.160	-0.042
	E	0.1	0.025	-0.037	0.010	0.069	0.237	0.042	-0.302	0.018	-0.142	-0.101	0.137	-0.047
		0.4	0.024	-0.032	0.011	-0.002	0.210	0.017	-0.233	0.076	-0.112	-0.097	0.143	-0.042
		0.7	0.018	-0.028	0.006	-0.044	0.185	0.000	-0.149	0.120	-0.073	-0.093	0.149	-0.039
30	G	0.1	0.008	-0.026	0.007	0.002	0.301	0.118	-0.167	0.025	-0.190	-0.077	0.176	-0.021
		0.4	0.005	-0.024	0.003	-0.008	0.257	0.060	-0.154	0.069	-0.143	-0.078	0.178	-0.019
		0.7	0.001	-0.021	-0.002	-0.027	0.224	0.021	-0.126	0.121	-0.066	-0.075	0.179	-0.014
	E	0.1	0.013	-0.023	0.003	0.034	0.290	0.094	-0.215	0.034	-0.151	-0.081	0.172	-0.019
		0.4	0.008	-0.023	-0.002	0.003	0.259	0.066	-0.182	0.067	-0.118	-0.079	0.173	-0.018
		0.7	0.003	-0.021	-0.005	-0.029	0.223	0.034	-0.135	0.116	-0.072	-0.076	0.174	-0.016
MS	Q-Matrisi	Korelasyon	LLM			R-RUM			G-DINA			HODINA		
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$
15	G	0.1	-0.122	0.109	-0.084	-0.089	0.206	0.016	-0.113	0.132	-0.063	-0.027	0.316	0.125
		0.4	-0.116	0.114	-0.079	-0.087	0.195	-0.003	-0.107	0.134	-0.048	-0.036	0.236	0.058
		0.7	-0.109	0.125	-0.060	-0.093	0.191	-0.021	-0.102	0.141	-0.045	-0.054	0.204	0.015
	E	0.1	-0.165	0.098	-0.072	-0.033	0.182	-0.019	-0.131	0.120	-0.060	0.077	0.238	0.042
		0.4	-0.150	0.115	-0.064	-0.049	0.169	-0.024	-0.123	0.132	-0.053	0.006	0.213	0.018
		0.7	-0.137	0.135	-0.051	-0.060	0.158	-0.031	-0.114	0.145	-0.044	-0.035	0.188	0.002
30	G	0.1	-0.093	0.147	-0.054	-0.057	0.214	0.026	-0.084	0.162	-0.041	0.002	0.303	0.119
		0.4	-0.094	0.149	-0.051	-0.061	0.211	0.017	-0.085	0.164	-0.034	-0.008	0.259	0.070
		0.7	-0.088	0.153	-0.037	-0.066	0.205	0.011	-0.080	0.166	-0.025	-0.027	0.228	0.025
	E	0.1	-0.107	0.145	-0.043	-0.047	0.205	0.015	-0.093	0.158	-0.029	0.036	0.292	0.096
		0.4	-0.105	0.146	-0.043	-0.050	0.201	0.010	-0.091	0.160	-0.029	0.006	0.262	0.069
		0.7	-0.097	0.153	-0.036	-0.056	0.192	0.003	-0.086	0.164	-0.024	-0.026	0.225	0.036

MS: Madde sayısı, G: Gerçek verinin faktör yapısına göre oluşturulan Q-matrisi, E: Nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi



Birey parametrelerine ilişkin yanlılık değerlerindeki değişim, her bir model için Şekil 10-Şekil17 üzerinde gösterilmiştir. Yanlılık değerlerindeki değişimler -0.3 ile 0.3 değerleri arasında ölçeklenmiş olmak üzere x-eksenine, boyutlar arası korelasyon değerleri (0.1, 0.4, 0.7) ise y-eksenine yerleştirilmiştir. Grafiklerde  $\alpha_1$ ,  $\alpha_2$  ve  $\alpha_3$  nitelikleri aşağıdan yukarı doğru sıralanmış ve her niteliğin altında Q-matrisi yapısı (E ve G) ile test uzunluğu (15 ve 30) değişkenleri yer almıştır. Her bir nitelikteki değişim madde sayısı, nitelikler arası korelasyon ve Q-matrisi yapısı koşullarına göre ayrı ayrı ele alınmıştır.

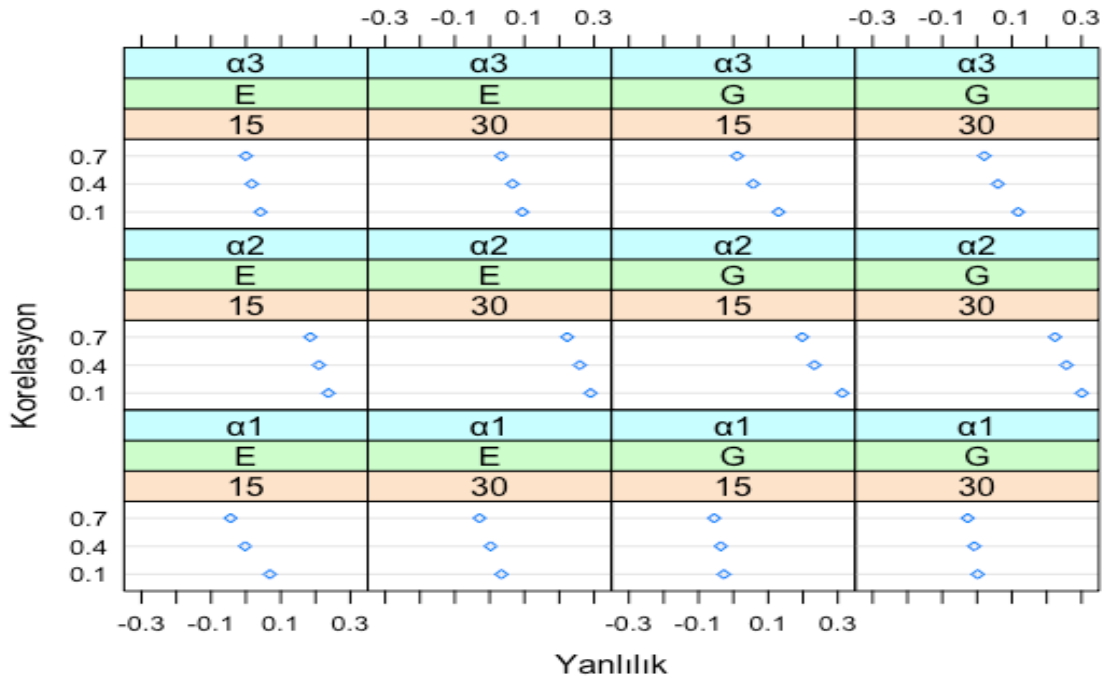
3 nitelikten oluşan tek boyutlu veri setleri için 3PL ÇB-MTK'ya göre kestirilen yanlılık değerleri Şekil 10'da görülmektedir.



Şekil 10. Üç nitelikten oluşan tek boyutlu veri setleri için ÇB-MTK'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

Şekil 10'da sunulan yanlılık değerleri incelendiğinde, 1. nitelik için yanlılık değerlerinin 0.001 ile 0.025; 2. nitelik için -0.040 ile -0.021 ve 3. nitelik için -0.005 ile 0.019 arasında değiştiği görülmektedir. Testteki madde sayısı ve nitelikler arası korelasyon değerleri arttıkça yanlılık değerlerinin düştüğü, ancak yanlılık değerlerinin düşmesindeki etkili değişkenin madde sayısı olduğu gözlenmiştir. Nitekim, diğer koşullar sabit tutulduğunda, madde sayısındaki artışa göre yanlılık kestirimleri arasındaki fark 0.006 ile 0.016 arasında değişirken; nitelikler arasındaki korelasyonun artmasına göre bu fark 0.000 ile 0.008 arasında değişmektedir. Q-

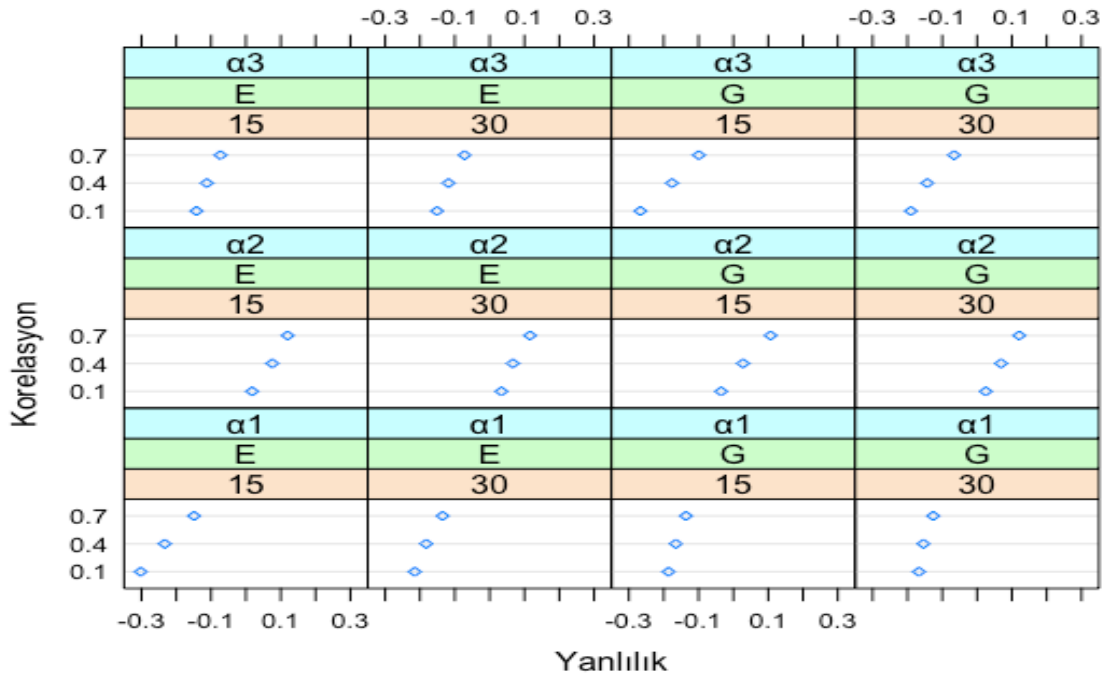
matrisi yapısının yanlılık üzerindeki etkisi incelendiğinde, gerçek verinin faktör yapısına göre oluşturulan Q-matrisi kullanıldığında 1. niteliğin daha düşük, diğer niteliklerin ise daha yüksek yanlılık değerlerine sahip olduğu görülmektedir. Q-matrisinin yapısına göre yanlılık kestirimleri arasındaki farklar 1. nitelik için 0.001-0.005, 2. nitelik için 0.000-0.003 ve 3. nitelik için 0.003-0.008 arasında değişmektedir. 3PL ÇB-MTK'nın, genel olarak, madde sayısının ve nitelikler arası korelasyonun daha yüksek olduğu koşullarda, daha yansız kestirim yaptığı söylenebilir.



Şekil 11. Üç nitelikten oluşan tek boyutlu veri setleri için DINA'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

Bireyin soruyu doğru cevaplayabilmesi için soruda ölçülen niteliklerin hepsine sahip olmasını gerektiren DINA modeline göre en yüksek yanlılık değerlerinin 2. nitelik için kestirildiği ve bunu sırasıyla 3. ve 1. niteliklerin takip ettiği görülmektedir. Her üç nitelik için en yüksek yanlılık değerleri test uzunluğunun 15 ve nitelikler arası korelasyonun 0.1 olduğu koşullarda gözlenmiştir. DINA modelinden elde edilen yanlılık değerleri testteki madde sayısı açısından karşılaştırıldığında, 1. nitelik için testteki madde sayısı arttıkça yanlılık değerlerinin azaldığı ve bu farkın 0.016-0.035 arasında değiştiği bulunmuştur. 2. ve 3. nitelikler için madde sayısının artması ise yanlılık değerlerini arttırmaktadır. 2. ve 3. nitelikler için kestirilen yanlılık değerlerinin, nitelikler arası korelasyon arttıkça önemli düzeyde düştüğü dikkat çekmektedir.

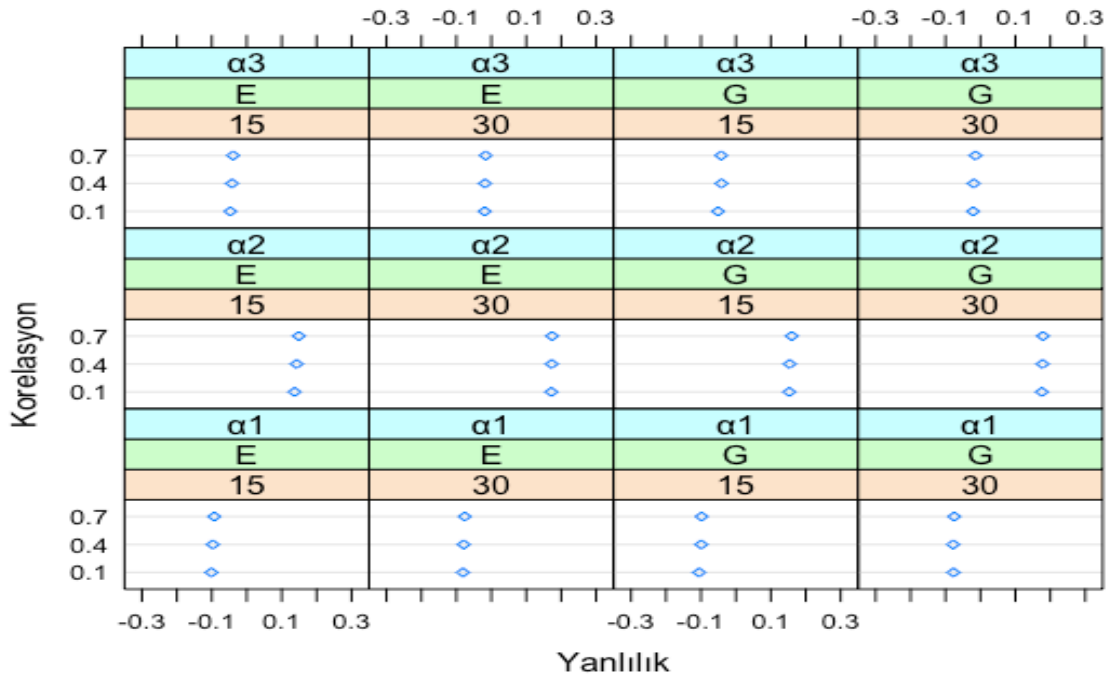
Nitelikler arası korelasyona göre yanlılık kestirimlerindeki değişim, 0.1-0.4 korelasyon düzeyleri arasında ve gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda ise daha fazladır. 1. nitelik için niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda 0.1-0.4 korelasyon düzeyleri arasındaki değişim yanlılığı düşürürken, diğer koşullar altında nitelikler arası korelasyonun artması ile yanlılık artmaktadır. Q-matrisi yapısına göre yanlılık incelendiğinde, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile 2. ve 3. nitelikler için kestirilen yanlılık değerlerinin düştüğü görülmektedir. Yanlılık değerlerindeki bu düşüş miktarı, madde sayısı ve nitelikler arası korelasyonun artması ile azalmaktadır.



Şekil 12. Üç nitelikten oluşan tek boyutlu veri setleri için DINO'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

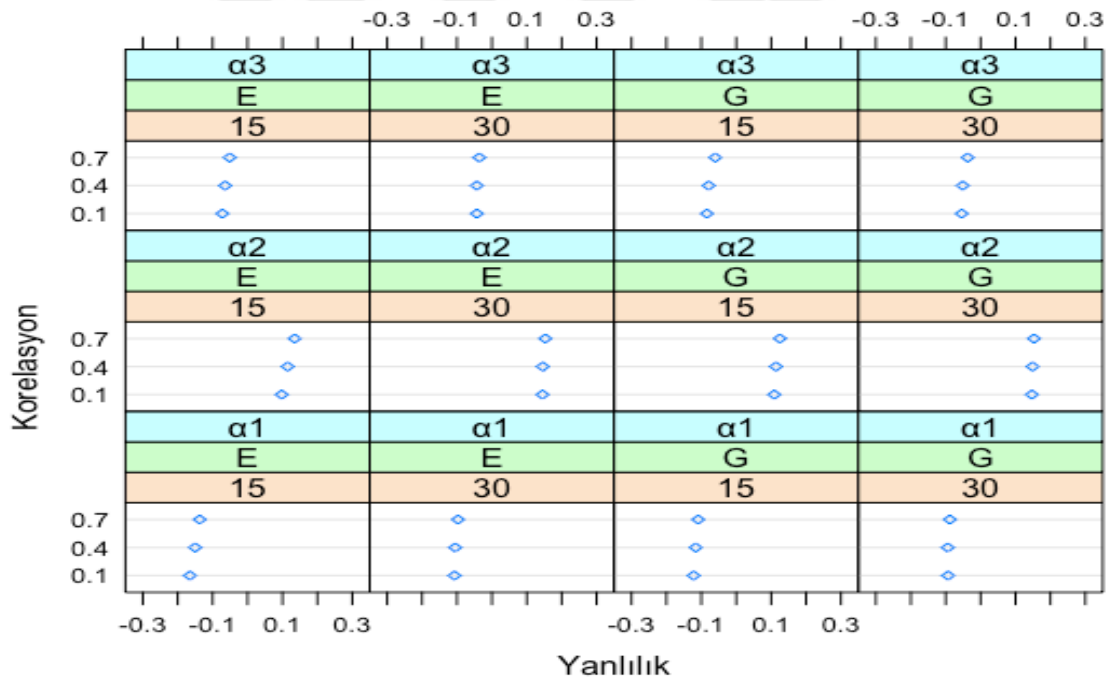
Tek boyutlu testler için madde sayısı, nitelikler arası korelasyon ve Q-matrisinin yapısına göre oluşturulan 12 koşul değerlendirildiğinde DINO modeli için kestirilen yanlılık değerlerinin 1. nitelik için -0.302 ile -0.126, 2. nitelik için -0.035 ile 0.121 ve 3. nitelik için -0.267 ile -0.066 arasında değiştiği görülmektedir. Diğer koşullar sabit tutulduğunda, testteki madde sayısı arttıkça 1. nitelik için kestirilen yanlılık değerlerinin 0.010-0.087 arasında azaldığı gözlenmiştir. Yanlılık değerlerindeki bu azalma, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı ve nitelikler arası korelasyon düşük olduğu koşullarda ise daha fazladır. 2. nitelik

için kestirilen yanlılık değerlerinin ise madde sayısındaki değişime göre bazı koşullarda arttığı bazı koşullarda ise azaldığı bulunmuştur. 3. nitelik için gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda madde sayısının artmasının yanlılık değerini 0.033-0.077 arasında düşürdüğü, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda ise 0.006-0.008 arasında arttırdığı görülmektedir. Nitelikler arasındaki korelasyonun 0.1'den 0.7'ye doğru artması ile 1. ve 3. nitelikler için kestirilen yanlılık değerlerinin sırasıyla 0.013-0.084 ve 0.030-0.091 arasında düştüğü gözlenmiştir. 2. niteliğin ise madde sayısının 15 olduğu ve gerçek veri yapısına göre oluşturulan Q-matrisinin kullanıldığı koşul dışındaki tüm düzeyler için korelasyonun artması ile yanlılık değerleri 0.033-0.079 arasında artmıştır. Q-matrisi yapısının yanlılık üzerindeki etkisi incelendiğinde, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda 1. nitelik için yanlılık değerlerinin 0.009-0.116 arasında arttığı, 3. nitelik için 0.025-0.125 arasında azaldığı, 2. nitelik için ise sürekli bir artış ya da azalış göstermediği bulunmuştur. Analiz sonuçlarına göre, DINO modeline göre kestirilen yanlılık değerleri üzerinde en etkili değişkenin Q-matrisi yapısı olduğu ve bu koşulu sırasıyla nitelikler arası korelasyon ve madde sayısı değişkenlerinin izlediği görülmektedir.



Şekil 13. Üç nitelikten oluşan tek boyutlu veri setleri için A-CDM'ye göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

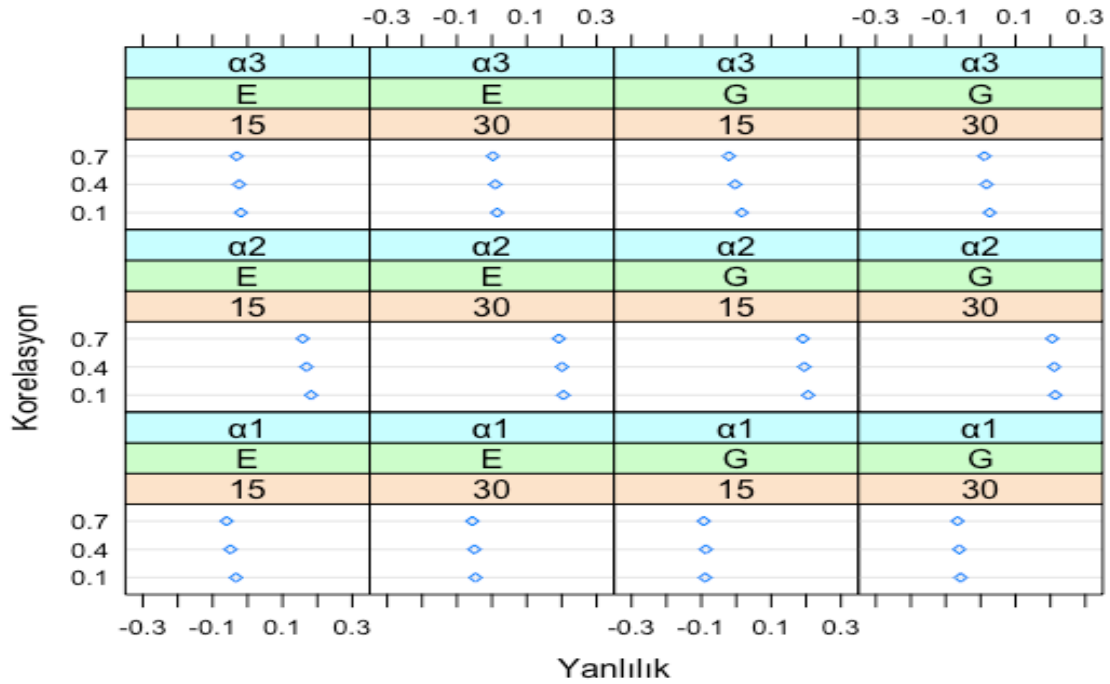
Şekil 13'te A-CDM'ye göre kestirilen yanlılık değerleri incelendiğinde, en düşük değerlerin 3. nitelik kapsamında kestirildiği ve bunu sırasıyla 1. ve 2. niteliğe ait kestirimlerin izlediği görülmektedir. Madde sayısı arttığında, 1. ve 3. niteliklerin tüm düzeyleri için kestirilen yanlılık değerlerinin sırasıyla 0.018-0.027 ve 0.022-0.030 arasında azaldığı, 2. nitelik için ise 0.019-0.034 arasında arttığı gözlenmiştir. Nitelikler arası korelasyon arttığında ise 1. ve 3. nitelikler için kestirilen yanlılık değerlerinin giderek düşüş gösterdiği bulunmuştur. 2. nitelik için ise nitelikler arası korelasyon düzeyinin 0.1'den 0.7'ye doğru artması ile yanlılık değerlerinin 0.001-0.006 arasında arttığı gözlenmiştir. Her bir nitelik eşit sayıda ölçülecek şekilde oluşturulan Q-matrisi kullanıldığında 1. nitelik için madde sayısının 15 olduğu koşullarda ve 2. niteliğin tüm düzeylerinde kestirilen yanlılık değerlerinin daha düşük olduğu görülmektedir. Sonuç olarak, A-CDM'ye göre kestirilen birey parametrelerinin yanlılığı üzerindeki en etkili faktörün, testteki madde sayısı olduğu tespit edilmiştir.



Şekil 14. Üç nitelikten oluşan tek boyutlu veri setleri için LLM'ye göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

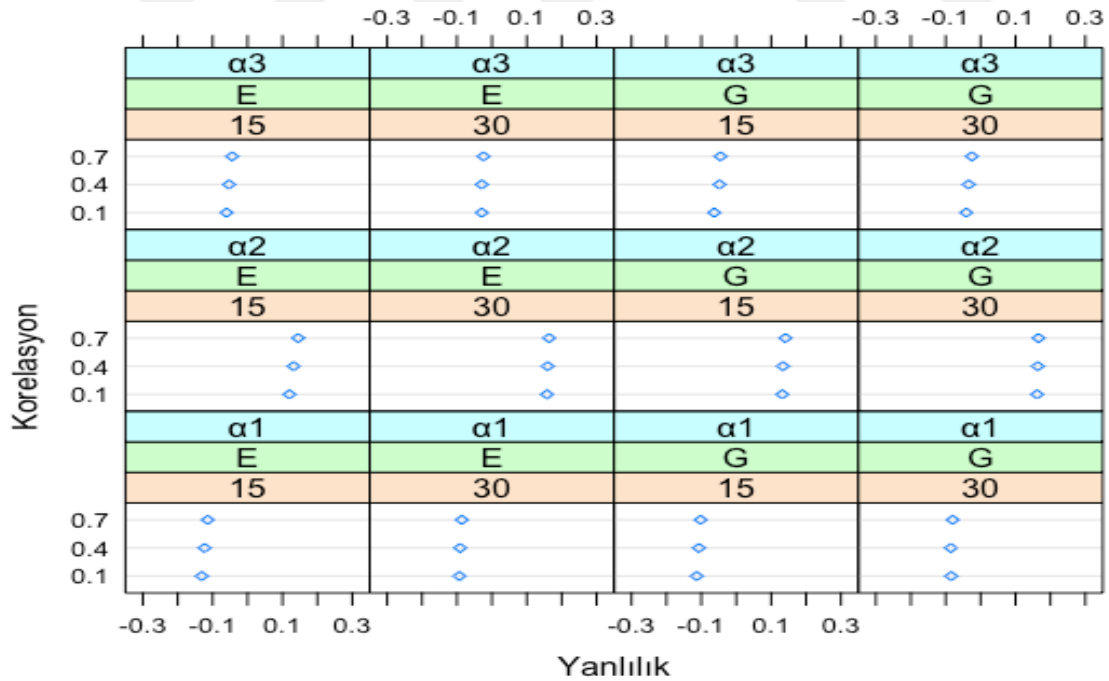
LLM modeli ile kestirilen yanlılık değerleri 1. nitelik için -0.165 ile -0.088, 2. nitelik için 0.098 ile 0.153 ve 3. nitelik için -0.084 ile -0.036 arasında değişmektedir. Model, 1. ve 3. niteliklerin tüm düzeyleri için negatif ve 2. nitelik için pozitif yanlılık kestirimi yapmıştır. Madde sayısının etkisi incelendiğinde, test uzunluğunun 15'ten

30'a çıkarılması ile 1. ve 3. nitelikler için kestirilen yanlılık değerlerinin düştüğü görülmektedir. 2. nitelik için madde sayısının artması ise yanlılık değerlerini 0.019 ile 0.047 arasında arttırmaktadır. Madde sayısının artmasına bağlı olarak yanlılık kestirimleri arasında oluşan farkın ise nitelikler arası korelasyonun artması ile azaldığı gözlenmiştir. Benzer şekilde, nitelikler arası korelasyonun artması 1. nitelik için kestirilen yanlılık değerlerini 0.002-0.016 ve 3. nitelik için kestirilen yanlılık değerlerini 0.003-0.019 arasında azaltmıştır. 2. nitelik kapsamında nitelikler arası korelasyonun artması ise daha yanlı kestirimler yapılmasına neden olmuştur. Testteki madde sayısı ve nitelikler arası korelasyon değişkenleri, nitelik bazında kestirilen yanlılık değerleri üzerinde benzer etkiye sahip olmasına rağmen madde sayısındaki değişimin yanlılık üzerinde daha büyük bir etkisi olduğu görülmektedir. Niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması, 1. niteliğin tüm düzeylerinde yanlılık değerlerini 0.008 ile 0.043 arasında arttırmıştır. 1. nitelik kapsamında yanlılık değerlerindeki bu değişim, test uzunluğunun 15 madde olduğu ve nitelikler arası korelasyonun düştüğü koşullarda daha büyüktür. 3. niteliğin tüm düzeyleri ve 2. niteliğin 30 maddelik test uzunluğu koşulu altında, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile LLM'nin daha yansız kestirimler yaptığı gözlenmiştir.



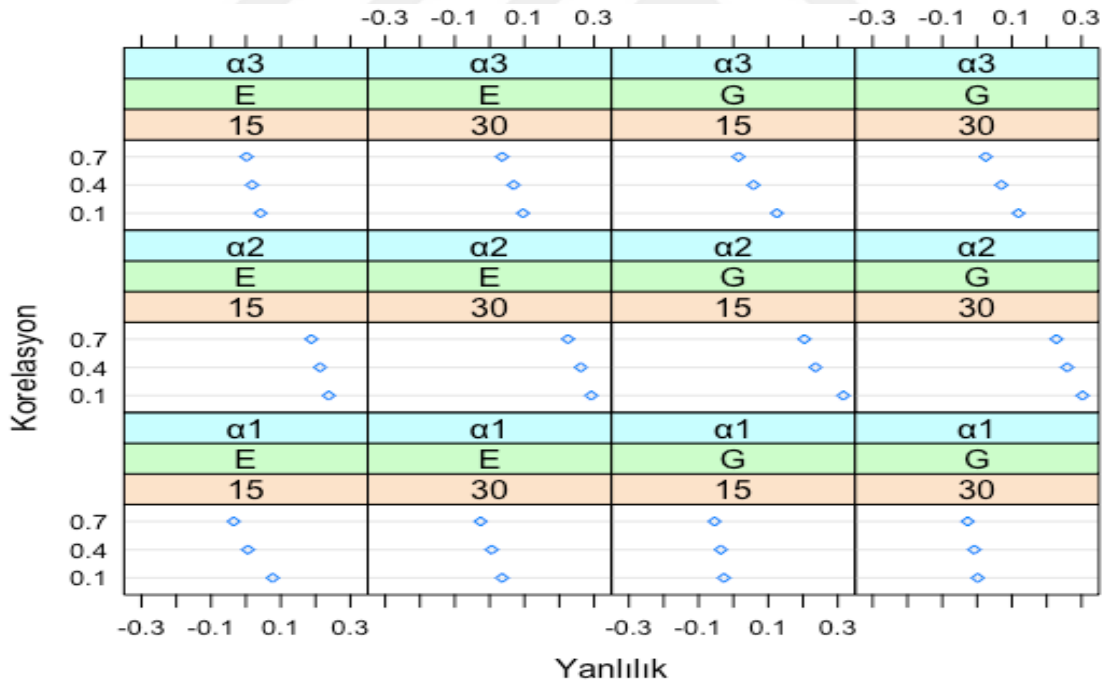
Şekil 15. Üç nitelikten oluşan tek boyutlu veri setleri için R-RUM'a göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

R-RUM modeline göre kestirilen birey parametrelerinin yanlılığı 1. nitelik için -0.093 ile -0.033, 2. nitelik için 0.158 ile 0.214 ve 3. nitelik için -0.031 ile 0.026 arasında değişmektedir. Test uzunluğunun kestirilen yanlılık değerleri üzerindeki etkisi, niteliklere göre farklılık göstermektedir. 1. nitelik için gerçek verinin faktör yapısına göre oluşturulan Q-matrisi kullanıldığında, test uzunluğu arttıkça yanlılığın azaldığı; niteliklerin eşit sayıda ölçüldüğü Q-matrisi koşulunda ise madde sayısının artması ile yanlılığın da arttığı dikkat çekmektedir. 2. niteliğin tüm düzeylerinde yanlılık 0.008-0.033 arasında artarken, 3. nitelik için gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı ve nitelikler arası korelasyonun 0.1-0.4 olduğu koşullar için yanlılık değerleri artmaktadır. Nitelikler arası korelasyonun artması ise 2. niteliğin tüm koşulları ve madde sayısının 30 olduğu koşullar altında 3. nitelik için yanlılık değerlerinin düşmesine neden olmuştur. Q-matrisi yapısının kestirilen yanlılık değerleri üzerindeki etkisi incelendiğinde, 3. nitelik için madde sayısının 15 olduğu koşullar dışındaki tüm nitelik düzeyleri için niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılmasının yanlılığı azalttığı gözlenmiştir. Kullanılan Q-matrisi yapısına göre yanlılık değerlerinde gözlenen düşüş, test uzunluğunun 15 madde olduğu koşullarda ise daha fazladır.



Şekil 16. Üç nitelikten oluşan tek boyutlu veri setleri için G-DINA'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

Birey parametrelerinin G-DINA modeline göre kestirimlerinden elde edilen yanlılık değerlerinin 1. nitelik için -0.131 ile -0.080, 2. nitelik için 0.120 ile 0.166 ve 3. nitelik için -0.063 ile -0.024 arasında değiştiği görülmektedir. Madde sayısı arttıkça, 1. ve 3. niteliklerin tüm düzeyleri için kestirimlerin yanlılığının azaldığı; 2. nitelik için kestirilen yanlılık değerlerinin ise arttığı gözlenmiştir. Benzer şekilde, nitelikler arası korelasyon düzeyi arttıkça 1. ve 3. niteliklere ait kestirimlerin yanlılığı düşerken, 2. niteliğe ilişkin yanlılık değerleri 0.002-0.013 arasında artmıştır. Testteki madde sayısı ve nitelikler arası korelasyon koşullarının yanlılık kestirimleri üzerinde benzer etkileri olmasına rağmen madde sayısındaki değişimin yanlılık üzerinde daha etkili olduğu tespit edilmiştir. Diğer koşullar sabit tutulduğunda, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması, 2. nitelik için kestirilen yanlılık değerlerinin 0.002 ile 0.013 arasında ve 3. nitelik için 0.001 ile 0.012 arasında düşmesine neden olmuştur. 1. nitelik için gerçek verinin faktör yapısına dayalı olarak oluşturulan Q-matrisinin kullanıldığı koşullarda ise yanlılık değeri diğer Q-matrisi yapısına göre 0.006 ile 0.018 arasında değişen, daha küçük kestirimler yapmıştır.



Şekil 17. Üç nitelikten oluşan tek boyutlu veri setleri için HO-DINA'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

HO-DINA modeline göre kestirilen yanlılık değerleri 1. nitelik için -0.054 ile 0.077, 2. nitelik için 0.188 ile 0.316 ve 3. nitelik için 0.002 ile 0.125 arasında değişmektedir. Madde sayısındaki değişimin yanlılık kestirimleri üzerindeki etkisinin



niteliklere göre deđiřtiđi grlmektedir. Test uzunluđunun artması ile 1. niteliđin tm dzeyleri iin yanlılık deđerleri 0.009-0.041 arasında dřerken, 2. ve 3. niteliklere ait yanlılık kestirimleri sırasıyla 0.023-0.053 ve 0.010-0.054 arasında artmıřtır. Bunun aksine, nitelikler arası korelasyonun artması 2. ve 3. niteliklerin tm dzeyleri iin kestirilen yanlılık deđerleri azalırken, 1. nitelik iin kestirilen yanlılık deđerleri artma eđilimi gstermiřtir. Q-matrisi yapısının yanlılık zerindeki etkisi incelendiđinde, 1. nitelik iin nitelikler arası korelasyonun 0.1 olduđu kořullarda gerek verinin faktr yapısına gre oluřturulan Q-matrisinin daha kk yanlılık deđerine sahip olduđu grlmektedir. 2. ve 3. nitelikler iin ise niteliklerin eřit sayıda lldđ Q-matrisinin kullanılması ile HO-DINA'nın daha yansız kestirimler yaptığı gzlenmiřtir. Q-matrisi yapısındaki deđiřimin yanlılık kestirimleri zerindeki etkisi ise test uzunluđunun 15 madde olduđu ve nitelikler arası korelasyonun daha dřk olduđu kořullarda daha fazladır.

### ** Nitelikten Oluřan Tek Boyutlu Veri Setlerinin Yanlılık Deđerleri: zet**

Tablo 17'deki verilere gre  nitelikten oluřan tek boyutlu veri setleri iin testteki madde sayısı, nitelikler arası korelasyon ve Q-matrisi yapısı kořullarının tm dzeylerine ait yanlılık deđerleri karřılařtırıldıđında, alıřma verilerinin retildiđi 3PL B-MTK modelinin her  nitelik bazında da diđer modellere gre daha yansız kestirimler yaptığı grlmektedir. Telafi edici olmayan modellerden DINA ve HO-DINA modellerinin ise tm nitelikler iin olduka benzer ve 2. nitelik iin diđer modellere gre daha yksek yanlılık deđerlerine sahip olduđu ve bunu R-RUM'un takip ettiđi gzlenmiřtir. 1. ve 3. nitelikler iin kestirilen yanlılık deđerleri incelendiđinde ise DINO modelinin diđer modellere gre olduka yanlı kestirimler yaptığı tespit edilmiřtir. Tm BTM'lerin 2. nitelik iin kestirdikleri yanlılık deđerlerinin pozitif olması ise bireylerin kestirilen sınıflama deđerlerinin gerek deđerlerinden daha byk olduđunu gstermektedir.

Tablo 17'ye gre test uzunluđunun kestirilen ve gerek birey parametreleri arasındaki farka etkisi incelendiđinde, madde sayısının 15'ten 30'a ıkarılması ile yanlılık deđerlerinin dřme eđiliminde olduđu gzlenmektedir. 2. nitelik iin BTM'lere gre kestirilen yanlılık deđerleri ise madde sayısının artmasına gre srekli bir artıř ya da azalıř gstermemektedir. Testteki madde sayısı ve Q-matrisi yapısı kořulları sabit tutulduđunda, nitelikler arası korelasyonun artması ile kestirilen

yanlılık değerleri ise genel olarak düşmektedir. Modellerin yanlılık değerleri Q-matrisi yapısına göre incelendiğinde, her bir nitelik ve nitelik profilinin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda 2. ve 3. nitelikler için kestirilen yanlılık değerlerinin daha küçük olduğu, 1. nitelik için ise gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda daha küçük yanlılık kestirimleri yapıldığı bulunmuştur. Bu durum nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi kullanıldığında 1. niteliği ölçen madde sayısının azalması, 2. ve 3. nitelikleri ölçen madde sayılarının ise artmasından kaynaklanabilir.

## **İkinci Alt Probleme İlişkin Bulgular ve Yorumlar**

### **Altı Nitelikten Oluşan İki Boyutlu Veri Setleri için Birey Parametrelerine İlişkin Hata Değerleri**

Her bir boyutun üç niteliği ölçtüğü iki boyutlu veri setleri için kestirilen hata değerleri testteki madde sayısı, nitelikler arası korelasyon ve Q-matrisi yapısı koşullarına göre 3PL ÇB-MTK ve BTM'ler kapsamında incelenmiş ve her bir modeldeki değişim Tablo 18'de sunulmuştur.

Altı nitelik için kestirilen birey parametrelerine ilişkin hata değerlerindeki değişim, her bir model için Şekil 18-Şekil 25 üzerinde gösterilmiştir. Hata değerlerindeki değişimler 0.35 ile 0.55 değerleri arasında ölçeklenmiş olmak üzere x-eksenine, boyutlar arası korelasyon değerleri (0.1, 0.4, 0.7) ise y-eksenine yerleştirilmiştir. Grafiklerde  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ ,  $\alpha_4$ ,  $\alpha_5$  ve  $\alpha_6$  nitelikleri aşağıdan yukarı doğru sıralanmış ve her niteliğin altında Q-matrisi yapısı (E ve G) ile test uzunluğu (15 ve 30) yer almıştır. Her bir nitelikteki değişim madde sayısı, nitelikler arası korelasyon ve Q-matrisi yapısı koşullarına göre ayrı ayrı ele alınmıştır.

Tablo 18

*Altı Nitelikten Oluşan İki Boyutlu Veri Setleri için Birey Parametrelerine İlişkin Hata Değerleri*

MS	Q-Matrisi	Korelasyon	ÇB-MTK						DINA					
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$
15	G	0.1	0.394	0.407	0.487	0.395	0.409	0.488	0.409	0.589	0.536	0.410	0.589	0.534
		0.4	0.382	0.391	0.459	0.383	0.391	0.459	0.396	0.535	0.487	0.398	0.536	0.484
		0.7	0.356	0.360	0.410	0.356	0.359	0.411	0.369	0.486	0.420	0.370	0.487	0.419
	E	0.1	0.447	0.400	0.428	0.448	0.402	0.428	0.492	0.538	0.449	0.492	0.535	0.449
		0.4	0.422	0.385	0.413	0.425	0.387	0.412	0.453	0.509	0.430	0.454	0.506	0.428
		0.7	0.381	0.355	0.378	0.384	0.355	0.378	0.397	0.474	0.388	0.398	0.472	0.388
30	G	0.1	0.340	0.352	0.424	0.338	0.350	0.427	0.361	0.563	0.486	0.360	0.561	0.487
		0.4	0.333	0.344	0.408	0.332	0.343	0.407	0.353	0.528	0.453	0.353	0.527	0.450
		0.7	0.316	0.325	0.372	0.315	0.324	0.371	0.333	0.491	0.397	0.331	0.490	0.395
	E	0.1	0.374	0.343	0.370	0.374	0.342	0.368	0.409	0.553	0.428	0.410	0.546	0.426
		0.4	0.362	0.337	0.360	0.362	0.336	0.357	0.392	0.525	0.407	0.394	0.523	0.403
		0.7	0.339	0.319	0.336	0.338	0.318	0.335	0.360	0.492	0.365	0.359	0.489	0.362
MS	Q-Matrisi	Korelasyon	DINO						A-CDM					
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$
15	G	0.1	0.467	0.436	0.566	0.469	0.436	0.567	0.416	0.474	0.499	0.420	0.479	0.500
		0.4	0.448	0.421	0.507	0.450	0.422	0.508	0.403	0.456	0.467	0.407	0.462	0.468
		0.7	0.411	0.408	0.435	0.413	0.408	0.438	0.382	0.436	0.416	0.383	0.439	0.416
	E	0.1	0.575	0.424	0.470	0.578	0.427	0.470	0.464	0.468	0.437	0.469	0.466	0.436
		0.4	0.523	0.422	0.442	0.529	0.424	0.442	0.438	0.455	0.420	0.445	0.456	0.418
		0.7	0.450	0.413	0.398	0.456	0.412	0.397	0.404	0.434	0.388	0.413	0.437	0.387
30	G	0.1	0.435	0.391	0.503	0.436	0.391	0.506	0.365	0.456	0.435	0.365	0.453	0.438
		0.4	0.421	0.391	0.466	0.423	0.391	0.467	0.357	0.449	0.416	0.358	0.447	0.415
		0.7	0.387	0.393	0.405	0.389	0.392	0.406	0.343	0.439	0.378	0.343	0.437	0.377
	E	0.1	0.490	0.381	0.444	0.494	0.381	0.446	0.398	0.448	0.380	0.398	0.444	0.378
		0.4	0.462	0.382	0.418	0.463	0.383	0.419	0.386	0.444	0.367	0.385	0.442	0.364
		0.7	0.414	0.386	0.372	0.415	0.386	0.373	0.365	0.437	0.342	0.363	0.432	0.340

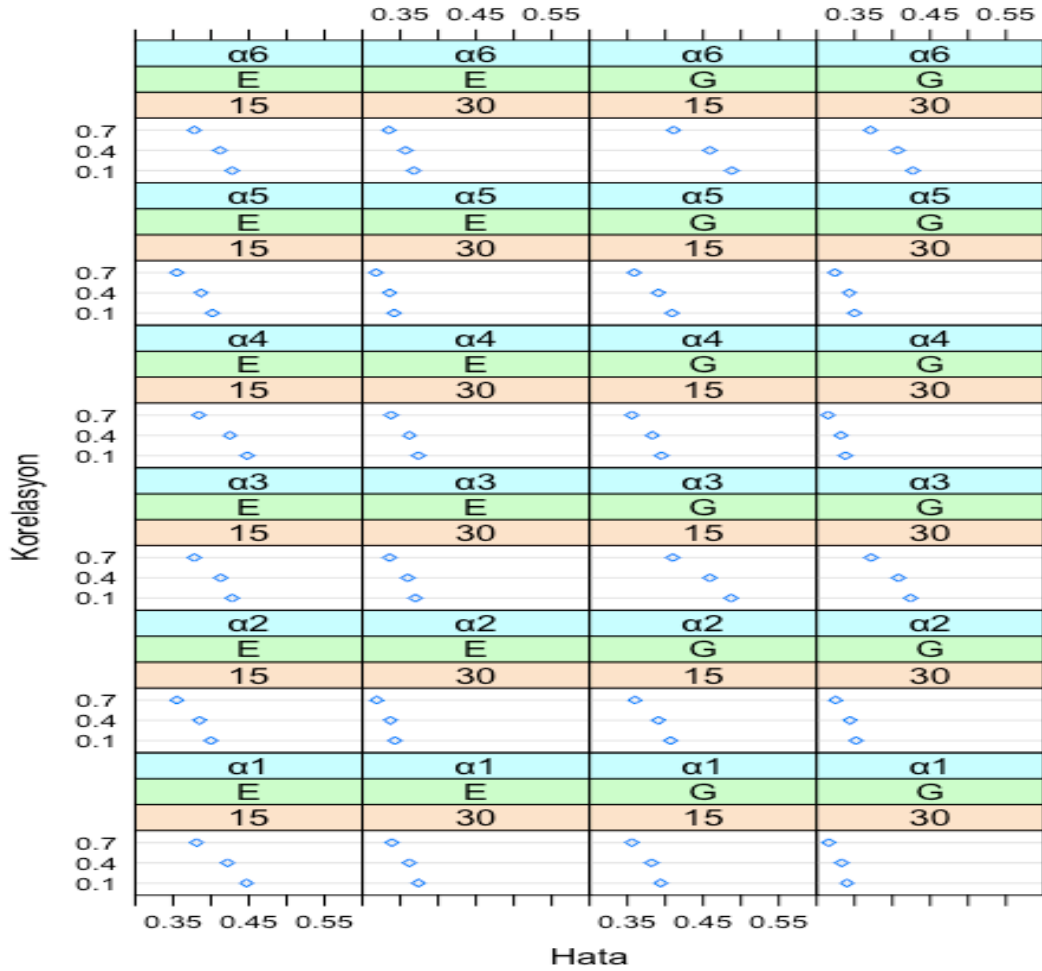
*MS: Madde sayısı, G: Gerçek verinin faktör yapısına göre oluşturulan Q-matrisi, E: Nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi*

Tablo 18'in devamı

MS	Q-Matrisi	Korelasyon	LLM						R-RUM					
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$
15	G	0.1	0.423	0.451	0.500	0.427	0.452	0.501	0.410	0.508	0.500	0.415	0.512	0.499
		0.4	0.410	0.433	0.470	0.414	0.434	0.471	0.398	0.486	0.467	0.403	0.492	0.465
		0.7	0.388	0.412	0.418	0.390	0.410	0.418	0.379	0.462	0.416	0.381	0.464	0.415
	E	0.1	0.485	0.441	0.440	0.491	0.443	0.439	0.457	0.493	0.435	0.460	0.493	0.435
		0.4	0.459	0.436	0.422	0.465	0.438	0.421	0.430	0.470	0.419	0.434	0.473	0.418
		0.7	0.427	0.424	0.388	0.435	0.426	0.388	0.392	0.436	0.388	0.399	0.443	0.387
30	G	0.1	0.374	0.433	0.437	0.374	0.431	0.440	0.358	0.486	0.439	0.357	0.484	0.441
		0.4	0.367	0.424	0.417	0.368	0.423	0.417	0.351	0.477	0.420	0.351	0.475	0.418
		0.7	0.352	0.415	0.379	0.352	0.414	0.378	0.339	0.463	0.381	0.338	0.461	0.379
	E	0.1	0.408	0.427	0.381	0.409	0.424	0.381	0.389	0.477	0.383	0.390	0.473	0.380
		0.4	0.396	0.421	0.368	0.397	0.421	0.367	0.377	0.469	0.370	0.377	0.467	0.366
		0.7	0.374	0.416	0.343	0.374	0.413	0.342	0.357	0.456	0.344	0.355	0.452	0.342
MS	Q-Matrisi	Korelasyon	G-DINA						HO-DINA					
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$
15	G	0.1	0.418	0.461	0.497	0.423	0.462	0.498	0.408	0.598	0.534	0.409	0.600	0.532
		0.4	0.405	0.445	0.465	0.410	0.446	0.466	0.395	0.543	0.485	0.397	0.545	0.482
		0.7	0.384	0.426	0.414	0.386	0.425	0.415	0.369	0.492	0.420	0.370	0.493	0.418
	E	0.1	0.472	0.453	0.436	0.475	0.456	0.436	0.493	0.543	0.448	0.493	0.539	0.448
		0.4	0.446	0.445	0.419	0.452	0.447	0.418	0.453	0.512	0.430	0.454	0.510	0.428
		0.7	0.415	0.429	0.387	0.421	0.430	0.386	0.396	0.475	0.388	0.397	0.473	0.387
30	G	0.1	0.369	0.444	0.433	0.369	0.441	0.436	0.360	0.570	0.486	0.359	0.567	0.487
		0.4	0.361	0.436	0.414	0.362	0.434	0.413	0.352	0.534	0.453	0.352	0.533	0.451
		0.7	0.347	0.427	0.376	0.346	0.426	0.375	0.332	0.495	0.396	0.331	0.494	0.395
	E	0.1	0.401	0.437	0.378	0.401	0.434	0.377	0.408	0.556	0.429	0.409	0.550	0.426
		0.4	0.388	0.433	0.365	0.389	0.432	0.363	0.391	0.529	0.407	0.392	0.526	0.403
		0.7	0.367	0.426	0.341	0.367	0.424	0.340	0.359	0.494	0.365	0.358	0.491	0.362

MS: Madde sayısı, G: Gerçek verinin faktör yapısına göre oluşturulan Q-matrisi, E: Nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi

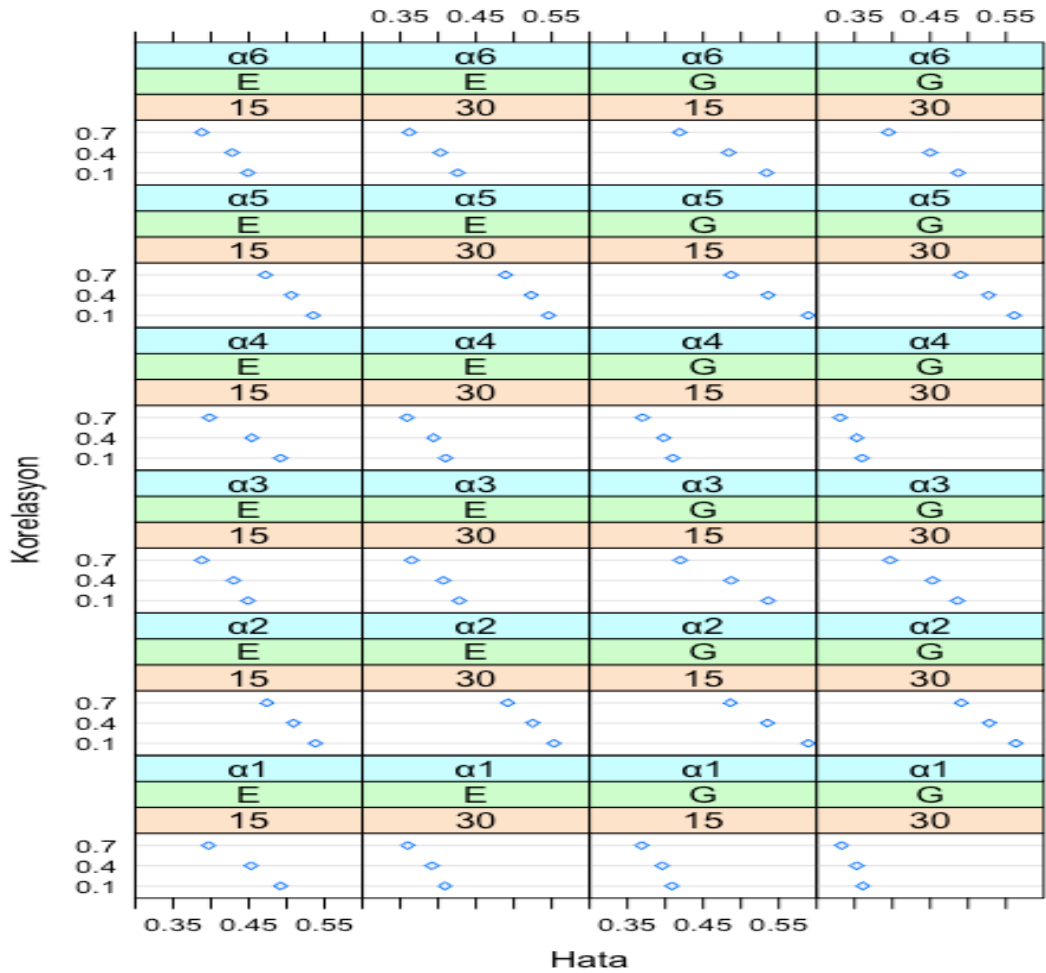
3PL ÇB-MTK kapsamında birey parametre kestirimlerine ilişkin hata değerlerindeki değişim, iki boyutlu veri setleri için testteki madde sayısı, nitelikler arası korelasyon ve Q-matrisinin yapısı koşullarına göre incelenmiş ve Şekil 18’de sunulmuştur.



Şekil 18. Altı nitelikten oluşan iki boyutlu veri setleri için ÇB-MTK'ya göre kestirilen birey parametrelerine ilişkin hata değerleri.

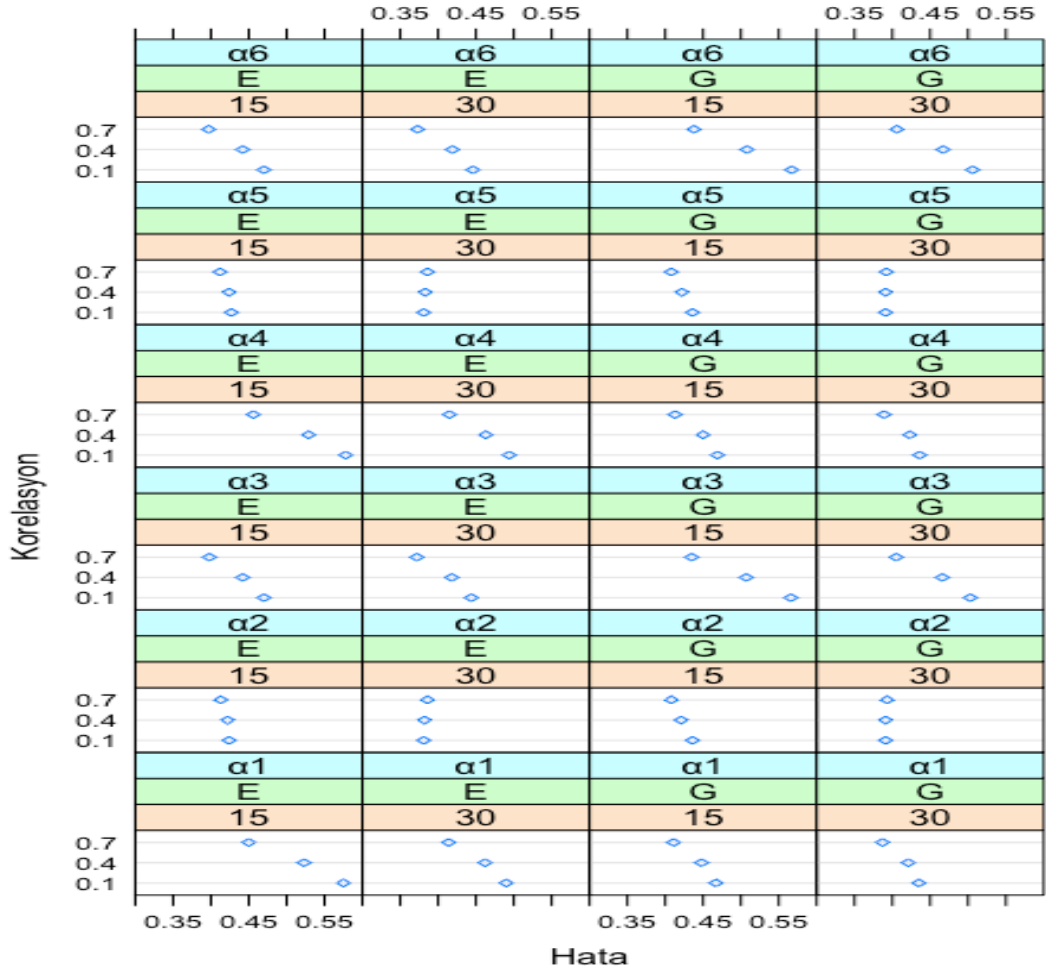
3PL ÇB-MTK'ya göre kestirilen hata değerleri 1. nitelik için 0.316-0.447, 2. nitelik için 0.319-0.407, 3. nitelik için 0.336-0.487, 4. nitelik için 0.315-0.448, 5. nitelik için 0.318-0.409 ve 6. nitelik için 0.335-0.488 arasında değişmektedir. 6 nitelik bazında en küçük hata kestirimleri test uzunluğunun 30 madde ve nitelikler arası korelasyonun 0.7 olduğu koşullarda, en yüksek hata değerleri ise test uzunluğunun 15 madde ve nitelikler arası korelasyonun 0.1 olduğu koşullarda elde etmiştir. Madde sayısının hata üzerine etkisi incelendiğinde, test uzunluğunun 15'ten 30'a çıkarılması ile tüm düzeyler için hata miktarının azaldığı görülmektedir. Nitelikler

arası korelasyonun düşmesi ile 15-30 maddelik test uzunluğu koşullarına göre kestirilen hata değerleri arasındaki fark, artmaktadır. Benzer şekilde, nitelikler arasındaki korelasyonun artması, 3PL ÇB-MTK'ya göre kestirilen hata değerlerinin tüm nitelik düzeyleri için düşmesine neden olmaktadır. Hata değerlerindeki azalma ise korelasyon düzeyinin 0.4-0.7 arasında değiştiği koşullarda 0.1-0.4 düzeylerindeki değişime göre daha fazladır. Q-matrisi yapısına göre hata kestirimindeki değişimler karşılaştırıldığında, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile 1. ve 4. nitelik için kestirilen hata değerlerinin arttığı, diğer 4 nitelik için ise hataların azaldığı görülmektedir. Hata kestirimleri arasındaki bu fark, nitelikler arası korelasyonun 0.1 olduğu koşullarda en büyük iken korelasyonun artması ile azalmaktadır.



Şekil 19. Altı nitelikten oluşan iki boyutlu veri setleri için DINA'ya göre kestirilen birey parametrelerine ilişkin hata değerleri.

Son derece sınırlı BTM'lerden DINA modeline göre kestirilen hata değerleri 1. nitelik için 0.333-0.492, 2. nitelik için 0.474-0.589, 3. nitelik için 0.365-0.536, 4. nitelik için 0.331-0.492, 5. nitelik için 0.472-0.589 ve 6. nitelik için 0.362-0.534 arasında değişmektedir. DINA modeli, tüm koşullar için en yüksek hata kestirimlerini ise 2. ve 5. nitelikler için yapmıştır. Madde sayısına göre hata değerlerindeki değişim incelendiğinde, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı tüm korelasyon düzeylerinde madde sayısı arttıkça 2. ve 5. nitelikler için kestirilen hata miktarının arttığı gözlenmiştir. Diğer niteliklerin tüm düzeyleri için ise madde sayısının artması ile hata miktarının azaldığı görülmektedir. Gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda madde sayısının artması ile hatalarda görülen düşüş, nitelikler arası korelasyonun artması ile azalmaktadır. Nitelikler arası korelasyonun artması ise 6 niteliğin tüm düzeylerinde kestirilen hata miktarını 0.007-0.067 arasında düşürmüştür. Korelasyon düzeyindeki değişimlere göre hata miktarında gözlenen düşüşler, test uzunluğunun 15 madde olduğu koşullarda ise daha fazladır. Q-matrisi yapısına göre hata değerleri karşılaştırıldığında, gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılması ile 1. ve 4. nitelikler için kestirilen hata değerlerinin düştüğü görülmektedir. Hata değerlerindeki bu düşüş test uzunluğunun 15 madde olduğu koşullarda daha fazla olmakla birlikte, nitelikler arası korelasyonun artması ile azalmaktadır. Diğer tüm koşullar için ise niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılmasının hatayı 0.001 ile 0.087 arasında azalttığı görülmektedir.

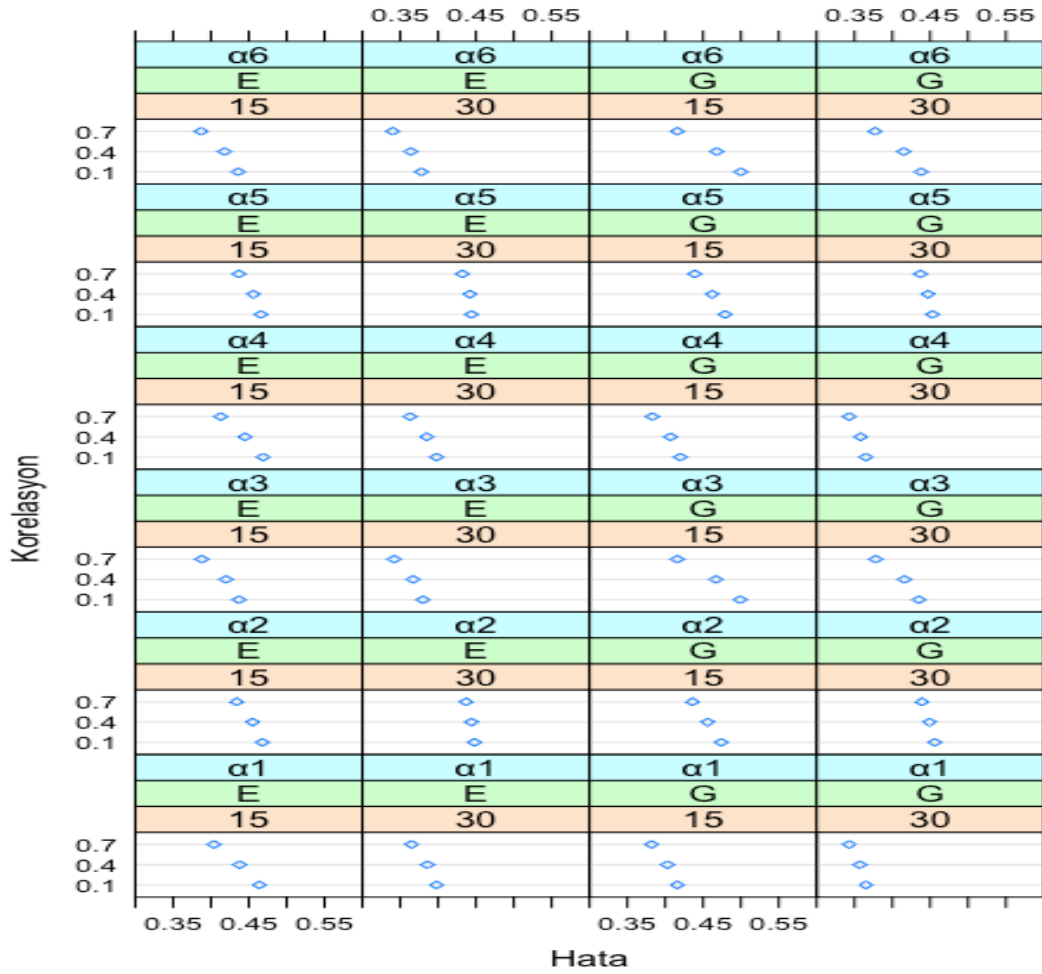


Şekil 20. Altı nitelikten oluşan iki boyutlu veri setleri için DINO'ya göre kestirilen birey parametrelerine ilişkin hata değerleri.

Birey parametrelerinin DINO modeline göre kestirimlerinden elde edilen hata değerlerinin 1. nitelik için 0.387-0.575, 2. nitelik için 0.381-0.436, 3. nitelik için 0.372-0.566, 4. nitelik için 0.389-0.578, 5. nitelik için 0.381-0.436 ve 6. nitelik için 0.373-0.567 arasında değiştiği görülmektedir. Test uzunluğunun 15'ten 30'a çıkarılmasının, tüm nitelik düzeyleri için kestirilen hata miktarını düşürdüğü gözlenmiştir. Madde sayısındaki artışa bağlı olarak hata miktarında gözlenen düşüşler, nitelikler arası korelasyonun artması ile azalmaktadır. Nitelikler arası korelasyonun artması ile test uzunluğunun 30 madde olduğu koşullarda 2. ve 5. nitelikler için kestirilen hata değerinin 0.001 ile 0.003 arasında arttığı görülmektedir. Diğer tüm nitelik düzeyleri için ise nitelikler arası korelasyonun artması ile hata düşmektedir. 0.4-0.7 korelasyon düzeyleri için kestirilen hata değerleri arasındaki farkın ise 0.1-0.4 korelasyon düzeylerine göre daha fazla olduğu tespit edilmiştir. Q-



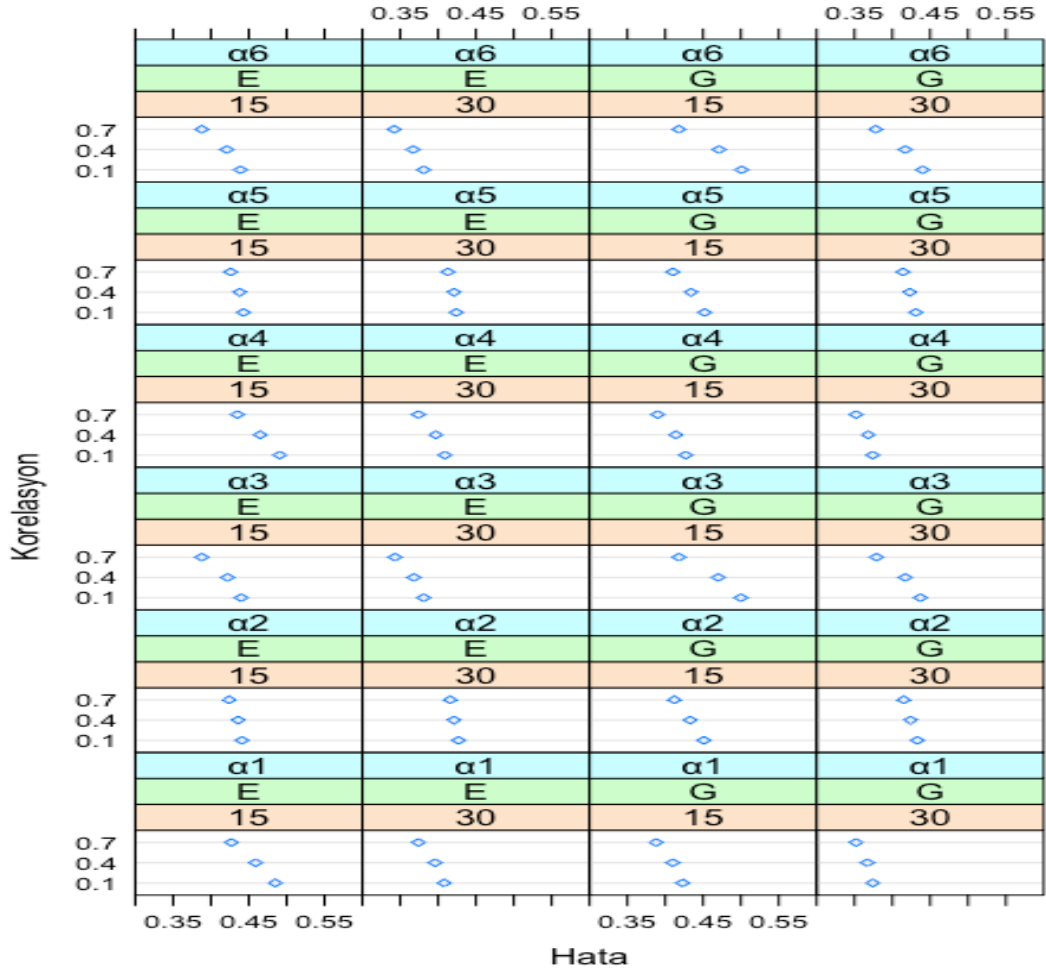
matrisi yapısının 1., 3., 4. ve 6. nitelikler için kestirilen hatalar üzerindeki etkisi oldukça büyüktür. Niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması, 1. ve 4. niteliklerin tüm düzeyleri için hatayı 0.026 ile 0.109 arasında; 2. ve 5. niteliklerin ise sadece test uzunluğunun 15 madde ve nitelikler arası korelasyonun 0.4-0.7 olduğu koşullar için hatayı 0.001 ile 0.004 arasında arttırmaktadır. Diğer nitelik düzeylerinde, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile kestirilen hata miktarı azalmaktadır.



Şekil 21. Altı nitelikten oluşan iki boyutlu veri setleri için A-CDM'ye göre kestirilen birey parametrelerine ilişkin hata değerleri.

Telafi edici BTM'lerden A-CDM modeli ile kestirilen hata değerlerinin 1. nitelik için 0.343-0.464, 2. nitelik için 0.434-0.474, 3. nitelik için 0.342-0.499, 4. nitelik için 0.343-0.469, 5. nitelik için 0.432-0.479 ve 6. nitelik için 0.340-0.500 arasında değiştiği görülmektedir. Nitelikler bazında en düşük hata kestirimleri madde sayısının 30 ve nitelikler arası korelasyonun 0.7 olduğu koşullarda, en yüksek hata

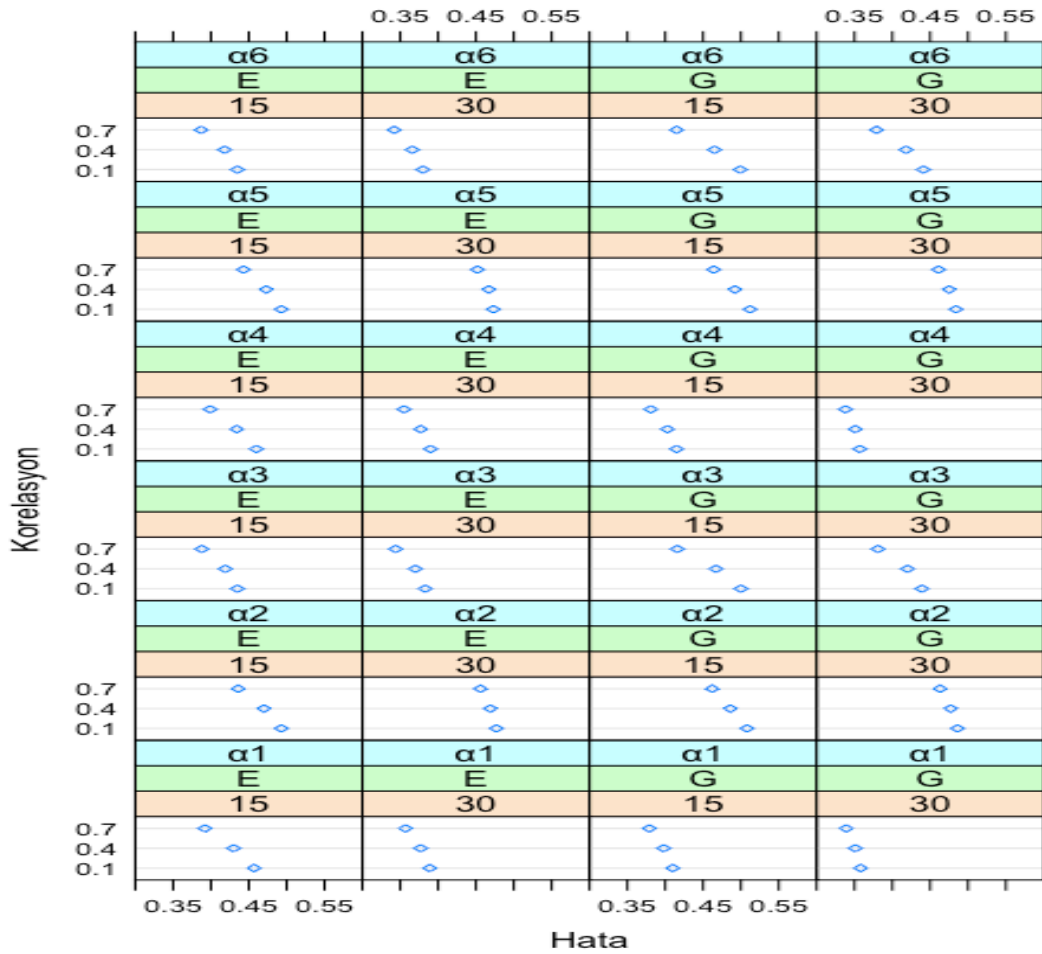
kestirimleri ise test uzunluğunun 15 ve nitelikler arası korelasyonun 0.1 olduğu koşullarda elde edilmiştir. Test uzunluğunun artması, 2. nitelik için nitelikler arası korelasyonun 0.7 olduğu koşullarda hata değerini 0.002-0.003 arasında arttırmıştır. Diğer tüm koşul ve nitelikler için ise madde sayısının artması, hatayı düşürmüştür. Madde sayısının artması ile hata miktarındaki azalma ise nitelikler arasındaki korelasyonun artması ile azalmaktadır. Madde sayısının kestirilen hata değerleri üzerindeki etkisi ise 1., 3., 4. ve 6. nitelikler için 2. ve 5. niteliklere göre daha fazladır. Diğer koşullar sabit tutulduğunda nitelikler arası korelasyonun artması, tüm nitelik düzeylerinde hata miktarını düşürmüştür. Diğer modellerin sonuçları ile benzer şekilde, orta-yüksek korelasyon düzeyleri için kestirilen hata değerleri arasındaki farkın, düşük-orta korelasyon düzeyleri arasındaki farktan daha fazla olduğu bulunmuştur. Q-matrisi yapısının kestirilen hata değerleri üzerindeki etkisi incelendiğinde, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılmasının 2., 3., 5. ve 6. niteliklerin tüm düzeyleri için kestirilen hata miktarını azalttığı, 1. ve 4. nitelik için ise hatayı 0.020-0.050 arasında arttırdığı görülmektedir. Tanımlanan Q-matrisi yapısının hata kestirimi üzerindeki etkisi ise 2. ve 5. nitelik için diğer niteliklere göre daha düşüktür.



Şekil 22. Altı nitelikten oluşan iki boyutlu veri setleri için LLM'ye göre kestirilen birey parametrelerine ilişkin hata değerleri.

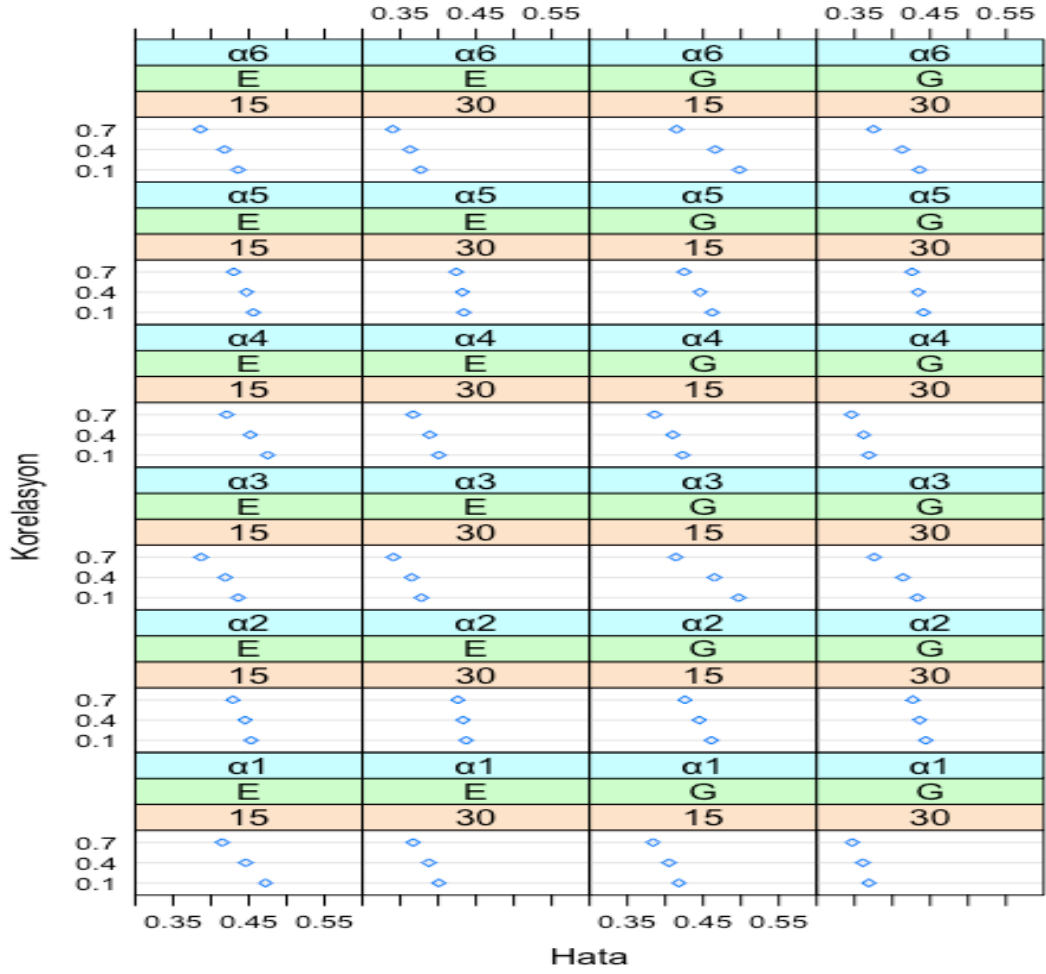
Eklemeli BTM'lerden LLM modeli için kestirilen hata değerlerinin 1. nitelik için 0.352-0.485, 2. nitelik için 0.412-0.451, 3. nitelik için 0.343-0.500, 4. nitelik için 0.352-0.491, 5. nitelik için 0.410-0.452 ve 6. nitelik için 0.342-0.501 arasında değiştiği görülmektedir. 6 nitelik için kestirilen hata değerleri incelendiğinde, en küçük hata değerlerinin nitelikler arası korelasyonun 0.7 olduğu ve en büyük hataların nitelikler arası korelasyonun 0.1 olduğu koşullarda elde edildiği gözlenmiştir. Testteki madde sayısının artması 2. ve 5. nitelikler için gerçek verinin faktör yapısına göre tanımlanan Q-matrisinin kullanıldığı ve nitelikler arası korelasyonun 0.7 olduğu koşullarda hata değerlerini sırasıyla 0.003 ve 0.004 arttırmıştır. Diğer nitelik düzeyleri için ise madde sayısının 15'ten 30'a çıkarılması ile hata miktarı azalmıştır. Madde sayısının artması ile hata kestirimlerinde en büyük azalma, Q-matrisi yapısından bağımsız olarak, nitelikler arasında 0,1 korelasyonun

olduğu koşullarda görülmektedir. Korelasyon düzeyinin artması ise madde sayısının hata kestirimi üzerindeki etkisini azaltmaktadır. Benzer şekilde, nitelikler arasındaki korelasyonun artmasının kestirilen hata değerlerini düşürdüğü ve 0.4-0.7 korelasyon düzeyleri arasındaki hata farklarının 0.1-0.4 korelasyon düzeylerinde yapılan kestirimler arası farktan daha fazla olduğu gözlenmiştir. Q-matrisi yapısının hata üzerindeki etkisi incelendiğinde, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılmasının 1. ve 4. niteliklerin tüm koşulları için kestirilen hata miktarını 0.022-0.064 arasında; 2. ve 5. niteliklerin ise test uzunluğunun 15 madde ve nitelikler arası korelasyonun 0.4 ve 0.7 olduğu koşullarda hatayı 0.003-0.015 arasında arttırdığı görülmektedir. 3. ve 6 nitelik için kestirilen hata değerleri ise niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile 0.030 ile 0.062 arasında azaltmaktadır. Q-matrisi yapısına göre kestirilen hata değerlerindeki değişim ise nitelikler arasındaki korelasyon arttıkça azalma eğilimi göstermektedir.



Şekil 23. Altı nitelikten oluşan iki boyutlu veri setleri için R-RUM'a göre kestirilen birey parametrelerine ilişkin hata değerleri.

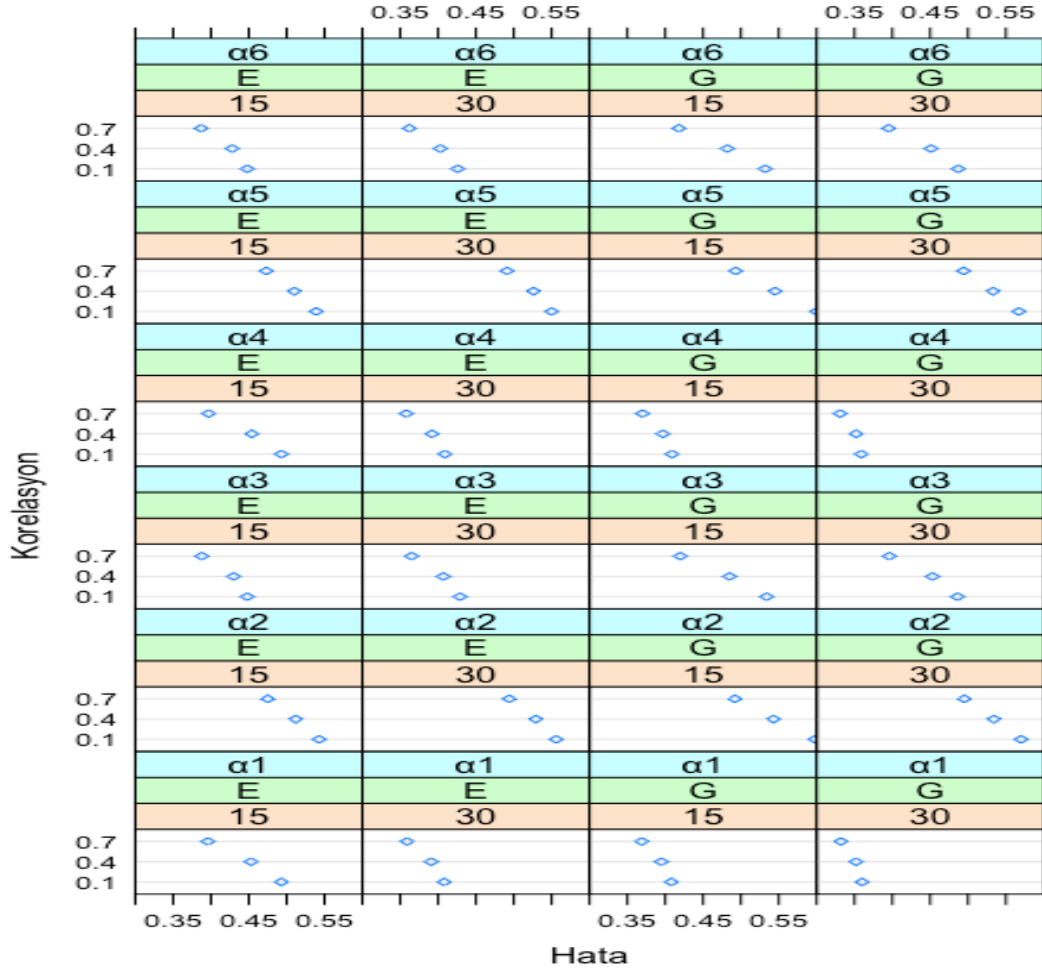
İki boyutlu veri setleri için R-RUM modeline göre kestirilen hata değerlerinin 1. nitelik için 0.339-0.457, 2. nitelik için 0.436-0.508, 3. nitelik için 0.344-0.500, 4. nitelik için 0.338-0.460, 5. nitelik için 0.443-0.512 ve 6. nitelik için 0.342-0.499 arasında değiştiği gözlenmektedir. Tüm koşullar için hata değerleri incelendiğinde, test uzunluğunun 15 madde ve nitelikler arası korelasyonun 0.1 olduğu koşullarda tüm niteliklerin en yüksek hata değerine sahip olduğu görülmektedir. Test uzunluğunun artması ile 1., 3., 4. ve 6. niteliklerin tüm düzeyleri için hata değerleri azalmaktadır. 2. nitelik için nitelikler arası korelasyonun 0.7 olduğu, 5. nitelik için niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı ve nitelikler arası korelasyonun 0.7 olduğu koşullarda kestirilen hata değerleri ise madde sayısının artması ile artmaktadır. Madde sayısındaki artışa bağlı olarak hatalarda gözlenen değişim, nitelikler arası korelasyon düzeyinin artması ile azalma eğilimi göstermektedir. Nitelikler arası korelasyonun 0.1'den 0.7'ye doğru artması, tüm nitelik düzeyleri için kestirilen hatayı düşürmüştür. Diğer modellerin yaptığı kestirimler ile benzer şekilde, nitelikler arasındaki korelasyonun artması doğrultusunda hata değerlerinde gözlenen değişim 0.4-0.7 korelasyon düzeyleri arasında 0.1-0.4 düzeylerine göre daha fazladır. Ayrıca madde sayısının artması ile hata miktarında gözlenen düşüşler, test uzunluğunun 15 madde olduğu koşullarda 30 madde olduğu koşullara göre ve niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullara göre daha fazladır. Q-matrisi yapısının hata üzerindeki etkisi incelendiğinde, niteliklerin eşit sayıda ölçüldüğü Q-matrisi koşulunda 1. ve 4. nitelik için kestirilen hataların arttığı, diğer 4 niteliğin tüm düzeyleri için kestirilen hataların ise azaldığı görülmektedir. Genel olarak, Q-matrisi yapısının kestirilen hata değerleri üzerindeki etkisinin test uzunluğunun 15 madde ve nitelikler arası korelasyonun düşük olduğu koşullarda ise daha yüksek olduğu bulunmuştur.



Şekil 24. Altı nitelikten oluşan iki boyutlu veri setleri için G-DINA'ya göre kestirilen birey parametrelerine ilişkin hata değerleri.

G-DINA modeline göre kestirilen hata değerleri 1. nitelik için 0.347-0.472, 2. nitelik için 0.426-0.461, 3. nitelik için 0.341-0.497, 4. nitelik için 0.346-0.475, 5. nitelik için 0.424-0.462 ve 6. nitelik için 0.340-0.498 arasında değişmektedir. 6 nitelik bazında kestirilen hatalar incelendiğinde, test uzunluğunun 15 madde ve nitelikler arası korelasyonun 0.1 olduğu koşullarda daha yüksek RMSE değerleri gözlenmektedir. Test uzunluğunun 30 madde ve nitelikler arası korelasyon düzeyinin 0.7 olduğu koşullarda ise en küçük hata değerleri elde edilmektedir. Test uzunluğunun kestirilen hata değerleri üzerindeki üzerindeki etkisi incelendiğinde, madde sayısının artması ile hata değerlerinin azaldığı görülmektedir. Hata değerlerindeki en büyük düşüş ise 0.040-0.074 arasındaki değişimler ile 4. nitelikte gözlenmektedir. Benzer şekilde, nitelikler arası korelasyonun artması, tüm nitelik düzeyleri için kestirilen hata miktarını düşürmektedir. Diğer koşullar sabit tutulduğunda, orta-yüksek korelasyon düzeyleri arasındaki hata kestirimlerindeki

değişim ise düşük-orta korelasyon düzeylerine göre daha fazladır. Gerçek verinin faktör yapısına göre tanımlanan Q-matrisinin kullanılması 1. ve 4. nitelikler için kestirilen hata değerlerini 0.020-0.053 arasında düşürürken, diğer nitelikler için kestirilen hata değerlerini arttırmaktadır. Q-matrisi yapısına göre 2. ve 5. nitelikler için hata değerlerinde gözlenen değişim 0.001-0.008 arasında, 3. ve 6. nitelikler için ise 0.027-0.061 arasında bir değer almaktadır.



Şekil 25. Altı nitelikten oluşan iki boyutlu veri setleri için HO-DINA'ya göre kestirilen birey parametrelerine ilişkin hata değerleri.

HO-DINA'ya göre kestirilen hata değerlerinin 1. nitelik için 0.332-0.493, 2. nitelik için 0.475-0.598, 3. nitelik için 0.365-0.534, 4. nitelik için 0.331-0.493, 5. nitelik için 0.473-0.600 ve 6. nitelik için 0.362-0.532 arasında değiştiği görülmektedir. Testteki madde sayısının artması 1., 3., 4. ve 6. niteliklerin tüm düzeyleri için kestirilen hata değerlerini önemli ölçüde düşürmektedir. 1. ve 4. nitelikler kapsamında hata değerlerindeki değişim, niteliklerin eşit sayıda ölçüldüğü Q-

matrisinin kullanıldığı ve nitelikler arası korelasyonun daha düşük olduğu koşullarda; 3. ve 6. nitelikler için ise gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda daha büyüktür. Madde sayısının artması ile 2. ve 5. nitelikler için nitelikler arası korelasyonun 0.7 olduğu veya niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda kestirilen hata değerlerinin daha fazla olduğu görülmektedir. Nitelikler arası korelasyonun artmasının ise hatayı düşürdüğü ve hata miktarındaki değişiminin 0.4-0.7 korelasyon düzeyleri arasında daha fazla olduğu bulunmuştur. Q-matrisi yapısının hata kestirimi üzerindeki etkisi değerlendirildiğinde, gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılmasının 1. ve 4. niteliklerin tüm düzeyleri için hata miktarını düşürdüğü ve bu değişimin test uzunluğunun 15 madde veya nitelikler arası korelasyonun daha düşük olduğu koşullarda daha büyük olduğu görülmektedir. Niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ise 2., 3., 5. ve 6. niteliklerin tüm düzeylerinde hata miktarını azaltmaktadır. Benzer şekilde hata değerlerindeki bu azalma, test uzunluğunun 15 madde ve nitelikler arası korelasyonun 0.1 olduğu koşullarda daha büyüktür.

### **Altı Nitelikten Oluşan İki Boyutlu Veri Setlerinin Hata Değerleri: Özet**

Tablo 18'de verilen iki boyutlu veri setleri için kestirilen hata değerleri incelendiğinde, en küçük hata kestirimlerini 3PL ÇB-MTK modelinin yaptığı tespit edilmiştir. 8 kestirim modelinin her bir nitelik için performansları karşılaştırıldığında 1., 3., 4. ve 6. nitelikler için DINO modelinin, 2. ve 5. nitelikler için ise HO-DINA modelinin en yüksek hata değerine sahip olduğu görülmektedir. Madde sayısının ve nitelikler arası korelasyonun artması ile 8 modele ait hata kestirimlerinin azaldığı gözlenmiştir. Nitelikler arası korelasyon düzeyinin 0.7 olduğu koşullarda modeller en düşük hata kestirimini yaparken, 0.1 olduğu koşullarda ise en yüksek hata değerlerine sahiptir. Ayrıca nitelikler arası korelasyonun hata üzerindeki etkisi incelendiğinde, 0.4-0.7 korelasyon düzeyleri için kestirilen hata değerleri arasındaki farkın, 0.1-0.4 korelasyon düzeylerine göre daha fazla olduğu görülmektedir. Son olarak, Q-matrisi yapısına göre hata kestirimindeki değişimler karşılaştırıldığında ise tüm modeller için nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile 1. ve 4. nitelikler için kestirilen hata değerlerini arttırırken, diğer 4 nitelik için hataları azaltmaktadır. Genel olarak, Q-matrisi yapısının kestirilen hata



değerleri üzerindeki etkisinin, nitelikler arası korelasyon düştükçe daha büyük olduğu gözlenmiştir.

### **Altı Nitelikten Oluşan İki Boyutlu Veri Setleri için Birey Parametrelerine İlişkin Yanlılık Değerleri**

3PL ÇB-MTK ve BTM'ler kapsamında birey parametre kestirimlerine ilişkin hata değerlerindeki değişim 6 nitelikten oluşan iki boyutlu veri setleri için testteki madde sayısı, nitelikler arası korelasyon ve Q-matrisi yapısı koşullarına göre incelenmiş ve Tablo 19'da sunulmuştur.

Altı nitelik için kestirilen birey parametrelerine ilişkin yanlılık değerlerindeki değişim, her bir model için Şekil 26-Şekil 33 üzerinde gösterilmiştir. Yanlılık değerlerindeki değişimler -0.30 ile 0.30 değerleri arasında ölçeklenmiş olmak üzere x-eksenine, boyutlar arası korelasyon değerleri (0.1, 0.4, 0.7) ise y-eksenine yerleştirilmiştir. Grafiklerde  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ ,  $\alpha_4$ ,  $\alpha_5$  ve  $\alpha_6$  nitelikleri aşağıdan yukarı doğru sıralanmış ve her niteliğin altında Q-matrisi yapısı (E ve G) ile test uzunluğu (15 ve 30) yer almıştır. Her bir nitelikteki değişim madde sayısı, nitelikler arası korelasyon ve Q-matrisi yapısı koşullarına göre ayrı ayrı ele alınmıştır.

Tablo 19

*Altı Nitelikten Oluşan İki Boyutlu Veri Setleri için Birey Parametrelerine İlişkin Yanlılık Değerleri*

MS	Q-Matrisi	Korelasyon	ÇB-MTK						DINA					
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$
15	G	0.1	0.021	-0.039	0.012	0.020	-0.042	0.015	-0.019	0.310	0.135	-0.021	0.309	0.122
		0.4	0.017	-0.035	0.012	0.016	-0.036	0.015	-0.028	0.246	0.062	-0.029	0.245	0.046
		0.7	0.009	-0.033	0.001	0.007	-0.033	0.002	-0.041	0.213	0.023	-0.042	0.213	0.016
	E	0.1	0.025	-0.033	0.012	0.026	-0.038	0.011	0.083	0.251	0.043	0.075	0.243	0.041
		0.4	0.019	-0.034	0.009	0.020	-0.034	0.007	0.018	0.222	0.024	0.009	0.217	0.022
		0.7	0.009	-0.034	-0.002	0.011	-0.034	-0.002	-0.021	0.204	0.012	-0.029	0.200	0.010
30	G	0.1	0.006	-0.029	0.005	0.004	-0.030	0.005	0.004	0.302	0.119	0.002	0.300	0.118
		0.4	-0.006	-0.034	-0.006	-0.007	-0.034	-0.006	-0.004	0.262	0.068	-0.004	0.261	0.063
		0.7	-0.015	-0.035	-0.018	-0.017	-0.035	-0.018	-0.022	0.231	0.035	-0.022	0.230	0.033
	E	0.1	0.011	-0.027	-0.001	0.009	-0.028	-0.001	0.034	0.294	0.100	0.037	0.287	0.096
		0.4	-0.003	-0.033	-0.012	-0.004	-0.033	-0.013	0.006	0.263	0.074	0.008	0.261	0.068
		0.7	-0.014	-0.036	-0.023	-0.016	-0.035	-0.022	-0.019	0.234	0.045	-0.020	0.232	0.041
MS	Q-Matrisi	Korelasyon	DINO						A-CDM					
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$
15	G	0.1	-0.175	-0.046	-0.260	-0.179	-0.046	-0.261	-0.091	0.146	-0.056	-0.098	0.153	-0.055
		0.4	-0.159	0.029	-0.168	-0.161	0.023	-0.174	-0.089	0.149	-0.046	-0.094	0.158	-0.060
		0.7	-0.130	0.096	-0.091	-0.131	0.095	-0.098	-0.084	0.154	-0.033	-0.087	0.158	-0.044
	E	0.1	-0.295	0.032	-0.139	-0.302	0.029	-0.138	-0.085	0.154	-0.047	-0.101	0.149	-0.043
		0.4	-0.228	0.081	-0.106	-0.237	0.081	-0.107	-0.084	0.155	-0.042	-0.100	0.156	-0.039
		0.7	-0.155	0.121	-0.066	-0.165	0.119	-0.067	-0.076	0.156	-0.034	-0.095	0.158	-0.029
30	G	0.1	-0.165	0.027	-0.187	-0.167	0.025	-0.195	-0.075	0.177	-0.018	-0.077	0.174	-0.025
		0.4	-0.153	0.070	-0.136	-0.155	0.068	-0.141	-0.074	0.179	-0.014	-0.076	0.177	-0.020
		0.7	-0.126	0.116	-0.074	-0.128	0.115	-0.078	-0.071	0.179	-0.011	-0.073	0.178	-0.012
	E	0.1	-0.212	0.037	-0.147	-0.217	0.039	-0.153	-0.080	0.177	-0.014	-0.079	0.172	-0.020
		0.4	-0.181	0.071	-0.114	-0.183	0.074	-0.120	-0.079	0.178	-0.012	-0.078	0.176	-0.018
		0.7	-0.139	0.113	-0.074	-0.140	0.113	-0.078	-0.077	0.180	-0.010	-0.074	0.176	-0.014

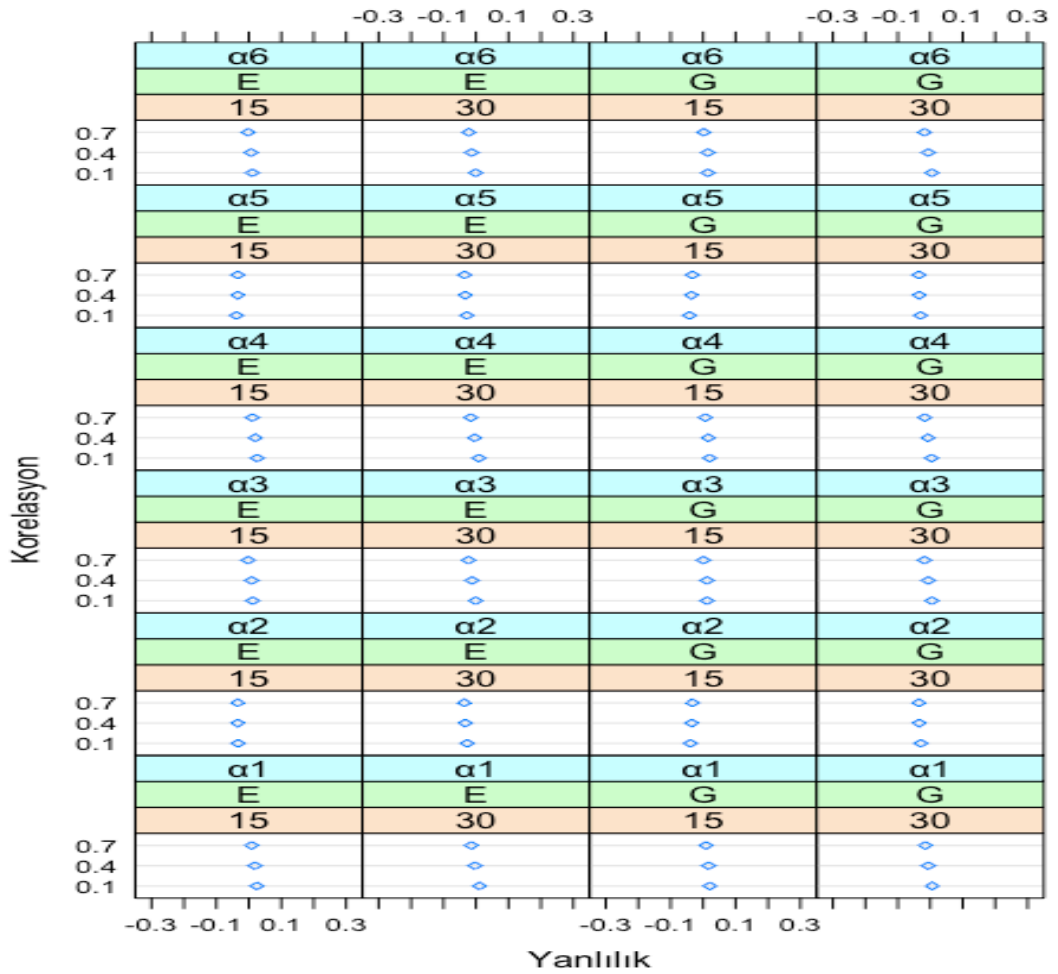
*MS: Madde sayısı, G: Gerçek verinin faktör yapısına göre oluşturulan Q-matrisi, E: Nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi*

Tablo 19'un devamı

MS	Q-Matrisi	Korelasyon	LLM						R-RUM					
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$
15	G	0.1	-0.107	0.107	-0.090	-0.115	0.106	-0.088	-0.073	0.202	0.007	-0.083	0.206	0.008
		0.4	-0.105	0.113	-0.081	-0.110	0.112	-0.090	-0.075	0.192	-0.008	-0.081	0.200	-0.020
		0.7	-0.096	0.120	-0.057	-0.099	0.117	-0.063	-0.077	0.188	-0.013	-0.079	0.191	-0.021
	E	0.1	-0.154	0.110	-0.071	-0.167	0.108	-0.070	-0.018	0.194	-0.022	-0.034	0.192	-0.019
		0.4	-0.141	0.126	-0.059	-0.152	0.128	-0.059	-0.033	0.176	-0.025	-0.051	0.180	-0.024
		0.7	-0.127	0.141	-0.043	-0.140	0.143	-0.042	-0.031	0.157	-0.030	-0.055	0.166	-0.024
30	G	0.1	-0.093	0.146	-0.050	-0.093	0.145	-0.063	-0.056	0.215	0.028	-0.057	0.212	0.021
		0.4	-0.092	0.149	-0.046	-0.094	0.148	-0.054	-0.057	0.211	0.021	-0.058	0.209	0.015
		0.7	-0.086	0.154	-0.034	-0.087	0.153	-0.039	-0.061	0.205	0.014	-0.062	0.204	0.012
	E	0.1	-0.103	0.150	-0.039	-0.104	0.147	-0.045	-0.045	0.211	0.020	-0.044	0.206	0.014
		0.4	-0.102	0.152	-0.037	-0.104	0.151	-0.044	-0.049	0.207	0.017	-0.049	0.204	0.010
		0.7	-0.096	0.158	-0.030	-0.095	0.156	-0.035	-0.056	0.200	0.011	-0.053	0.196	0.006
MS	Q-Matrisi	Korelasyon	G-DINA						HO-DINA					
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$
15	G	0.1	-0.098	0.129	-0.077	-0.106	0.127	-0.073	-0.018	0.326	0.134	-0.019	0.329	0.119
		0.4	-0.095	0.135	-0.060	-0.101	0.134	-0.069	-0.026	0.259	0.058	-0.028	0.261	0.039
		0.7	-0.089	0.140	-0.043	-0.093	0.138	-0.048	-0.040	0.221	0.027	-0.042	0.222	0.015
	E	0.1	-0.121	0.132	-0.058	-0.129	0.132	-0.057	0.101	0.259	0.046	0.093	0.252	0.044
		0.4	-0.113	0.141	-0.048	-0.124	0.143	-0.049	0.033	0.228	0.026	0.025	0.224	0.024
		0.7	-0.103	0.148	-0.036	-0.114	0.149	-0.036	-0.012	0.206	0.013	-0.022	0.202	0.011
30	G	0.1	-0.084	0.161	-0.037	-0.085	0.159	-0.049	0.006	0.312	0.125	0.004	0.310	0.125
		0.4	-0.083	0.164	-0.030	-0.084	0.163	-0.038	-0.003	0.270	0.077	-0.003	0.270	0.073
		0.7	-0.079	0.167	-0.022	-0.079	0.165	-0.026	-0.020	0.236	0.039	-0.021	0.236	0.038
	E	0.1	-0.090	0.163	-0.026	-0.091	0.160	-0.032	0.039	0.300	0.105	0.042	0.292	0.101
		0.4	-0.088	0.166	-0.023	-0.089	0.165	-0.030	0.011	0.269	0.079	0.013	0.266	0.074
		0.7	-0.084	0.169	-0.019	-0.084	0.167	-0.023	-0.017	0.237	0.048	-0.018	0.234	0.044

MS: Madde sayısı, G: Gerçek verinin faktör yapısına göre oluşturulan Q-matrisi, E: Nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi

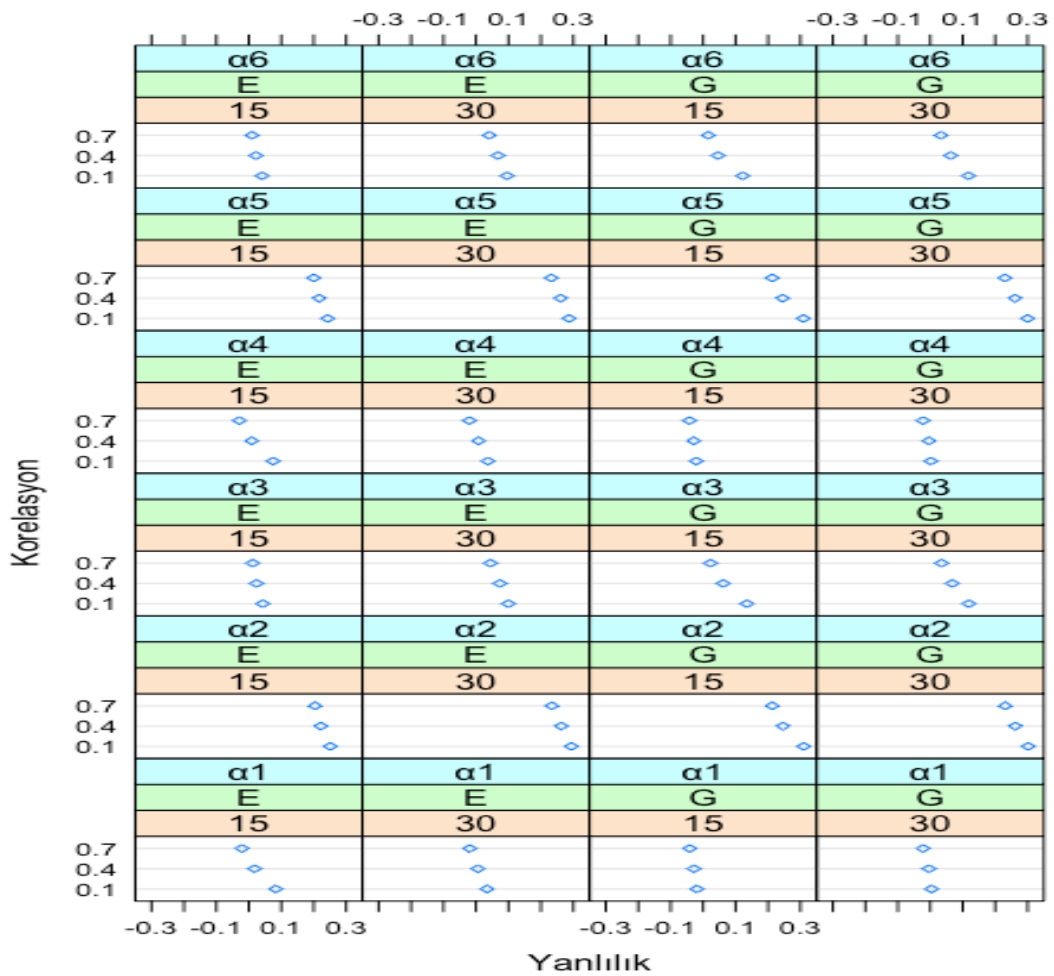
Araştırma koşullarına göre iki boyutlu veri setleri için 3PL ÇB-MTK için kestirilen yanlılık değerleri Şekil 26'da görülmektedir.



Şekil 26. Altı nitelikten oluşan iki boyutlu veri setleri için ÇB-MTK'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

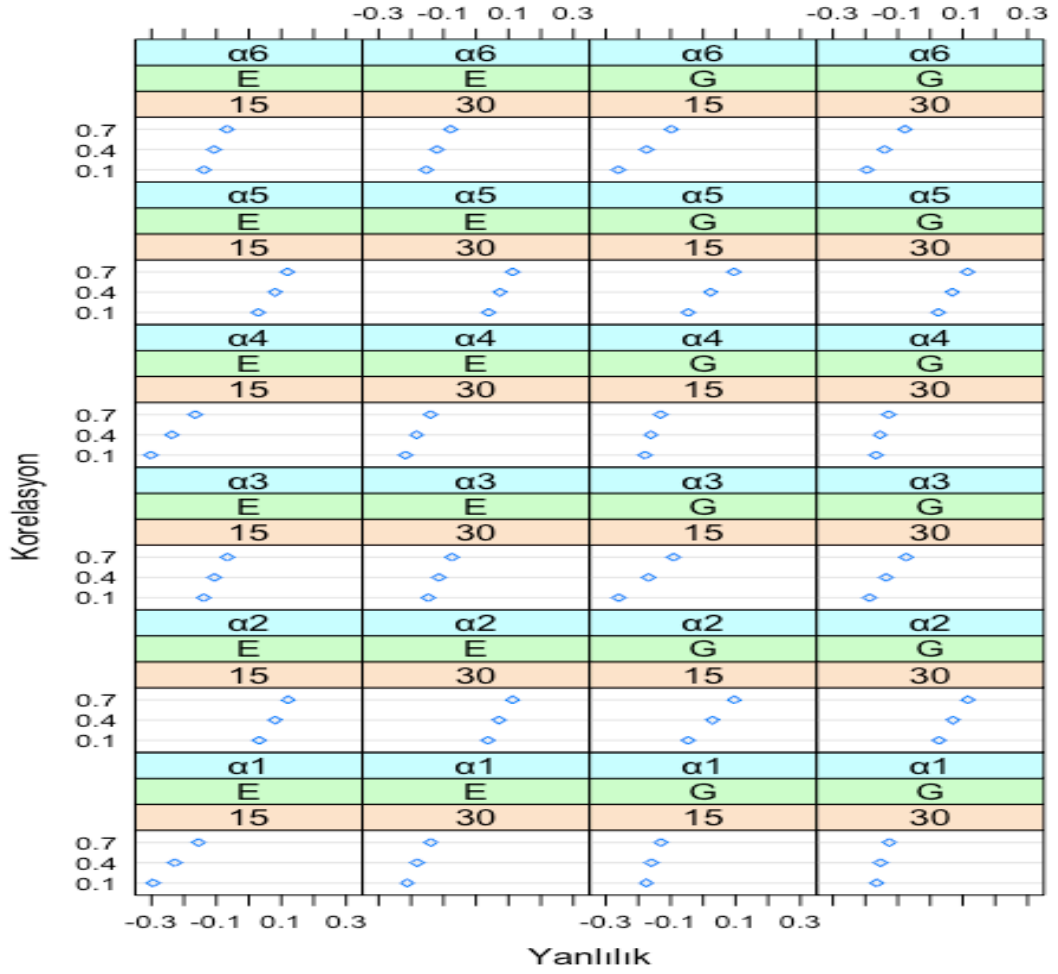
3PL ÇB-MTK'ya göre kestirilen yanlılık değerlerinin 1. nitelik için -0.015 ile 0.025, 2. nitelik için -0.039 ile -0.027, 3. nitelik için -0.023 ile 0.012, 4. nitelik için -0.017 ile 0.026, 5. nitelik için -0.042 ile -0.028 ve 6. nitelik için -0.022 ile 0.015 arasında değiştiği görülmektedir. 2. ve 5. niteliklerin tüm düzeyleri için kestirilen yanlılık değerlerinin negatif olması, bireylerin kestirilen sınıflama değerlerinin gerçek sınıflama değerlerinden daha küçük olduğunu göstermektedir. Madde sayısının kestirilen yanlılık değerleri üzerindeki etkisi incelendiğinde, madde sayısının artması ile yanlılık değerlerinin 0'a yaklaşma eğilimi gösterdiği gözlenmiştir. Ancak nitelikler arası korelasyonun 0.7 olduğu koşullarda, 6 nitelik için de kestirilen yanlılık değerleri 0.002 ile 0.021 arasında bir artış göstermiştir. Nitelikler arası korelasyonun artması

ile test uzunluğunun 15 madde olduğu koşullarda kestirilen yanlılık değerleri 0.002-0.012 arasında azalmıştır. Test uzunluğunun 30 madde olduğu koşullarda ise nitelikler arası korelasyonun artması ile yanlılığının arttığı gözlenmektedir. Yanlılık değerlerindeki bu artış, 0.001 ile 0.012 arasında değişmektedir. Kullanılan Q-matrisi yapısına göre yanlılık değerlerindeki değişimler karşılaştırıldığında, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile yanlılığın giderek azaldığı görülmektedir. 1. ve 4. nitelikler için ise test uzunluğunun 15 madde olduğu koşullarda gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılması, yanlılığın 0.001 ile 0.004 arasında daha küçük kestirilmesine neden olmaktadır.



Şekil 27. Altı nitelikten oluşan iki boyutlu veri setleri için DINA'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

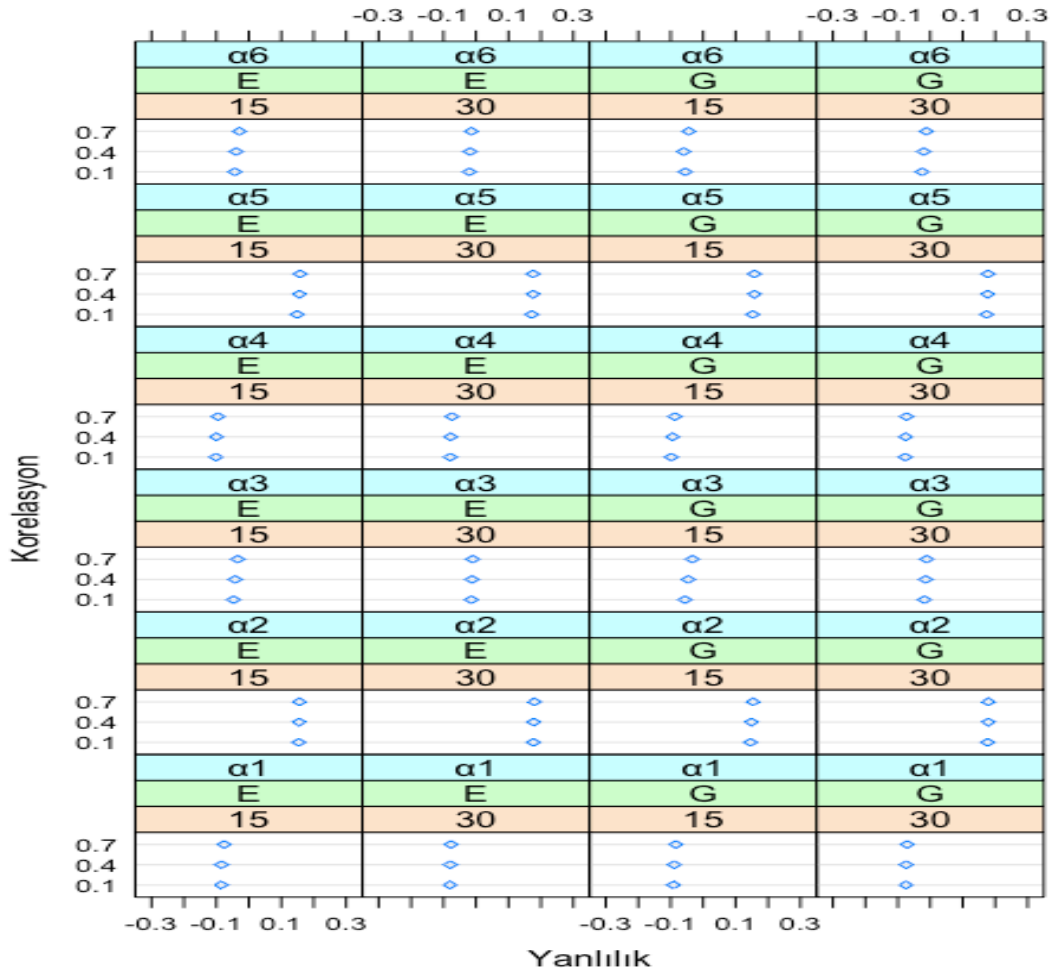
Telafi edici olmayan modellerden DINA'ya göre kestirilen yanlılık değerinin 1. nitelik için -0.041 ile 0.083, 2. nitelik için 0.204 ile 0.310, 3. nitelik için 0.012 ile 0.135, 4. nitelik için -0.042 ile 0.075, 5. nitelik için 0.200 ile 0.309 ve 6. nitelik için 0.010 ile 0.122 arasında değiştiği görülmektedir. Altı nitelik için kestirilen yanlılık değerleri oldukça yüksek olmakla birlikte, en yüksek yanlılık kestirimleri test uzunluğunun 15 ve nitelikler arası korelasyonun 0.1 olduğu koşullarda elde edilmiştir. Madde sayısının artması ile 1. ve 4. niteliğin tüm düzeyleri için kestirilen yanlılık değerleri azalmaktadır. Madde sayısının artmasına göre yanlılık değerlerinde görülen en büyük farklar, gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullar için nitelikler arasında 0,4 korelasyon olduğunda, niteliklerin eşit sayıda ölçüldüğü Q-matrisi kullanıldığında ise nitelikler arası korelasyonun 0.1 olduğu koşullarda görülmektedir. Diğer nitelikler için madde sayısının artması, gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı ve nitelikler arası korelasyonun 0.1 olduğu koşullar dışında, yanlılık değerlerini 0.005 ile 0.057 arasında arttırmaktadır. 2., 3., 5. ve 6. niteliklerin tüm düzeyleri için kestirilen yanlılık değerlerinin nitelikler arası korelasyonun artması ile azaldığı görülmektedir. 0.1-0.4 düzeylerindeki korelasyon değişiminin yanlılık üzerindeki etkisi ise daha fazladır. 1. ve 4. nitelikler için kestirilen yanlılık değerleri ise gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda, nitelikler arası korelasyonun artması ile artma eğilimi göstermektedir. Niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması, test uzunluğunun 15 madde olduğu koşullarda yanlılık değerini önemli ölçüde azaltmaktadır. Yanlılık değerlerindeki bu fark 0.010-0.092 arasında değişmekte ve nitelikler arası korelasyonun düşmesi ile artmaktadır. Test uzunluğunun 30 madde olduğu koşullarda ise gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılmasının, yanlılık değerini düşürdüğü söylenebilir.



Şekil 28. Altı nitelikten oluşan iki boyutlu veri setleri için DINO'ya göre kestirilen birey parametrelerine ilişkin yanılılık değerleri.

DINO modeli ile kestirilen yanılılık değerlerinin 1. nitelik için -0.295 ile -0.126, 2. nitelik için -0.046 ile 0.121, 3. nitelik için -0.260 ile -0.066, 4. nitelik için -0.302 ile -0.128, 5. nitelik için -0.046 ile 0.119 ve 6. nitelik için -0.261 ile -0.067 arasında değiştiği ve en yüksek yanılılık değerlerinin test uzunluğunun 15 madde olduğu koşullarda elde edildiği görülmektedir. Madde sayısının 15'ten 30'a çıkarılması ile 1. ve 4. niteliğin tüm düzeylerinde yanılılık azalmıştır. Yanılılık miktarındaki azalmanın, nitelikler arası korelasyon düştükçe ve niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda ise daha fazla olduğu gözlenmektedir. 2. ve 4. nitelikler için nitelikler arası korelasyonun 0.4-0.7 olduğu ve gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda, madde sayısının artması ile yanılılık değeri 0.020 ile 0.045 arasında artmıştır. Madde sayısının artması ile 3. ve 6. nitelikler için kestirilen yanılılık değerleri, gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda 0.017-0.072 arasında

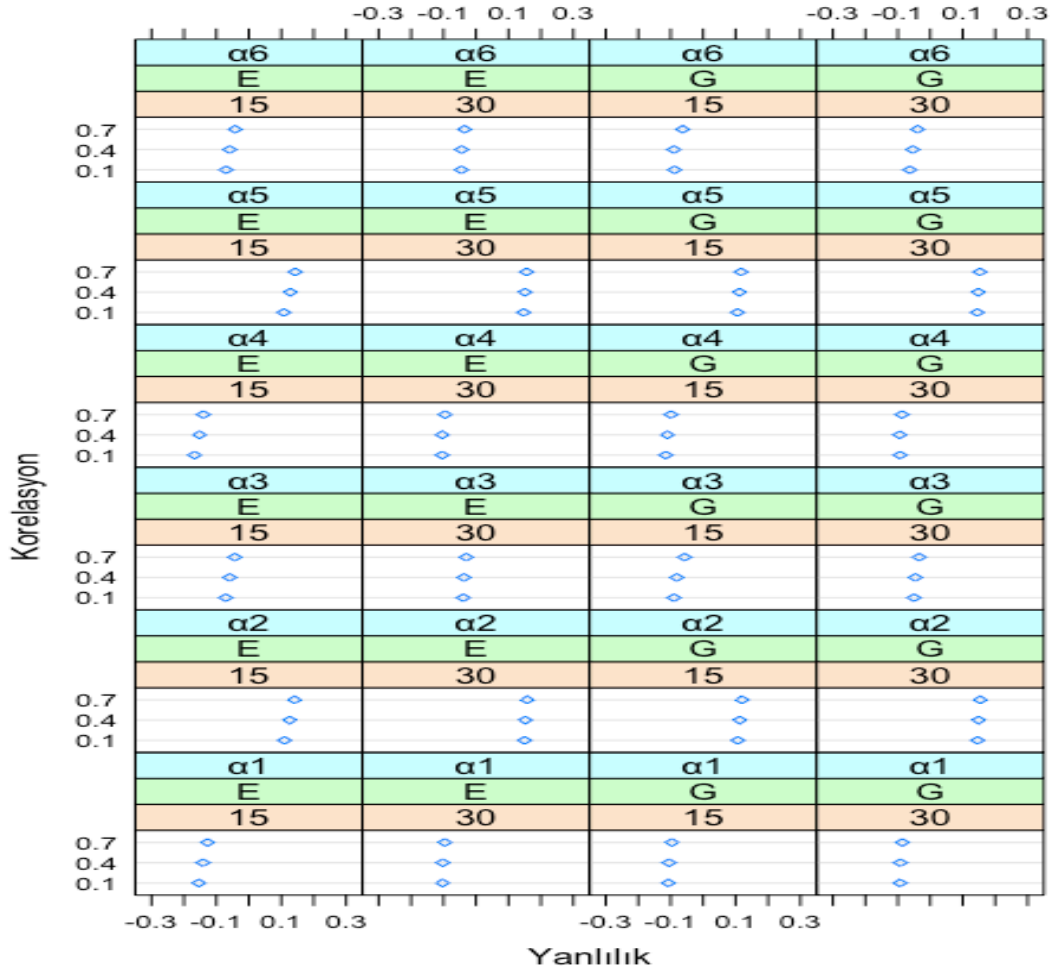
azalırken, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda 0.007-0.015 arasında artmaktadır. Nitelikler arası korelasyonun yanlılık üzerindeki etkisi incelendiğinde, 2. ve 5. nitelikler için kestirilen yanlılık değerlerinin korelasyon düzeyi arttıkça 0.034 ile 0.072 arasında arttığı, diğer niteliklerin tüm düzeylerinde ise nitelikler arası korelasyonun artmasının yanlılığı düşürdüğü bulunmuştur. Nitelikler arası korelasyona göre yanlılık değerlerindeki değişimin, 0.4-0.7 korelasyon düzeylerinde daha fazla olduğu görülmektedir. Son olarak, gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılması 1., 2., 4. ve 5. nitelikler için kestirilen yanlılık değerlerini azaltırken, 3. ve 6. nitelikler için yanlılık değerlerini arttırmaktadır.



Şekil 29. Altı nitelikten oluşan iki boyutlu veri setleri için A-CDM'ye göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.



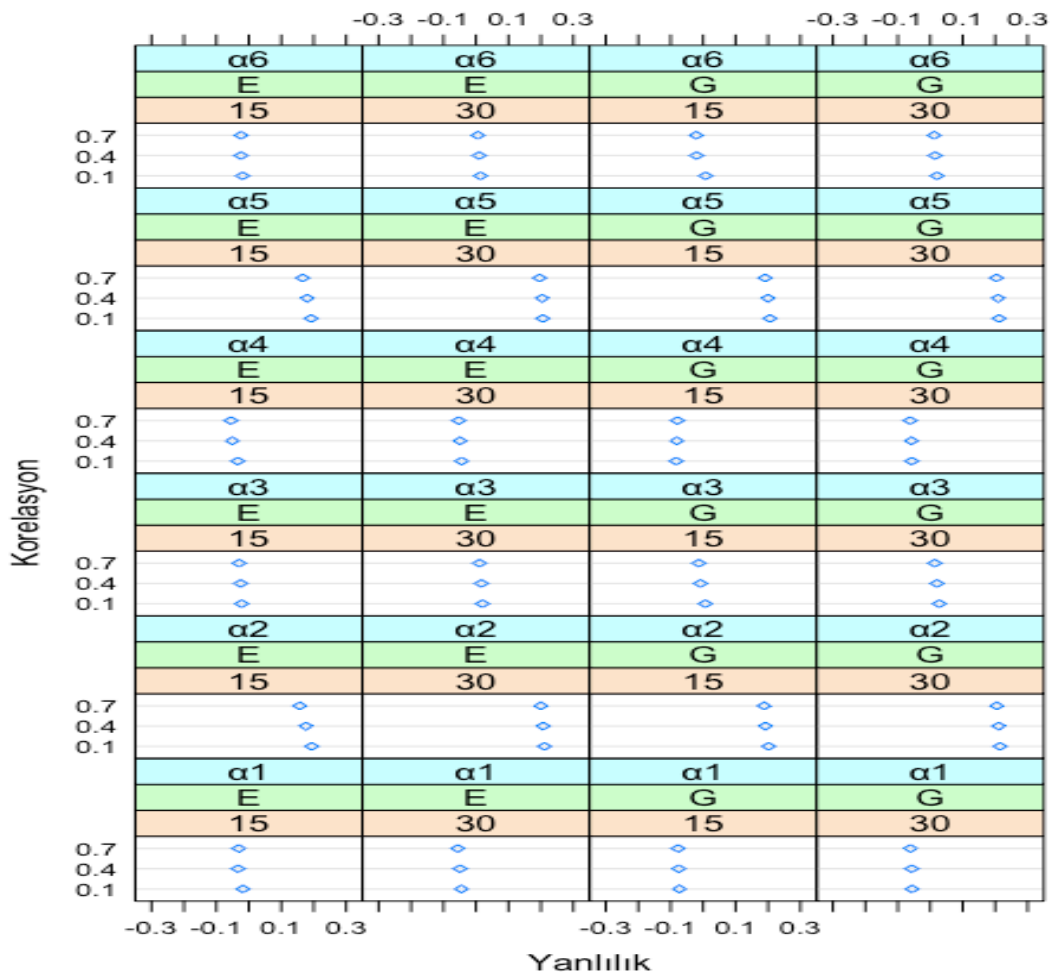
İki boyutlu veri setleri için A-CDM'ye göre kestirilen yanlılık değerlerinin 1. nitelik için -0.091 ile -0.071, 2. nitelik için 0.146 ile 0.180, 3. nitelik için -0.056 ile -0.010, 4. nitelik için -0.101 ile -0.073, 5. nitelik için 0.149 ile 0.178 ve 6. nitelik için -0.060 ile -0.012 arasında değiştiği gözlenmektedir. Madde sayısının artması ile 2. niteliğin tüm düzeyleri için kestirilen yanlılık değerleri 0.022-0.031, 5. niteliğin yanlılık değerleri ise 0.018-0.023 arasında artmıştır. Diğer nitelik düzeyleri için kestirilen yanlılık değerleri ise madde sayısının artması ile azalmaktadır. Yanlılık değerlerindeki azalma, 3. ve 6. nitelikler için 1. ve 4. niteliklere göre daha fazladır. Nitelikler arası korelasyonun artması ile 2. ve 5. nitelikler için kestirilen yanlılık değerleri artarken, diğer nitelikler için yanlılık düşmektedir. Ancak nitelikler arasındaki korelasyonun yanlılık üzerindeki etkisi oldukça küçüktür. Q-matrisi yapısının yanlılık kestirimleri üzerindeki etkisi incelendiğinde, gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılması ile 1. nitelik için test uzunluğunun 30 madde ve 2. nitelik için test uzunluğunun 15 madde olduğu koşullarda ve 4. niteliğin tüm düzeylerinde yanlılığın düştüğü görülmektedir. Diğer nitelik düzeylerinde ise niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması, yanlılığı düşürmektedir.



Şekil 30. Altı nitelikten oluşan iki boyutlu veri setleri için LLM'ye göre kestirilen birey parametrelerine ilişkin yanılılık değerleri.

Eklemeli BTM modellerinden LLM'ye göre kestirilen yanılılık değerlerinin 1. nitelik için -0.154 ile -0.086, 2. nitelik için 0.107 ile 0.158, 3. nitelik için -0.090 ile -0.030, 4. nitelik için -0.167 ile -0.087, 5. nitelik için 0.106 ile 0.156 ve 6. nitelik için -0.090 ile -0.035 arasında değiştiği gözlenmektedir. Tüm simülasyon koşulları dikkate alındığında 2. ve 5. nitelikler için kestirilen yanılılık değerlerinin pozitif olduğu, diğer nitelikler için ise yanılılık değerlerinin negatif olduğu dikkat çekmektedir. Madde sayısının artması ile 2. ve 5. nitelikler için kestirilen yanılılık değerlerinin 0.013-0.040 arasında arttığı görülmektedir. Yanılılık değerlerindeki bu artış en büyük değerini nitelikler arası korelasyonun 0.1 olduğu koşullarda elde etmekte ve nitelikler arası korelasyonun artması ile yanılılık kestirimleri arasındaki fark azalmaktadır. 1., 3., 4. ve 6. niteliklerin tüm düzeyleri için ise madde sayısının 15'ten 30'a çıkarılması ile yanılılık düşmektedir. 1. ve 4. nitelikler için yanılılık değerlerindeki değişim niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda daha büyük iken, 3. ve 6.

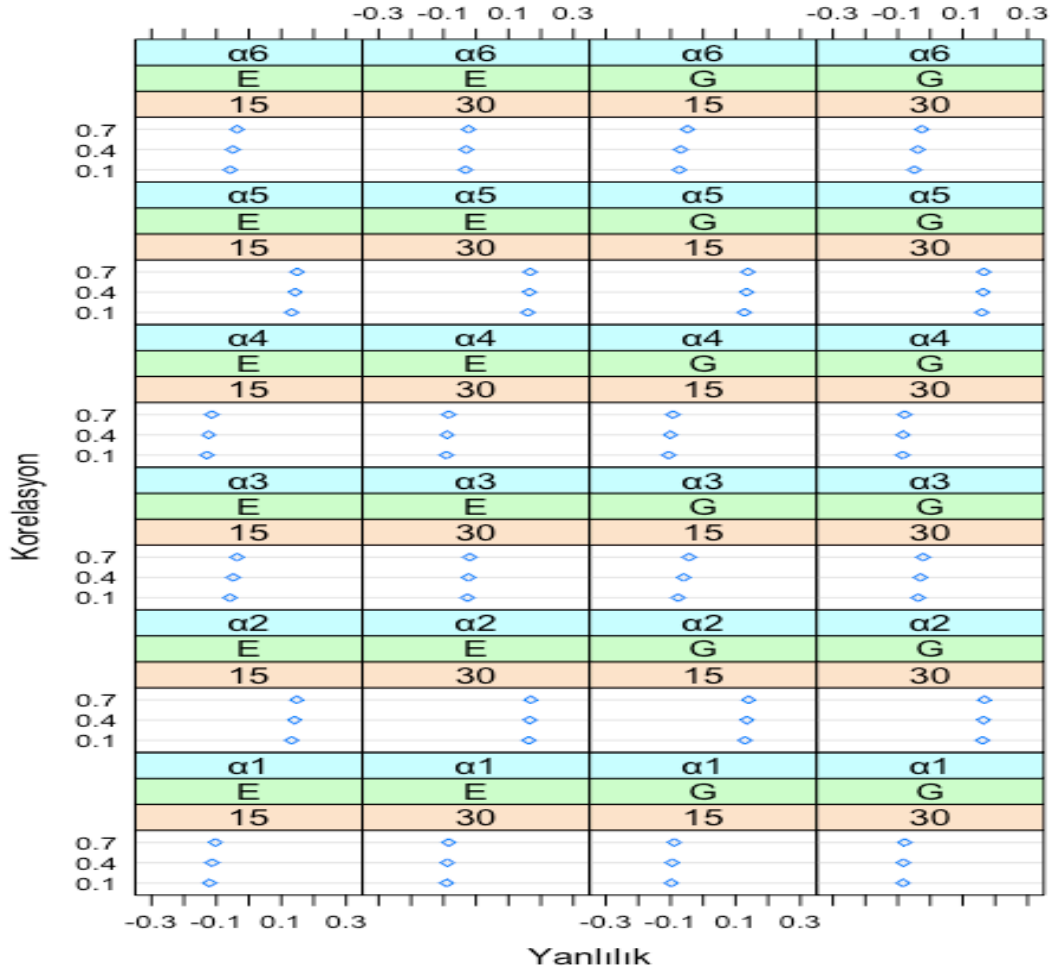
nitelikler için gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda daha büyüktür. Benzer şekilde, nitelikler arası korelasyonun artması 2. ve 5. niteliklerin tüm düzeylerinde kestirilen yanlılık değerlerini arttırırken, diğer 4 nitelik için kestirilen yanlılık değerlerinin düşmesine neden olmuştur. Niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması, 3. ve 6. nitelikler için yanlılığı 0.004 ile 0.031 arasında azaltmaktadır. Diğer 4 nitelik için ise gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılması, yanlılığı düşürmektedir. Kullanılan Q-matrisi yapısının yanlılık kestirimleri üzerindeki etkisinin, test uzunluğunun 15 madde olduğu koşullarda daha fazla olduğu tespit edilmiştir.



Şekil 31. Altı nitelikten oluşan iki boyutlu veri setleri için R-RUM'a göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

R-RUM modeline göre kestirilen birey parametrelerinin yanlılığı 1. nitelik için -0.077 ile -0.018, 2. nitelik için 0.157 ile 0.215, 3. nitelik için -0.030 ile 0.028, 4. nitelik için -0.083 ile -0.034, 5. nitelik için 0.166 ile 0.212 ve 6. nitelik için -0.024 ile 0.021

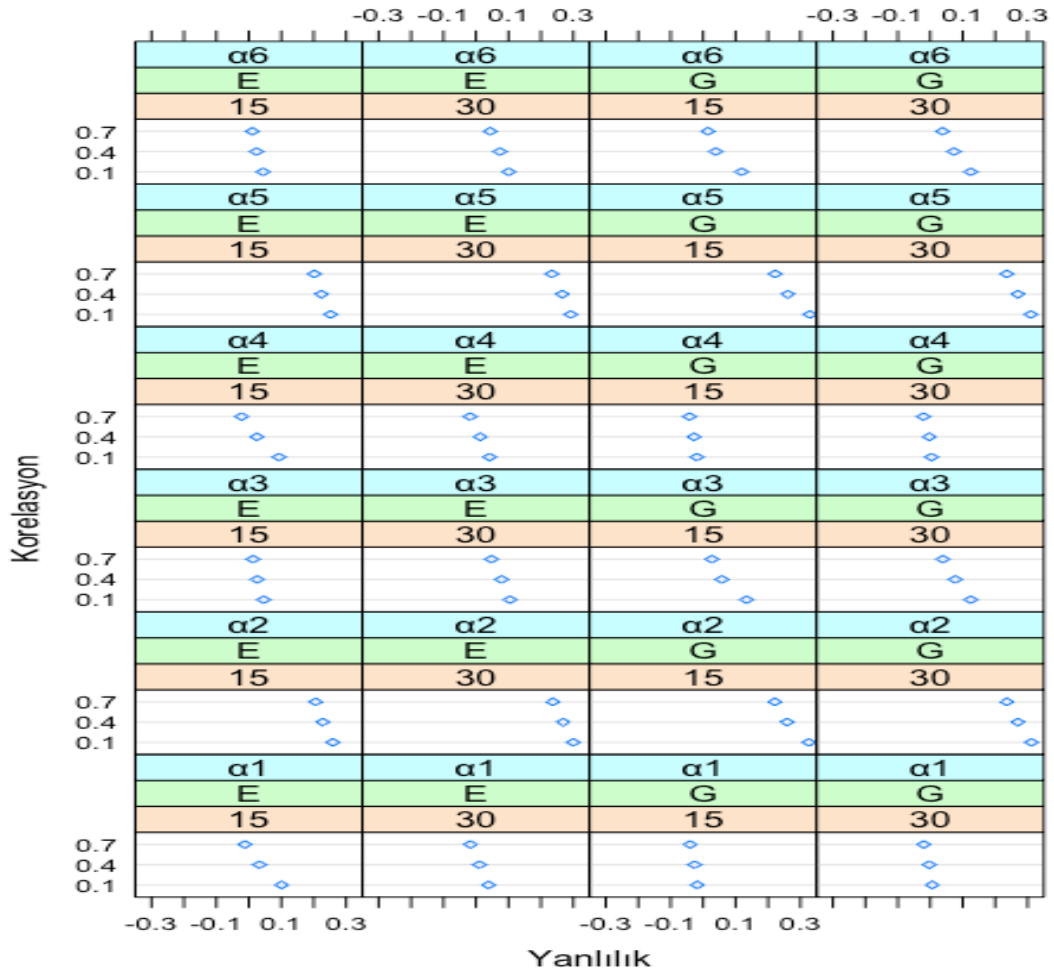
arasında değişmektedir. Model en düşük yanlılık kestirimlerini 3. ve 6. nitelikler için yaparken, en yüksek yanlılık kestirimlerini 2. ve 5. nitelikler için yapmıştır. Test uzunluğunun yanlılık kestirimleri üzerine etkisi incelendiğinde, madde sayısının 30'a çıkarılması ile 1. nitelik için niteliklerin eşit sayıda ölçüldüğü Q-matrisinin ve 3. nitelik için gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda, 2. ve 5. nitelikler için ise tüm koşullarda yanlılığın arttığı gözlenmektedir. 2. ve 5. nitelikler için yanlılık değerlerindeki değişim, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda daha büyük olmakla birlikte, nitelikler arası korelasyonun artması ile artmaktadır. Nitelikler arası korelasyonun artması ile 1. niteliğin tüm düzeylerinde yanlılığın 0.001-0.014 arasında, test uzunluğunun 15 madde olduğu koşullarda 3. nitelik için kestirilen yanlılık değerlerinin 0.001-0.005 arasında ve 6. nitelik için yanlılık değerlerinin 0.002-0.012 arasında arttığı bulunmuştur. 4. nitelik için ise test uzunluğunun 15 madde ve niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullar ile test uzunluğunun 30 madde olduğu tüm koşullar için kestirilen yanlılık değerleri, nitelikler arası korelasyonun artması ile 0.004-0.017 arasında artmıştır. Nitelikler arasındaki korelasyonun artması ile 2. ve 5. niteliklerin tüm düzeyleri için kestirilen yanlılık değerleri ise azalmıştır. Q-matrisi yapısının yanlılık üzerindeki etkisi incelendiğinde, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile 1., 2., 4. ve 5. niteliklerin tüm düzeylerinde ve test uzunluğunun 30 madde olduğu koşullarda 3. ve 6. nitelikler için yanlılığın azaldığı görülmektedir. Yanlılık kestirimindeki farklar, test uzunluğunun 15 madde ve nitelikler arası korelasyonun düşük olduğu koşullarda ise daha büyük olma eğilimi göstermiştir.



Şekil 32. Altı nitelikten oluşan iki boyutlu veri setleri için G-DINA'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

G-DINA modeline göre birey parametrelerinin kestirimlerinden elde edilen yanlılık değerleri 1. nitelik için -0.121 ile -0,079, 2. nitelik için 0.129 ile 0.169, 3. nitelik için -0.077 ile -0.019, 4. nitelik için -0.129 ile -0.079, 5. nitelik için 0.127 ile 0.167 ve 6. nitelik için -0.073 ile -0.023 arasında değişmektedir. Diğer koşullar sabit tutulduğunda madde sayısının artması, 2. ve 5. nitelikler için kestirilen yanlılık değerlerini 0.017-0.033 arasında arttırırken, diğer 4 niteliğin tüm düzeylerinde kestirilen yanlılık değerleri azalmıştır. Test uzunluğuna göre yanlılık değerlerindeki değişim miktarı ise nitelikler arası korelasyonun artması ile küçülmektedir. Madde sayısındaki değişimin yanlılık üzerindeki etkisi ile benzer şekilde, nitelikler arası korelasyonun artması ile 2. ve 5. niteliklerin tüm düzeyleri için yanlılığın arttığı, 1., 3., 4. ve 6. niteliklerin tüm düzeyleri için ise yanlılığın azaldığı görülmektedir. Niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması 3. ve 6. niteliklerin tüm

düzeylerinde kestirilen yanlılık değerlerini düşürmekle birlikte kestirimlerdeki değişim, test uzunluğunun 15 madde ve nitelikler arası korelasyonun düşük olduğu koşullarda daha büyüktür. Diğer 4 nitelik için kestirilen yanlılık değerleri ise gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda daha düşüktür. Q-matrisi yapısındaki farklılıklara ilişkin 1. ve 4. niteliklerin yanlılık kestirimlerinde görülen değişim ise testteki madde sayısının 15 olduğu koşullarda daha büyüktür.



Şekil 33. Altı nitelikten oluşan iki boyutlu veri setleri için HO-DINA'ya göre kestirilen birey parametrelerine ilişkin yanlılık değerleri.

HO-DINA'ya göre kestirilen yanlılık değerlerinin 1. nitelik için -0.040 ile 0.101, 2. nitelik için 0.206 ile 0.326, 3. nitelik için 0.013 ile 0.134, 4. nitelik için -0.042 ile 0.093, 5. nitelik için 0.202 ile 0.329 ve 6. nitelik için 0.011 ile 0.125 arasında değiştiği gözlenmektedir. Altı nitelik için en yüksek yanlılık kestirimleri, test uzunluğunun 15 madde ve nitelikler arası korelasyonun 0.1 olduğu koşullarda elde edilmiştir. Testteki

madde sayısının artmasına bağılı olarak yanlılık deęerlerinde gözlenen deęişim, niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda daha fazladır. Madde sayısının artması ile 1. ve 4. nitelikler için kestirilen yanlılık deęerlerinin düştüğü, dięer 4 nitelik için ise arttığı görülmektedir. Nitelikler arası korelasyonun artması 2., 3., 5. ve 6. niteliklerin tüm düzeyleri için yanlılığı düşürmektedir. Yanlılık kestirimindeki deęişimlerin, 0.1-0.4 korelasyon düzeyleri arasında ve gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda daha fazla olduğu bulunmuştur. Nitelikler arası korelasyonun artması ile 1. ve 4. nitelikler için kestirilen yanlılık deęerlerinin, test uzunluğunun 15 madde olduğu ve gerçek verinin faktör yapısına göre oluşturulan Q-matrisi kullanıldığı koşullarda ise arttığı görülmektedir. Test uzunluğunun 30 madde olduğu koşullarda ise korelasyon düzeyinin 0.1-0.4 arasında deęişmesi 1. ve 4. nitelikler için yanlılığı azaltırken, 0.4-0.7 arasında deęişmesi yanlılığı arttırmaktadır. Niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması, 1. nitelik için nitelikler arası korelasyonun 0.1 ve 0.4 olduğu; 3. ve 6. nitelikler için test uzunluğunun 30 madde ve nitelikler arası korelasyonun 0.4 ve 0.7 olduğu; 4. nitelik için test uzunluğunun 15 madde ve nitelikler arası korelasyonun 0.1 olduğu koşul ile test uzunluğunun 30 madde ve nitelikler arası korelasyonun 0.1 ve 0.4 olduğu koşullarda yanlılığı arttırmaktadır. Dięer nitelik düzeylerinde ise niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile yanlılık azalmaktadır. Q-matrisi yapısına göre yanlılık miktarındaki en büyük deęişimler ise test uzunluğunun 15 madde ve nitelikler arası korelasyonun 0.1 olduğu koşullarda görülmektedir.

### **Altı Nitelikten Oluşan İki Boyutlu Veri Setlerinin Yanlılık Deęerleri: Özet**

Tablo 19 incelendiğinde iki boyutlu veri setleri için testteki madde sayısı, nitelikler arası korelasyon ve Q-matrisi yapısı koşullarının tüm düzeylerine ait yanlılık deęerleri karşılaştırıldığında, 3PL ÇB-MTK'nın 6 nitelik için yaptığı yanlılık kestirimlerinin BTM modellerine göre daha küçük olduğu görülmektedir. 1. ve 4. nitelikler için DINA ve HO-DINA'nın, 2. ve 5. nitelikler için DINO'nun, 3. ve 6. nitelikler için ise R-RUM'un 3PL ÇB-MTK'ya en yakın yanlılık deęerlerine sahip olduğu gözlenmiştir.

Test uzunluğunun yanlılık deęerleri üzerindeki etkisi incelendiğinde, madde sayısının artması ile 2. ve 5. nitelikler için kestirilen yanlılık deęerlerinin arttığı, dięer

4 nitelik için yanlılık değerlerinin ise 0'a yaklaşma eğilimi gösterdiği görülmüştür. Madde sayısındaki artışa bağlı olarak yanlılık değerlerinde gözlenen değişimler, nitelikler arası korelasyonun 0.1 olduğu koşullarda daha fazladır ve korelasyonun artması ile azalmaktadır. Nitelikler arası korelasyonun artması ile yanlılık kestirimlerinde gözlenen değişim, kestirim modeline ve niteliklere göre farklılık göstermektedir. Madde sayısının artması ile kestirilen yanlılık değerlerinin, 3PL ÇB-MTK'ya göre test uzunluğunun 30 madde olduğu tüm düzeyler için; telafi edici olmayan BTM modellerinden DINA, R-RUM ve HO-DINA'ya göre 1. ve 4. nitelikler için; G-DINA ve telafi edici BTM modellerinden DINO, A-CDM ve LLM'ye göre ise 2. ve 5. nitelikler için arttığı tespit edilmiştir. Araştırma kapsamındaki 8 modele göre kestirilen yanlılık değerleri Q-matrisinin yapısına göre incelendiğinde, gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılması ile DINO, LLM ve G-DINA'ya göre 1., 2., 4. ve 5. nitelikler için; A-CDM'ye göre 1.,2. ve 4. nitelikler için; 3PL ÇB-MTK ve HO-DINA'ya göre 1. ve 4. nitelikler için; R-RUM'a göre test uzunluğunun 15 madde olduğu koşullarda 3. ve 6. nitelikler için yanlılığın azaldığı görülmektedir. Diğer nitelik düzeyleri için ise nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanılması yanlılığı düşürmektedir.

### **Üçüncü ve Dördüncü Alt Problemlere İlişkin Bulgular ve Yorumlar**

Araştırmanın ikinci alt problemi kapsamında, ÇB-MTK'ya göre üretilen verinin 3PL ÇB-MTK ve BTM modelleri ile analizinden elde edilen birey parametrelerinin doğru sınıflanma oranları her bir nitelik ve nitelik profili için testteki madde sayısı, nitelikler arası korelasyon ve Q-matrisi yapısına göre 3 niteliğin ve 6 niteliğin ölçüldüğü veri setleri için incelenmiştir.

### **Üç Nitelikten Oluşan Tek Boyutlu Veri Setleri için Doğru Sınıflama Oranları**

3PL ÇB-MTK ve BTM'lerin tek boyutlu veri setleri için kestirdikleri birey parametrelerinin nitelik ve nitelik profili bazındaki doğru sınıflama oranlarını Tablo 20'de görülmektedir.



Tablo 20

## Üç Nitelikten Oluşan Tek Boyutlu Veri Setleri için Doğru Sınıflama Oranları

MS	QM	r	ÇB-MTK		DINA		DINO		A-CDM	
			N	NP	N	NP	N	NP	N	NP
15	G	0.1	0.812	0.531	0.730	0.363	0.754	0.416	0.782	0.471
		0.4	0.827	0.567	0.772	0.461	0.784	0.485	0.799	0.517
		0.7	0.852	0.632	0.813	0.569	0.816	0.573	0.823	0.591
	E	0.1	0.818	0.546	0.759	0.415	0.754	0.411	0.792	0.491
		0.4	0.832	0.579	0.785	0.481	0.781	0.476	0.807	0.530
		0.7	0.857	0.641	0.821	0.580	0.820	0.577	0.829	0.597
30	G	0.1	0.859	0.631	0.772	0.430	0.800	0.501	0.822	0.550
		0.4	0.867	0.652	0.797	0.497	0.815	0.540	0.831	0.578
		0.7	0.883	0.698	0.829	0.588	0.840	0.613	0.847	0.630
	E	0.1	0.868	0.652	0.782	0.447	0.804	0.506	0.833	0.572
		0.4	0.875	0.670	0.802	0.501	0.821	0.548	0.840	0.594
		0.7	0.889	0.711	0.833	0.590	0.845	0.617	0.854	0.640
MS	QM	r	LLM		R-RUM		G-DINA		HO-DINA	
			N	NP	N	NP	N	NP	N	NP
15	G	0.1	0.787	0.482	0.772	0.449	0.786	0.480	0.730	0.362
		0.4	0.803	0.525	0.792	0.502	0.803	0.525	0.772	0.460
		0.7	0.829	0.600	0.817	0.581	0.828	0.599	0.813	0.568
	E	0.1	0.791	0.489	0.787	0.477	0.794	0.495	0.758	0.414
		0.4	0.806	0.529	0.805	0.524	0.808	0.533	0.785	0.480
		0.7	0.825	0.593	0.830	0.599	0.828	0.597	0.821	0.580
30	G	0.1	0.826	0.562	0.813	0.528	0.825	0.560	0.771	0.429
		0.4	0.836	0.589	0.823	0.558	0.835	0.587	0.797	0.495
		0.7	0.851	0.640	0.840	0.615	0.850	0.637	0.829	0.587
	E	0.1	0.835	0.580	0.826	0.551	0.835	0.579	0.782	0.446
		0.4	0.843	0.603	0.834	0.578	0.843	0.602	0.802	0.500
		0.7	0.856	0.648	0.850	0.630	0.856	0.646	0.833	0.590

MS: Madde sayısı, QM: Q-matrisi, r: Nitelikler arası korelasyon, G: Gerçek verinin faktör yapısına göre oluşturulan Q-matrisi, E: Nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi, N: Nitelik için doğru sınıflama oranı, NP: Nitelik profili için doğru sınıflama oranı

Nitelik bazında yapılan doğru sınıflama oranlarının 0.730 ile 0.889 arasında değiştiği ve 3PL ÇB-MTK'nın diğer modellere göre daha yüksek doğru sınıflama oranına sahip olduğu görülmektedir. En düşük doğru sınıflama oranına sahip model ise HO-DINA'dır. Tüm modellerin 0.833 ile 0.889 arasında değişen değerlerle en yüksek oranda doğru sınıflamayı test uzunluğunun 30 madde, nitelikler arası korelasyonun 0.7 olduğu ve niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda elde ettiği tespit edilmiştir. En düşük doğru sınıflama oranları ise test uzunluğunun 15 madde, nitelikler arası korelasyonun 0.1 olduğu ve gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda görülmekte ve 0.730 ile 0.812 arasında değişmektedir. Test uzunluğunun 15'ten 30'a çıkarılması ile tüm modeller için doğru sınıflama oranları 0.012-0.050 arasında bir artış göstermiştir. Madde sayısındaki artışa göre doğru sınıflama oranlarında gözlenen değişim, nitelikler arası korelasyonun artması ile azalmaktadır. Benzer şekilde nitelikler arası korelasyonun artması, tüm modeller için doğru sınıflama

oranlarının 0.007 ile 0.042 arasında artmasına neden olmuştur. Nitelikler arası korelasyonun artması ile doğru sınıflama oranlarında gözlenen değişim, test uzunluğunun 15 madde olduğu koşullarda ve 0.4-0.7 korelasyon düzeylerindeki değişimde daha fazladır. Q-matrisi yapısının modellerin yaptığı doğru sınıflama oranları üzerindeki etkisi incelendiğinde ise nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda doğru sınıflama oranlarının 0.003 ile 0.029 arasında artığı gözlenmiştir.

Tek boyutlu veri setleri için nitelik profili bazında yapılan doğru sınıflama oranlarının 0.362 ile 0.711 arasında değiştiği ve 3PL ÇB-MTK'nın BTM'lere daha yüksek doğru sınıflama oranına sahip olduğu görülmüştür. Tüm modellerin nitelik profili için en yüksek doğru sınıflama oranına madde sayısının 30, nitelikler arası korelasyonun 0.7 olduğu ve niteliklerin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda sahip olduğu tespit edilmiştir. En düşük doğru sınıflama oranları ise test uzunluğunun 15 madde, nitelikler arası korelasyonun 0.1 olduğu ve gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda gözlenmiştir. Diğer koşullar sabit tutulduğunda test uzunluğunun artması ile doğru sınıflama oranları 0.010 ile 0.106 arasında bir artış göstermiştir. Madde sayısının artması ile doğru sınıflama oranlarında gözlenen değişimler, nitelikler arası korelasyonun artması ile azalmaktadır. Nitelikler arası korelasyonun nitelik profillerinin doğru sınıflanması üzerindeki etkisi incelendiğinde, nitelikler arası korelasyonun artması ile doğru sınıflama oranlarının 0.018 ile 0.108 arasında arttığı tespit edilmiştir. Doğru sınıflama oranlarında gözlenen bu artış, 0.4-0.7 korelasyon düzeyleri arasındaki değişimde ve test uzunluğunun 15 madde olduğu koşullarda daha fazladır. Nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ise modellerin doğru sınıflama oranlarını 0.002 ile 0.052 arasında arttırma eğilimi göstermiştir.

### **Altı Nitelikten Oluşan İki Boyutlu Veri Setleri için Doğru Sınıflama Oranları**

3PL ÇB-MTK ve BTM'ler kapsamında birey parametre kestirimlerine ilişkin doğru sınıflama oranlarındaki değişim iki boyutlu veri setleri için testteki madde sayısı, nitelikler arası korelasyon ve Q-matrisi yapısı koşullarına göre incelenmiş ve Tablo 21'de sunulmuştur.

Tablo 21

*Altı Nitelikten Oluşan İki Boyutlu Veri Setleri için Doğru Sınıflama Oranları*

MS	Q	r	ÇB-MTK		DINA		DINO		A-CDM	
			N	NP	N	NP	N	NP	N	NP
15	G	0.1	0.813	0.286	0.731	0.137	0.756	0.181	0.782	0.225
		0.4	0.830	0.343	0.773	0.247	0.787	0.268	0.802	0.296
		0.7	0.858	0.456	0.816	0.407	0.824	0.414	0.829	0.424
	E	0.1	0.818	0.298	0.755	0.171	0.754	0.174	0.790	0.239
		0.4	0.834	0.352	0.784	0.259	0.782	0.259	0.806	0.302
		0.7	0.862	0.462	0.822	0.410	0.822	0.410	0.830	0.424
30	G	0.1	0.860	0.401	0.772	0.189	0.800	0.254	0.823	0.307
		0.4	0.868	0.439	0.797	0.277	0.817	0.317	0.832	0.355
		0.7	0.886	0.523	0.831	0.420	0.843	0.440	0.849	0.454
	E	0.1	0.869	0.428	0.782	0.203	0.804	0.258	0.832	0.328
		0.4	0.876	0.459	0.802	0.279	0.821	0.322	0.840	0.368
		0.7	0.890	0.535	0.832	0.416	0.846	0.443	0.854	0.459

MS	QM	r	LLM		R-RUM		G-DINA		HO-DINA	
			N	NP	N	NP	N	NP	N	NP
15	G	0.1	0.788	0.237	0.772	0.206	0.787	0.235	0.728	0.135
		0.4	0.807	0.305	0.794	0.283	0.806	0.303	0.770	0.246
		0.7	0.834	0.431	0.822	0.415	0.832	0.429	0.815	0.406
	E	0.1	0.790	0.240	0.785	0.226	0.792	0.245	0.753	0.169
		0.4	0.805	0.302	0.805	0.297	0.807	0.306	0.782	0.257
		0.7	0.827	0.423	0.833	0.426	0.830	0.425	0.822	0.411
30	G	0.1	0.827	0.318	0.814	0.283	0.826	0.316	0.769	0.186
		0.4	0.837	0.366	0.825	0.336	0.836	0.364	0.795	0.275
		0.7	0.853	0.463	0.842	0.442	0.852	0.461	0.829	0.418
	E	0.1	0.835	0.337	0.825	0.306	0.835	0.337	0.781	0.202
		0.4	0.843	0.378	0.834	0.351	0.843	0.377	0.801	0.278
		0.7	0.856	0.467	0.849	0.449	0.856	0.465	0.832	0.415

MS: Madde sayısı, QM: Q-matrisi, r: Nitelikler arası korelasyon, G: Gerçek verinin faktör yapısına göre oluşturulan Q-matrisi, E: Nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi, N: Nitelik için doğru sınıflama oranı, NP: Nitelik profili için doğru sınıflama oranı

Nitelik bazında yapılan doğru sınıflama oranlarının 0.728 ile 0.890 arasında değiştiği, 3PL ÇB-MTK'nın en yüksek ve HO-DINA'nın en düşük orana sahip olduğu gözlenmiştir. Tüm modeller en düşük doğru sınıflama oranını madde sayısının 15 ve nitelikler arası korelasyonun 0.1 olduğu koşullarda elde etmiştir. Bu koşullar altında doğru sınıflama oranları 0.728 ile 0.818 arasında değişmektedir. Test uzunluğunun 30 madde, nitelikler arası korelasyonun 0.7 ve nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda ise modellerin doğru sınıflama oranları 0.832-0.890 arasında değişmekte ve her bir model için en yüksek oranına ulaşmaktadır. Test uzunluğunun nitelik bazında yapılan doğru sınıflama oranları üzerindeki etkisi incelendiğinde, madde sayısının 15'ten 30'a çıkması ile doğru sınıflama oranlarının tüm modeller için arttığı ve artış oranının 0.010 ile 0.051 arasında değiştiği bulunmuştur. Madde sayısındaki değişime bağlı olarak doğru sınıflama oranlarında gözlenen bu artış, nitelikler arası korelasyonun artması ile azalmaktadır. Benzer şekilde, nitelikler arası korelasyonun artması ile

tüm model ve düzeyler için doğru sınıflama oranlarının 0.007 ile 0.045 arasında arttığı gözlenmiştir. Nitelikler arası korelasyonun artması ile doğru sınıflama oranlarında gözlenen değişim test uzunluğunun 15 madde olduğu koşullarda daha fazladır. Ayrıca 0.4-0.7 korelasyon düzeyleri arasındaki değişimin doğru sınıflama oranları üzerindeki etkisinin 0.1-0.4 korelasyon düzeyleri arasındaki değişimden daha fazla olduğu görülmüştür. Q-matrisi yapısının doğru sınıflama oranı üzerindeki etkisi incelendiğinde, test uzunluğunun 15 madde olduğu koşullarda gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılmasının DINO ve LLM için doğru sınıflama oranlarını sırasıyla 0.002-0.005 ve 0.002-0.007 arasında arttırdığı tespit edilmiştir. Diğer koşullar altında ise tüm modellerin nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda daha yüksek doğru sınıflama oranına sahip olduğu gözlenmiştir. Kullanılan Q-matrisi yapısına göre niteliklerin doğru sınıflama oranları arasında gözlenen farklar ise nitelikler arası korelasyonun artması ile azalmaktadır.

Nitelik profilleri için yapılan doğru sınıflama oranlarının 0.135 ile 0.535 arasında değiştiği görülmektedir. İlgili modeller en düşük doğru sınıflama oranını, test uzunluğunun 15 madde, nitelikler arası korelasyonun 0.1 olduğu ve gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı koşullarda elde etmektedir. En yüksek doğru sınıflama oranları ise test uzunluğunun 30 madde ve nitelikler arası korelasyonun 0.7 olduğu koşullarda gözlenmektedir. Nitelik sınıflamaları ile benzer şekilde, 3PL ÇB-MTK en yüksek doğru sınıflama oranına sahip iken HO-DINA en düşük doğru sınıflama oranına sahiptir. Madde sayısının artması ile doğru sınıflama oranlarının 0.004 ile 0.130 arasında arttığı tespit edilmiştir. Test uzunluğundaki değişime göre doğru sınıflama oranlarında gözlenen farklar, nitelikler arasındaki korelasyonun 0.1 olduğu koşullarda daha fazla olmakla birlikte, korelasyonun artması ile azalmaktadır. Benzer şekilde, nitelikler arası korelasyonun artması ile doğru sınıflama oranları 0.031 ile 0.160 arasında artmıştır. Korelasyon düzeyindeki artışa bağlı olarak doğru sınıflama oranlarında gözlenen değişimler, orta-yüksek korelasyon düzeyleri için daha fazla olmakla birlikte, test uzunluğunun 15 madde olduğu koşullarda 30 madde olduğu koşullara göre daha fazladır. Gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılması ile telafi edici BTM'lerden DINO ve LLM için doğru sınıflama oranları sırasıyla 0.004-0.009 ve 0.003-0.008 oranında artmıştır. Diğer koşullarda ise nitelik ve nitelik

profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile doğru sınıflama oranlarının 0.002 ile 0.034 arasında arttığı gözlenmiştir.

### Beşinci Alt Probleme İlişkin Bulgular ve Yorumlar

Araştırmanın üçüncü alt problemi kapsamında TIMSS 2015 uygulaması matematik testi verilerinin 3PL ÇB-MTK ve BTM ile analizinden kestirilen birey parametreleri arasındaki ilişkiler her bir nitelik ve nitelik sınıfı için incelenmiş ve geçen kişi oranları Tablo 22’de sunulmuştur.

Tablo 22

*Nitelik ve Nitelik Sınıfları için Geçen Kişi Oranları*

Model	Nitelik			Nitelik Sınıfı							
	$\alpha_1$	$\alpha_2$	$\alpha_3$	000	100	010	001	110	101	011	111
ÇB-MTK	0.53	0.25	0.46	0.42	0.11	0.00	0.05	0.01	0.18	0.00	0.23
DINA	0.53	0.46	0.49	0.45	0.04	0.00	0.02	0.03	0.04	0.01	0.42
DINO	0.49	0.22	0.40	0.48	0.11	0.00	0.02	0.00	0.16	0.00	0.22
A-CDM	0.53	0.23	0.49	0.41	0.09	0.00	0.05	0.00	0.22	0.00	0.22
LLM	0.52	0.22	0.48	0.41	0.11	0.00	0.05	0.00	0.20	0.01	0.22
R-RUM	0.53	0.22	0.52	0.40	0.08	0.00	0.06	0.00	0.24	0.00	0.22
G-DINA	0.53	0.24	0.47	0.41	0.11	0.00	0.05	0.00	0.19	0.01	0.23
HO-DINA	0.53	0.46	0.49	0.44	0.04	0.00	0.02	0.03	0.04	0.01	0.42

Tablo 22 incelendiğinde 1. niteliği geçen kişi yüzdelerinin %49 ile %53 arasında, 2. niteliği geçen kişi yüzdelerinin %22 ile %46 arasında ve 3. niteliği geçen kişi yüzdelerinin %40 ile %52 arasında değiştiği görülmektedir. 1. ve 3. nitelikler için geçen kişi yüzdelerinin ÇB-MTK ve BTM’lere göre benzer olduğu ancak 2. nitelik için telafi edici olmayan BTM’lerden DINA ve HO-DINA’ya göre geçen kişi yüzdesinin diğer modellere göre oldukça fazla olduğu dikkat çekmektedir. Nitelikler bazında geçen kişi yüzdeleri karşılaştırıldığında, 2. niteliğin en zor olduğu ve bunu sırasıyla 3. ve 1. niteliklerin takip ettiği gözlenmektedir. Her bir nitelik sınıfındaki kişi yüzdeleri incelendiğinde ise bireylerin büyük bir çoğunluğunun iki sınıfta toplandığı görülmektedir. Bunlar; bireylerin %40-%48’inin niteliklerin hiçbirine sahip olmadığı (0,0,0) sınıfı ve %22-%42’sinin niteliklerin hepsine sahip olduğu (1,1,1) sınıfıdır. (0,0,0) nitelik sınıfı için DINA ve HO-DINA modellerine göre geçen kişi yüzdeleri, diğer modellerden fazla olmakla birlikte aralarında çok büyük farklar yoktur. Ancak (1,1,1) nitelik sınıfı için diğer modellere göre geçen kişi yüzdeleri %22 ile %23

arasında deęişirken, DINA ve HO-DINA'ya gre geen kiři yzdeleri %42'dir. Benzer Őekilde dięer nitelik sınıfları iin de DINA ve HO-DINA, dięer 6 modelden farklı kestirimde bulunmuřtur. Gerek verinin yapısına en iyi model ve madde uyumunu gsteren G-DINA modeli ise DINA ve HO-DINA dıřındaki modeller ile olduka benzer kestirimlerde bulunmuřtur.



## Bölüm 5

### Sonuç, Tartışma ve Öneriler

Çalışmanın bu bölümünde, araştırma kapsamında elde edilen bulgulara dayalı olarak ulaşılan sonuçlara ve sonuçların ilgili literatür desteğiyle yorumlanmasına, ulaşılan sonuçlar doğrultusunda geliştirilen önerilere yer verilmiştir.

#### Sonuç ve Tartışma

ÇB-MTK'ya göre üretilen verilerin 3PL ÇB-MTK ve BTM'ler ile analizinden elde edilen hata değerleri incelendiğinde, verinin üretildiği 3PL ÇB-MTK modelinin hem üç hem de altı niteliğin ölçüldüğü test koşulları için en düşük hatayla kestirim yaptığı bulunmuştur. Bu durum BTM'lerin ÇB-MTK verisine uyarlanması kaynaklanmaktadır. Bu çalışmanın sonuçları ile paralel şekilde de la Torre ve Karelitz'in (2009) hiyerarşik nitelik yapısı altında DINA ve 2PL tek boyutlu MTK'yı birbirine karşılıklı uyarladıkları çalışmada, tüm koşullar için verinin üretildiği modelin veriye daha iyi uyum sağladığı gözlenmiştir. Benzer şekilde Yakar'ın (2017) bilişsel tanı ve çok boyutlu MTK modellerini birbirine karşılıklı uyarladığı çalışmada, tüm koşullar için verilerin üretildiği modellerin en yüksek doğru sınıflama oranı ile kestirim yaptığını belirtmiştir. Nitelik bazında kestirilen hata değerleri incelendiğinde, en fazla (1. ve 4. nitelikler) ve en az madde (3. ve 6. nitelikler) ile ölçülen nitelikler için en hatalı kestirimleri telafi edici BTM'lerden DINO modelinin, diğer nitelikler (2. ve 5. nitelikler) için ise DINA ve HO-DINA modellerinin yaptığı görülmüştür. Bu durum DINA ve DINO modellerinin oldukça sınırlandırıcı olmasından, DINA modelinin katılımcıları madde ile ölçülen niteliklerden hepsine sahip olan ve en az birine sahip olmayanlar, DINO modelinin ise madde ile ölçülen niteliklerden en az birine sahip olan ve niteliklerden hiçbirine sahip olmayanlar şeklinde iki sınıfa ayırmasından kaynaklanabilir. Bu çalışmanın sonuçlarından farklı olarak Tu, Wang, Cai, Douglas ve Chang (2018), nitelikler arasında tanımlanan hiyerarşik ilişkilerin madde parametreleri için kestirilen RMSE değerleri üzerine etkisini inceledikleri çalışmada, DINA modelinin en düşük RMSE değerine sahip olduğunu ve bunu sırasıyla R-RUM, A-CDM ve G-DINA modellerinin takip ettiğini bulmuştur.

Çalışmada testteki madde sayısı ve nitelikler arası korelasyon değerleri arttıkça kestirilen hata değerlerinin düştüğü bulunmuştur. Ancak hata değerlerinin

düşmesindeki en etkili değişkenin madde sayısı olduğu, nitelikler arası korelasyonun artması ile hata değerlerinde gözlenen azalmanın çok küçük olduğu görülmüştür. Madde sayısının artması ile hata değerlerinde gözlenen düşüş ise nitelikler arasındaki korelasyonun artması ile azalmıştır. Nitelikler arası korelasyonun hata üzerindeki etkisi incelendiğinde 0.4-0.7 korelasyon düzeyleri için kestirilen hata değerleri arasındaki farkın, 0.1-0.4 korelasyon düzeylerine göre daha fazla olduğu tespit edilmiştir. Tüm koşullar için modellerin test uzunluğunun 30 madde ve nitelikler arası korelasyon düzeyinin 0.7 olduğu koşullarda en düşük hata kestirimi yaptığı, test uzunluğunun 15 madde ve nitelikler arası korelasyonun 0.1 olduğu koşullarda ise en yüksek hata değerlerine sahip olduğu dikkat çekmektedir. Q-matrisi yapısının kestirilen hata değerleri üzerindeki etkisi ise niteliklere göre farklılık göstermiştir. Genel olarak, 1. nitelik için nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanılmasının hatayı arttırdığı, 2. ve 3. nitelikler için ise hatayı azalttığı gözlenmiştir. 6 niteliğin ölçüldüğü koşullar için nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi koşulunda 1. ve 4. nitelikler için kestirilen hataların arttığı, diğer 4 nitelik için kestirilen hataların ise azaldığı görülmüştür. 1. ve 4. nitelikler kapsamında kestirilen hata değerlerinin artması, gerçek verinin faktör yapısına göre oluşturulan Q-matrisinde bu nitelikleri ölçen madde sayısının diğer Q-matrisi yapısına göre daha fazla olmasından kaynaklanabilir. Nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda hata değerlerinin azalma eğilimi göstermesi ise olası tüm madde-nitelik ilişkilerinin yeterli sayıda soru ile temsil edilmesinden kaynaklanabilir. Genel olarak, Q-matrisi yapısının kestirilen hata değerleri üzerindeki etkisi ise nitelikler arası korelasyon düştükçe artmıştır.

Testteki madde sayısı, nitelikler arası korelasyon ve Q-matrisi yapısı koşullarının tüm düzeylerine ait yanlılık değerleri karşılaştırıldığında, RMSE sonuçları ile benzer şekilde, çalışma verilerinin üretildiği 3PL ÇB-MTK modelinin tüm nitelik düzeyleri için BTM'lere göre daha yansız kestirim yaptığı gözlenmiştir. Üç niteliğin ölçüldüğü veri setlerinde 1. nitelik için telafi edici olmayan modellerden DINA ve HO-DINA'nın ÇB-MTK ile en yakın ve DINO'nun ise en yanlı kestirimde bulunduğu; 2. nitelik için DINO modelinin ÇB-MTK ile benzer kestirimde bulunduğu, DINA ve HO-DINA'nın ise en yüksek yanlılık değerine sahip olduğu bulunmuştur. 3. nitelik için ise eklemeli modellerden R-RUM ve A-CDM, 3PL ÇB-MTK ile en yakın



yanlılık kestirimi yapmıştır. Altı nitelik için yanlılık değerlerinin kestirildiği veri setlerinde 1. ve 4. nitelikler için DINA ve HO-DINA'nın, 2. ve 5. nitelikler için DINO'nun, 3. ve 6. nitelikler için ise R-RUM'un 3PL ÇB-MTK ile en yakın yanlılık değerlerine sahip olduğu belirlenmiştir.

Test uzunluğunun artmasına bağlı olarak yanlılık değerlerinde gözlenen değişimler incelendiğinde, RMSE'deki değişim gibi sürekli bir azalma görülmemiştir. Madde sayısının artması ile yanlılık değerleri genel olarak 0'a yaklaşma eğilimi göstermesine rağmen üç niteliğin ölçüldüğü veri setleri için 2. nitelik ve altı niteliğin ölçüldüğü veri setleri için 2. ve 5. nitelikler için kestirilen yanlılık değerlerinde zaman zaman artışlar gözlenmiştir. Benzer şekilde, nitelikler arası korelasyonun artması ile yanlılık değerlerinin genel olarak düşüş gösterdiği ancak yanlılık kestirimlerinde gözlenen değişimin, kestirim modeli ve niteliklere göre farklılık gösterdiği belirlenmiştir. Nitelik sayısının üç olduğu test koşullarında, nitelikler arası korelasyonun artması ile 2. nitelik için kestirilen bazı yanlılık değerlerinde küçük artışlar gözlenmiştir. Altı nitelik için yanlılık değerlerinin kestirildiği test koşulları için de telafi edici modellerden DINO, A-CDM ve LLM ile G-DINA'ya göre 2. ve 5. nitelikler için kestirilen yanlılık değerleri, korelasyonun artması ile küçük oranda artmıştır. Üç niteliğin ölçüldüğü test koşullarında, nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile 2. ve 3. nitelikler için kestirilen yanlılık değerlerinin azaldığı, 1. nitelik için ise gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılması ile daha küçük yanlılık değerlerinin elde edildiği görülmüştür. Altı niteliğin ölçüldüğü test koşulları için Q-matrisi yapısına göre kestirilen yanlılık değerleri incelendiğinde, gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanılması ile DINO, LLM ve G-DINA'ya göre 1., 2., 4. ve 5. nitelikler için; A-CDM'ye göre 1., 2. ve 4. nitelikler için; 3PL ÇB-MTK ve HO-DINA'ya göre 1. ve 4. nitelikler için; R-RUM'a göre test uzunluğunun 15 madde olduğu koşullarda 3. ve 6. nitelikler için yanlılığın azaldığı belirlenmiştir. Diğer nitelik düzeylerinde ise nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanılması yanlılığı azaltmıştır.

Bu çalışmanın sonuçları ile benzer şekilde Huang (2017), çok düzeyli G-DINA modeli için kestirilen parametrelere ilişkin bias ve RMSE değerlerinin test uzunluğunun 15'ten 30'a çıkarılması ile azalma eğilimi gösterdiğini bulmuştur. Bradshaw ve Templin (2014), madde parametreleri için kestirilen bias ve RMSE

değerlerinin test uzunluğundan etkilendiğini, test uzunluğunun artması ile hata ve yanlılığın azaldığını belirtmiştir. Wang, Song, Chen ve Ding (2019), test uzunluğunun artması ile test düzeyinde kestirilen sınıflama doğrulukları için yanlılık, mutlak yanlılık ve RMSE değerlerinin düşme eğiliminde olduğunu gözlemiştir. Chen ve de la Torre (2018) kestirilen ortalama yanlılık ve RMSE değerlerinin, madde sayısının artması ile azaldığını tespit etmiştir. Chen ve de la Torre (2013) ise test uzunluğunun artması ve Q-matrisinin daha az kompleks olması ile yanlılık değerlerinin düşme eğilimi gösterdiğini bulmuştur.

Nitelik bazında yapılan doğru sınıflama oranları incelendiğinde bu değerlerin 0.728 ile 0.890 arasında değiştiği ve tüm simülasyon koşulları için 3PL ÇB-MTK'nın en yüksek doğru sınıflama oranına sahip olduğu gözlenmiştir. En düşük doğru sınıflama oranına sahip model ise HO-DINA olarak bulunmuştur. Araştırma kapsamındaki tüm modellerin en yüksek oranda doğru sınıflamayı, test uzunluğunun 30 madde, nitelikler arası korelasyonun 0.7 olduğu ve nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı test koşullarında yaptığı görülmüştür. En düşük doğru sınıflama oranları ise test uzunluğunun 15 madde ve nitelikler arası korelasyonun 0.1 olduğu test koşullarında elde edilmiştir. Test uzunluğunun artması ile tüm modeller için doğru sınıflama oranları artış göstermiştir. Benzer şekilde, nitelikler arası korelasyonun artması ile tüm test koşulları için ÇB-MTK ve BTM'lere göre yapılan doğru sınıflama oranları, üç niteliğin ölçüldüğü testler için 0.007-0.042 ve altı niteliğin ölçüldüğü testler için 0.007 ile 0.045 arasında artmıştır. Q-matrisi yapısının doğru sınıflama oranı üzerindeki etkisi incelendiğinde, nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile daha yüksek doğru sınıflama oranlarının elde edildiği tespit edilmiştir. Q-matrisi yapısına göre niteliklerin doğru sınıflama oranları arasında gözlenen farklar ise nitelikler arası korelasyonun artması ile azalma eğilimi göstermiştir.

Nitelik profili bazında yapılan doğru sınıflama oranlarının ise nitelik bazında yapılan doğru sınıflama oranlarına göre daha düşük olduğu gözlenmiştir. Üç niteliğin ölçüldüğü test koşullarında bu oranların 0.362 ile 0.711 ve altı niteliğin ölçüldüğü test koşullarında 0.135 ile 0.535 arasında değiştiği belirlenmiştir. Tüm koşullar için 3PL ÇB-MTK'nın BTM'lere göre daha yüksek doğru sınıflama oranına sahip olduğu tespit edilmiştir. BTM'ler arasından LLM ve G-DINA modellerinin ÇB-MTK'ya en yakın oranda doğru sınıflama yaptığı, DINA, DINO ve HO-DINA modellerinin ise en

düşük doğru sınıflama oranına sahip olduğu gözlenmiştir. Benzer şekilde Gao, Miller ve Liu (2017), DINA modelinin tüm koşullar için G-DINA, A-CDM'ye göre en düşük doğru sınıflama oranına sahip olduğunu bulmuştur. Tu, Wang, Cai, Douglas ve Chang (2018) ise DINA, R-RUM, A-CDM ve G-DINA modellerinin nitelik profili bazında doğru sınıflandırma oranlarını karşılaştırmış ve DINA modelinin en yüksek doğru sınıflama oranına sahip olduğu sonucuna varmıştır.

ÇB-MTK ve BTM'ler kapsamında kestirilen nitelik profilleri için en yüksek doğru sınıflama oranları, madde sayısının 30 ve nitelikler arası korelasyonun 0.7 olduğu koşullarda elde edilmiştir. En düşük doğru sınıflama oranları ise test uzunluğunun 15 madde, nitelikler arası korelasyonun 0.1 olduğu ve gerçek verinin faktör yapısına göre oluşturulan Q-matrisinin kullanıldığı test koşullarında görülmüştür. Test uzunluğunun 15'ten 30'a çıkarılması ile doğru sınıflama oranlarının üç niteliğin ölçüldüğü veri setleri için 0.010-0.106 ve altı niteliğin ölçüldüğü veri setleri için 0.004 ile 0.130 arasında arttığı, test uzunluğunun artması ile doğru sınıflama oranlarında gözlenen değişimlerin ise nitelikler arası korelasyonun artması ile azaldığı sonucuna varılmıştır. Benzer şekilde, nitelikler arası korelasyonun artması ile doğru sınıflama oranlarının arttığı ve doğru sınıflama oranlarında gözlenen bu artışın 0.4-0.7 korelasyon düzeylerindeki değişimde ve test uzunluğunun 15 madde olduğu koşullarda daha fazla olduğu gözlenmiştir. Son olarak, test uzunluğu ve nitelikler arasındaki korelasyona göre etkisi daha az olmakla birlikte, nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanılması ile doğru sınıflama oranlarının üç niteliğin ölçüldüğü test koşullarında 0.002 ile 0.052 arasında, altı niteliğin ölçüldüğü test koşullarında ise 0.002 ile 0.034 arasında arttığı tespit edilmiştir.

Bu çalışmadan elde edilen bulgularla paralellik gösterecek şekilde Tu, Zheng, Cai, Gao ve Wang (2018), DINA modelini temel alarak çok kategorili veriler için geliştirdikleri DINA-GD bilişsel tanı modelinin nitelik ve nitelik profili bazında birey parametrelerini doğru sınıflama oranlarının nitelik sayısının artması ile düştüğünü, test uzunluğunun artması ile arttığını belirtmiştir. Bradshaw ve Templin'nin (2014) çalışmasında da nitelik ve nitelik vektörü bazında yapılan doğru sınıflama oranlarının, test uzunluğunun artması ve nitelik sayısının azalması ile arttığı gözlenmiştir. Benzer şekilde Wang, Song, Chen, Meng ve Ding (2015), nitelik ve nitelik vektörü bazında kestirilen doğru sınıflama oranı, sınıflama doğruluğu ve

sınıflama tutarlığı değerlerinin test uzunluğunun artması ile arttığını, nitelik sayısının artması ile sınıflama doğruluğu ve tutarlığı değerlerinin ise düştüğünü tespit etmiştir. Gao, Miller ve Liu (2017), test uzunluğunun artması ile bireylerin genel doğru sınıflama oranlarının arttığını, modellerin doğru sınıflamaları üzerindeki en etkili değişkenin test uzunluğu olduğunu bulmuştur. Nitelikler arası korelasyonun artması ise birkaç istisnai durum dışında, doğru sınıflama oranlarını biraz arttırmıştır. Yakar (2017) de testteki madde sayısı ve nitelikler arası korelasyonun artması ile modellerin doğru sınıflama oranlarının arttığını ancak doğru sınıflama oranları üzerinde en etkili faktörün madde sayısı olduğunu belirtmiştir. Chen ve de la Torre (2018), Huang (2017) ve Wang, Song, Chen ve Ding (2019), test uzunluğunun artması ile nitelik ve nitelik vektörü için doğru sınıflama oranlarının monotonik olarak arttığını bulmuştur. Ayrıca Chen ve de la Torre (2013) ve Wang, Song, Chen ve Ding (2019), nitelik ve nitelik vektörü için kestirilen doğru sınıflama oranlarının daha basit yapıdaki Q-matrislerinin kullanılması ile arttığını belirtmiştir.

Gerçek veri uygulamasına dayalı sonuçlar incelendiğinde, G-DINA modelinin diğer 6 BTM'ye göre mutlak ve göreceli model uyumu ve madde uyumu indekslerine göre daha iyi uyum gösterdiği belirlenmiştir. Chen ve Chen (2016) PISA 2006 uygulaması okuma becerileri testi için G-DINA modeline göre oluşturdukları Q-matrisini kullanarak, DINA, DINO, A-CDM, LLM ve R-RUM'un model ve madde uyumunu incelendiği çalışmada, LLM'in en iyi model uyumunu gösterdiğini ancak testteki maddelerin hepsine uyum sağlayan bir BTM'nin olmadığını bulmuştur. Ayrıca bu çalışmanın sonuçlarından farklı olarak, birden fazla özelliği ölçen maddelerin hiçbirinin DINA ve DINO modeline uyum sağlamadığını tespit etmiştir. Benzer şekilde Chen, de la Torre ve Zhang (2013), 536 orta okul öğrencisi ile yaptıkları gerçek veri uygulamasında, AIC ve BIC değerlerine göre LLM'nin doygun model, DINA, DINO, A-CDM ve R-RUM'dan daha iyi model uyumu gösterdiğini tespit etmiştir. Yi (2017) ise İngilizce gramer testi verisi için en uygun bilişsel tanı modelini incelediği çalışmada, LCDM ve C-RUM'un diğer BTM'lere göre daha iyi göreceli model uyumu gösterdiğini bulmuştur. Mutlak uyum indeksleri incelendiğinde ise araştırma kapsamındaki bütün BTM'lerin (LCDM, C-RUM, DINA, DINO ve NIDO) veriye kabul edilebilir düzeyde uyum sağladığı ancak LCDM ve C-RUM'un daha iyi uyuma sahip olduğu görülmüştür.

Çalışmada, gerçek veriye göre nitelik bazında geçen kişi yüzdelerinin birbirinden farklı olduğu, katılımcıların en fazla 1. nitelikte başarılı olduğu, 2. nitelik ile ölçülen özelliklerin ise diğer konu alanlarına göre daha zor olduğu bulunmuştur. 1. ve 3. nitelikler için geçen kişi yüzdelerinin ÇB-MTK ve BTM'lere göre benzer olduğu ancak 2. nitelik için telafi edici olmayan BTM'lerden DINA ve HO-DINA'ya göre geçen kişi yüzdesinin diğer modellere göre daha yüksek olduğu gözlenmiştir. Ayrıca katılımcıların büyük bir çoğunluğunun (1,1,1) ve (0,0,0) nitelik sınıflarında toplandığı görülmüştür. Bu durum, TIMSS matematik testine göre tanımlanan nitelikler arası korelasyonun ortalama olarak 0.70 civarında olmasından kaynaklanabilir. Tu, Wang, Cai, Douglas ve Chang (2018) ise gerçek veri uygulaması yaptıkları çalışmada, geçen kişi yüzdelerinin kestirim yapılan modele göre değiştiği ve katılımcıların büyük bir çoğunluğunun %40-%48 ve %22-%42 arasında değişen yüzdelerle sırasıyla (1,1,1) ve (0,0,0) sınıflarında toplandığı görülmüştür. Benzer şekilde, Tu, Zheng, Cai, Gao ve Wang'ın (2018) PISA 2006 fen okuryazarlığı test verisini analiz ettikleri çalışmada, Q-matrisi 3 nitelik için tanımlanmış ve 8 nitelik profili ( $2^3$ ) için geçen kişi yüzdeleri karşılaştırıldığında en fazla kişinin (1,1,1) ve (0,0,0) nitelik sınıflarında toplandığını bulunmuştur. Chen ve Chen (2016), her bir nitelik için geçen kişi yüzdelerinin %49.6 ile %65.5 arasında değiştiğini, örtük sınıfların dağılımının ise birbirinden oldukça farklı olmakla birlikte katılımcıların büyük bir çoğunluğunun %30.7 değeri ile (1,1,1,1,1,1,1) örtük sınıfında toplandığını belirtmiştir. Yi (2017), örtük sınıflara göre geçen kişi yüzdelerinin, katılımcıların niteliklerin hepsine sahip olduğu (1,1,1) ve niteliklerin hiçbirine sahip olmadığı (0,0,0) iki uç sınıfta toplandığını göstermiştir.

## Öneriler

Bu bölümde, araştırmadan elde edilen sonuçlar doğrultusunda uygulamaya ve gelecek araştırmalara yönelik önerilere yer verilmiştir.

### Uygulamaya Dönük Öneriler

- 1) Verilerin üretildiği 3PL ÇB-MTK'nın tüm test koşulları için BTM'lere göre daha hatasız ve yansız kestirim yaptığı ve en yüksek doğru sınıflama oranına sahip olduğu bulunmuştur. Ancak madde sayısının ve nitelikler arası korelasyonun daha yüksek olduğu koşullarda BTM'lerin ÇB-MTK ile daha yakın kestirimler yaptığı gözlenmiştir. Bu nedenle BTM'nin ÇB-MTK verisine uyarlanması,

madde sayısının ve nitelikler arası korelasyonun yüksek olduğu test koşullarda daha uygun olabilir.

- 2) Çalışmada DINA, HO-DINA ve DINO modellerinin diğer modellere göre daha hatalı ve yanlı kestirimler yaptığı, nitelik ve nitelik profili bazında daha düşük doğru sınıflama oranına sahip olduğu gözlenmiştir. Tanılayıcı dönüt elde etmek amacıyla BTM'nin mevcut büyük ölçekli testlere uyarlanması durumunda sınırlayıcı BTM'ler yerine eklemeli ya da genel modeller kullanılabilir.
- 3) Araştırma kapsamında ele alınan TIMSS 2015 uygulaması 8. kitapçıkta yer alan soruların hepsi yayınlanmadığı için Q-matrisinin oluşturulmasında istatistiksel yöntemlerden faktör analizi kullanılmıştır. Faktör analizi sonuçlarına göre sorular karmaşık bir yapı sergilediği için niteliklerin ölçtükleri konu alanları isimlendirilememiştir. BTM'nin tanılayıcı gücünden faydalanabilmek için Q-matrisinin yayınlanan sorular üzerinden oluşturularak, niteliklerin açık ve net bir şekilde tanımlanması önerilebilir.
- 4) Çalışmada faktör analitik yöntemlerle gerçek verinin faktör yapısına göre oluşturulan Q-matrisi ve nitelik ve maddeler arası ilişkilerin random olarak oluşturulması ile nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisi kullanılmıştır. Çalışma sonuçları nitelik ve nitelik profillerinin eşit sayıda ölçüldüğü Q-matrisinin kullanıldığı koşullarda modellerin daha hatasız, yansız ve yüksek doğru sınıflama oranı ile kestirim yaptığını göstermiştir. Bu doğrultuda Q-matrisinin oluşturulmasında  $2^k-1$  nitelik profilinin her birinin yeterli sayıda ölçülmesine dikkat edilmesi gerekmektedir.

#### **Araştırmaya Dönük Öneriler**

- 1) Çalışmada madde ve nitelikler arasındaki ilişkilere yönelik ortak varsayımları temsil eden 7 bilişsel tanı modeli kullanılmıştır. Analizlerin uzun zaman alması nedeniyle araştırmaya dahil edilen model sayısı azaltılıp test koşulları arttırılarak, modellerin sınıflama doğruluğu üzerinde daha fazla değişkenin (örneğin, madde niteliği, verilerin üretilmesinde kullanılan model vb. gibi) etkisi incelenebilir.

- 2) Çalışmada kullanılan Q-matrisleri iki kategorili olarak tanımlanmıştır. Benzer bir çalışma, bireyin madde ile ölçülen niteliğe sahip olma derecesini çok kategorili olarak tanımlayan Q-matrisi yapısı ile tekrarlanabilir.
- 3) Çalışma kapsamında niteliklerin birbirinden bağımsız olduğu iki kategorili Q-matrisi yapısı kullanılmıştır. Gelecek araştırmalar farklı Q-matrisi yapılarının (nitelikler arası hiyerarşik ilişkilerin tanımlandığı ya da bireyin madde ile ölçülen niteliğe sahip olma derecesinin çok kategorili olarak belirlendiği) sınıflama doğruluğu üzerine etkisini araştırabilir.
- 4) ÇB-MTK ve BTM'lerin yaptığı sınıflamaların karşılaştırılması için ÇB-MTK çerçevesinde elde edilen sürekli yetenek puanlarının belirlenen bir kesme noktasına göre kategorik hale getirilmesi gerekmektedir. Gelecek araştırmalar, farklı kesme noktası belirleme yöntemlerinin sınıflama doğruluğu üzerine etkisini inceleyebilir.
- 5) Bu çalışmada katılımcılar, ölçülen özelliğe sahip olma ve olmama durumlarına göre iki gruba ayrılmıştır. Benzer bir çalışma, grup sayısının ikiden fazla olması durumunda modellerin nasıl sınıflama yaptığını araştırabilir.

## Kaynaklar

- Ackerman, T. (1996). Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement*, 20, 311-329.
- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*, 22(3), 37-53. Retrieved from <https://doi.org/10.1111/j.1745-3992.2003.tb00136.x>
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723. Retrieved from [https://doi.org/10.1007/978-1-4612-1694-0\\_16](https://doi.org/10.1007/978-1-4612-1694-0_16)
- Aryadoust, V. (2018). A cognitive diagnostic assessment study of the listening test of the Singapore-Cambridge general certificate of education o-level: application of DINA, DINO, G-DINA, HO-DINA, and RRUM. *International Journal of Listening*, 1-24. Retrieved from <https://doi.org/10.1080/10904018.2018.1500915>
- Başokçu, O. T. (2010). Öğrenme eksiklerinin belirlenmesinde klasik test teorisine dayalı yöntemler ve DINA modelin karşılaştırılması. *Ege Eğitim Dergisi*, 11(1), 59-83.
- Başokçu, O. T. ve Kelecioğlu, H. (2014). Bağlı ve mutlak değerlendirme ile DINA modele göre yapılan sınıflamaların geçerliğinin karşılaştırılması. *e-Uluslararası Eğitim Araştırmaları Dergisi*, 5(1), 1-19. <http://www.e-ijer.com/tr/download/article-file/89789> adresinden erişilmiştir.
- Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability*. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395-479). Reading, MA: Addison-Wesley.
- Böckenholt, U., & Tsai, R. C. (2006). 14 Random-Effects Models for Preference Data. In C. R. Rao, & S. Sinharay (Eds.), *Psychometrics* (pp. 447-468). (Handbook of Statistics; Vol. 26). Retrieved from [https://doi.org/10.1016/S0169-7161\(06\)26014-0](https://doi.org/10.1016/S0169-7161(06)26014-0)



- Bolt, D. (2007). The present and future of irt-based cognitive diagnostic models (ICDMs) and related methods. *Journal of Educational Measurement, 44*(4), 377- 383.
- Bradshaw, L., & Templin, J. (2014). Combining item response theory and diagnostic classification models: a psychometric model for scaling ability and diagnosing misconceptions. *Psychometrika, 79*(3), 403-425.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford.
- Buck, G., & Tatsuoka, L. (1998). Application of the rule-space procedure to language testing examining attributes of a free response listening test. *Language Testing, 15*, 119–157.
- Byrne, B.M. (1998). *Structural equation modeling with LISREL, PRELIS and SIMPLIS: basic concepts, applications and programming*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Chalmers, P., Pritikin, J., Robitzsch, A., Zoltak, M., Kim, K., Falk, C. F., Meade, A., Schneider, L., King, D., Liu, C. W., & Oguzhan, O. (2019). Mirt: Multidimensional Item Response Theory. R package version 1.30. Retrieved from <http://CRAN.R-project.org/package=mirt>
- Chen, H., & Chen, J. (2016). Retrofitting non-cognitive-diagnostic reading assessment under the generalized DINA model framework. *Language Assessment Quarterly, 13*(3), 218-230.
- Chen, J., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. *Applied Psychological Measurement, 37*(6), 419-437.
- Chen, J., & de la Torre, J. (2018). Introducing the general polytomous diagnosis modeling framework. *Frontiers in Psychology, Front. Psychol.* 9:1474. Retrieved from <https://doi.org/10.3389/fpsyg.2018.01474>
- Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement, 50*(2), 123-14.

- Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika*, 74(4), 619-632.
- Choi, H. J. (2010). *A model that combines diagnostic classification assessment with mixture item response theory models* (Doctoral dissertation). University of Georgia, Athens.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Florida: Harcourt Brace Jovanovich College Publishers.
- Cui, Y., Gierl, M. J., & Chang, H. H. (2012). Estimating classification consistency and accuracy for cognitive diagnostic assessment. *Journal of Educational Measurement*, 49, 19-38.
- Dayton, C. M., & Macready, G. (2006). *Latent class analysis in psychometrics*. In Rao, C. R., Sinharay, S. (Eds.). *Handbook of statistics: Psychometrics*, 26, 421-446. Netherlands: Elsevier.
- de Ayala, R. J. (2009). *Theory and practice of item response theory*. Guilford Publications.
- de la Torre, J. (2008). An empirically based method of q-matrix validation for the DINA model: development and applications. *Journal of Educational Measurement*, 45(4), 343-362.
- de la Torre, J. (2009). DINA model and parameter estimation: a didactic. *Journal of Educational and Behavioral Statistics*, 34, 115-130.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199.
- de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69, 333-353.
- de la Torre, J., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, 73, 595.

- de la Torre, J., & Karelitz, T. M. (2009). Impact of diagnosticity on the adequacy of models for cognitive diagnosis under a linear attribute structure: a simulation study. *Journal of Educational Measurement*, 46, 450-469.
- de la Torre, J., & Lee, Y. S. (2013). Evaluating the Wald test for item-level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement*, 50(4), 355–373.
- de la Torre, J., & Minchen, N. (2014). Cognitively Diagnostic Assessments and the Cognitive Diagnosis Model Framework. *Psicología Educativa*, 20(2), 89-97. Retrieved from <https://doi.org/10.1016/j.pse.2014.11.001>
- de la Torre, J., & Patz, R. J. (2005). Making the most of what we have: a practical application of multidimensional item response theory in test scoring. *Journal of Educational and Behavioral Statistics*, 30(3), 295–311. Retrieved from <https://doi.org/10.3102/10769986030003295>
- de la Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement*, 47(2), 227-249.
- Diamantopoulos, A., & Siguaw, J. A. (2000), *Introducing LISREL*. London: Sage Publications.
- Dziuban, C. D., & Shirkey, E. C. (1974). when is a correletion matrix appropriate for factor analysis? some decision rules. *Psychological Bulletin*, 81(6), 358-361.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Embretson, S. E., & Yang, X. (2013). A multicomponent latent trait model for diagnosis. *Psychometrika*, 78, 14-36.
- Fay, R. H. (2018). *Application of the fusion model for cognitive diagnostic assessment with non-diagnostic algebra-geometry readiness test data* (Doctoral dissertation). University of South Florida, Florida.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286–299. Retrieved from <https://doi.org/10.1037/1040-3590.7.3.286>

- Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: a Latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading, 10*(3), 301-322.
- Gao, M., Miller, M. D., & Liu, R. (2017). The Impact of q-matrix misspecification and model misuse on classification accuracy in the generalized DINA model. *Journal of Measurement and Evaluation in Education and Psychology, 8*(4), 391-403.
- García, P. E., Olea, J., & de la Torre, J. (2014). Application of cognitive diagnosis models to competency-based situational judgment tests. *Psicothema, 26*(3), 372-377.
- George, A. C., & Robitzsch, A. (2015). Cognitive diagnosis models in R: a didactic. *Tutorials in Quantitative Methods for Psychology, 11*(3), 189-205.
- George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R package CDM for cognitive diagnosis models. *Journal of Statistical Software, 74*(2), 1-24.
- Gierl, M. J., & Cui, Y. (2008) Defining characteristics of diagnostic classification models and the problem of retrofitting in cognitive diagnostic assessment. *Measurement: Interdisciplinary Research and Perspectives, 6*(4), 263-268.
- Gierl, M. J., Alves, C., & Majeau, R. T. (2010). Using the attribute hierarchy method to make diagnostic inferences about examinees' knowledge and skills in mathematics: An operational implementation of cognitive diagnostic assessment. *International Journal of Testing, 10*, 318-341.
- Gierl, M. J., Alves, C., Roberts, M., & Gotzmann, A. (2009, April). *Using judgments from content specialists to develop cognitive models for diagnostic assessments*. In J. Gorin (Chair), How to build a cognitive model for educational assessments. Paper presented in symposium conducted at the 2009 annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Gierl, M. J., Leighton, J. P., & Hunka, S. M. (2007). *Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills*. In

- J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (p. 242–274). Cambridge University Press. Retrieved from <https://doi.org/10.1017/CBO9780511611186.009>
- Gierl, M. J., Tan, X., & Wang, C. (2005). *Identifying content and cognitive dimensions on the SAT* (Research Report No. 2005-11). New York: College Examination Board.
- Gierl, M. J., Wang, C., & Zhou, J. (2008). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in algebra on the SAT. *Journal of Technology, Learning, and Assessment*, 6(6), 4-50.
- Haberman, S. J., & Sinharay, S. (2010). Reporting of subscores using multidimensional item response theory. *Psychometrika*, 75, 209–227.
- Haberman, S. J., Sinharay, S., & Puhan, G. (2009). Reporting subscores for institutions. *British Journal of Mathematical and Statistical Psychology*, 62, 79–95.
- Haladyna, T. M., & Kramer, G. A. (2004). The validity of subscores for a credentialing test. *Evaluation & The Health Professions*, 27(4), 349–368.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston: Academic Publishers Group.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Thousand Oaks: Sage Publications.
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139–164.
- Hartig, J., & Hohler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation*, 35, 57–63.
- Hartz, S. M. (2002). *A bayesian framework for the unified model for assessing cognitive abilities: blending theory with practice* (Doctoral dissertation). University of Illinois, Urbana-Champaign.
- Henard, D. H. (2000). Item response theory, in *Reading and Understanding MORE Multivariate Statistics*, Vol. II, Larry Grimm and Paul Yarnold (Eds). Washington, DC: American Psychological Association, 67-97.

- Henning, G. (1987). *A guide to language testing: development, evaluation, research*. Rowley, MA: Newbury House Publishers.
- Henson, R. (2004). *Test discrimination and test construction for cognitive diagnostic models* (Doctoral dissertation). University of Illinois, Urbana-Champaign.
- Henson, R. A. (2009). Diagnostic classification models: thoughts and future directions. *Measurement: Interdisciplinary Research and Perspectives*, 7, 34-36.
- Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74(2), 191-210.
- Henson, R., & Douglas, J. (2005) Test construction for cognitive diagnosis. *Applied Psychological Measurement*, 29(4), 262-277.
- Henson, R., & Templin, J. (2007, April). Large-scale language assessment using cognitive diagnosis models. Paper presented at the annual meeting of the National Council for Measurement in Education, Chicago, Illinois.
- Henson, R., Roussos, L., Douglas, J., & He, X. (2008). Cognitive diagnostic attribute-level discrimination indices. *Applied Psychological Measurement*, 32(4), 275-288.
- Henson, R., Templin, J., & Douglas, J. (2007). Using efficient model based sum scores for conducting skills diagnoses. *Journal of Educational Measurement*, 44(4), 361-376.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Huang, H. Y., & Wang, W. C. (2014). The random-effect dina model. *Journal of Educational Measurement*, 51(1), 75-97.
- Huebner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research & Evaluation*, 15, 1-7.

- Huebner, A., & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement, 71*(2), 407-419.
- Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: validity arguments for applying fusion model to LanguEdge assessment. *Language Testing, 26*(1), 31-73.
- Jang, E. E., Dunlop, M., Wagner, M., Kim, Y. H., & Gu, Z. (2013). Elementary school ELLs' reading skill profiles using cognitive diagnosis modeling: roles of length of residence and home language environment. *Language Learning, 63*, 400-436.
- Jansen, B. R., & van der Maas, H. L. (1997). Statistical test of the rule assessment methodology by latent class analysis. *Developmental Review, 17*(3), 321-357.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*, 258-272.
- Jurich, D. P. (2014). *Assessing model fit of multidimensional item response theory and diagnostic classification models using limited-information statistics* (Doctoral dissertation). James Madison University, Virginia.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31-36. Retrieved from <https://doi.org/10.1007/BF02291575>
- Karasar, N. (2009). *Bilimsel araştırma yöntemi* (20. Baskı). Ankara: Nobel Yayıncılık.
- Kim, A. Y. (2014). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing, 32*, 227-258.
- Kruidenier, J. (2002). Literacy assessment in adult basic education. In J. Comings, B. Garner and C. Smith (Eds.) *Annual Review of Adult Learning and Literacy*, vol.3 [pp. 84–151]. San Francisco, CA: Jossey-Bass.
- Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2009). A practical illustration of multidimensional diagnostic skills profiling: comparing results from

- confirmatory factor analysis and diagnostic classification models. *Studies in Educational Evaluation*, 35(2), 64–70.
- Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2012). The impact of model misspecification on parameter estimation and item-fit assessment in log-linear diagnostic classification models. *Journal of Educational Measurement* 49, 59-81.
- Lee, S. H. (2007). *Multiimensional item response theory: a SAS MDIRT MACRO and emprical study of PIAT MATH test* (Doctoral dissertation). The University of Oklahoma, Norman.
- Lee, Y., & Sawaki, Y. (2009). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly*, 6, 239–263.
- Leighton, J. P., & Gierl, M.J. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*, 26, 3-16.
- Leighton, J. P., Gierl, M. J., & Hunka, S. (2004). The attribute hierarchy model: an approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement*, 41, 205–237.
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936–949. Retrieved from <https://doi.org/10.3758/s13428-015-0619-7>
- Li, F., Cohen, A. S., Kim, S. H., & Cho, S. J. (2009). Model selection methods for mixture dichotomous IRT models. *Applied Psychological Measurement*, 33(5), 353-373.
- Li, H., Hunter, C. V., & Lei, P. W. (2015). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*, 33(3), 391-409.
- Liao, Y. (2007). Investigating the construct validity of the grammar and vocabulary section and the listening section of the ECCE: Lexico-grammatical ability as a predictor of L2 listening ability. *Spaan Fellow Working Papers in Second or*



- Foreign Language Assessment*, 5, 37–78. English Language Institute, University of Michigan, MI.
- Lin, C. J., & Chang, H. H. (2018). Item selection criteria with practical constraints in cognitive diagnostic computerized adaptive testing. *Educational and Psychological Measurement*, 79(2), 335–357.
- Liu, R. (2017). Misspecification of attribute structure in diagnostic measurement. *Educational and Psychological Measurement*, 1–28.
- Liu, R., Huggins-Manley, A. C., & Bradshaw, L. (2017). The impact of q-matrix designs on diagnostic classification accuracy in the presence of attribute hierarchies. *Educational and Psychological Measurement*, 77(2) 220–240.
- Liu, R., Huggins-Manley, A. C., & Bulut, O. (2017). Retrofitting diagnostic classification models to responses from IRT-based assessment forms. *Educational and Psychological Measurement*, 1-27.
- Longabach, T., & Peyton, V. (2017). A comparison of reliability and precision of subscore reporting methods for a state English language proficiency assessment. *Language Testing*, 1–21.
- Lord, F. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates Publishers.
- Lord, F. M., & Novick M. R. (1968). *Statistical theories of mental test scores*. New York: Addison- Wesley Publishing Company.
- Ma, W., & de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology*, 69, 253-275.
- Ma, W., de la Torre, J., Sorrel, M., & Jiang, Z. (2019). GDINA: the generalized DINA model framework. R package version 2.7.3. Retrieved from <https://cran.r-project.org/web/packages/GDINA>.
- Ma, W., Iaconangelo, C., & de la Torre, J. (2016). Model similarity, model selection, and attribute classification. *Applied Psychological Measurement*, 40, 200-217. Retrieved from <https://doi.org/10.1177/0146621615621717>

- Maris, E. (1995). Psychometric latent response models. *Psychometrika*, 60(4), 523-547.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 54, 184-212.
- Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models, measurement. *Interdisciplinary Research and Perspectives*, 11(3), 71-101.
- McDonald, R. P. (1999). *Test theory: a unified treatment*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- McKinley, R. L., & Reckase, M. D. (1982). *Multidimensional latent trait models*. Paper presented at the National Council on Measurement in Education, New York.
- MEB (2016). TIMSS 2015 ulusal matematik ve fen bilimleri ön raporu 4. ve 8. sınıflar. Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü, Ankara.
- [http://timss.meb.gov.tr/wp-content/uploads/TIMSS\\_2015\\_Ulusal\\_Rapor.pdf](http://timss.meb.gov.tr/wp-content/uploads/TIMSS_2015_Ulusal_Rapor.pdf) adresinden erişilmiştir.
- Minchen, N. D., de la Torre, J., & Liu, Y. (2017). A cognitive diagnosis model for continuous response. *Journal of Educational and Behavioral Statistics*, 42(6), 651-677.
- Minchen, N., & de la Torre, J. (2018). A general cognitive diagnosis model for continuous-response data. *Measurement: Interdisciplinary Research and Perspectives*, 16:1, 30-44.
- Mîndrilă, D. (2010). Maximum likelihood (ml) and diagonally weighted least squares (DWLS) estimation procedures: a comparison of estimation bias with ordinal and multivariate non-normal data. *International Journal of Digital Society (IJDS)*, 1(1), 60-66.
- Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 257-305). Washington, DC: American Council on Education.

- Moseley, D. (2004). The diagnostic assessment of word recognition and phonic skills in five-year-olds. *Journal of Research in Reading*, 27 (2), 132-140. Retrieved from <https://doi.org/10.1111/j.1467-9817.2004.00221.x>
- Muthén, B. O., & Kaplan, D. (1985). A comparison of some methodologies for the factor-analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171–180.
- National Research Council (2001). Knowing what students know: The science and design of educational assessment. Committee on the Foundations of Assessment. J. Pellegrino, N. Chudowsky, & R. Glaser, (Eds.), Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- Neuman, W. L. (2014). *Toplumsal araştırma yöntemleri nitel ve nicel yaklaşımlar* (7. Baskı). Ankara: Yayın Odası.
- Ravand, H. (2016). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment*, 34, 782–799.
- Ravand, H., & Robitzsch, A. (2015). Cognitive diagnostic modeling using R. *Practical Assessment, Research & Evaluation*, 20(11), 1-12. Retrieved from <http://pareonline.net/getvn.asp?v=20&n=11>
- Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9(4), 401-412. Retrieved from <https://doi.org/10.1177/014662168500900409>
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.
- Reckase, M. D., & Xu, J. R. (2015). The evidence for a subscore structure in a test of English language competency for english language learners. *Educational and Psychological Measurement*, 75, 805-825.
- Revelle, W. (2019). Psych: procedures for psychological, psychometric, and personality research. R package version 1.8.12. Retrieved from <https://cran.r-project.org/web/packages/psych>
- Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2019). MASS: support functions and datasets for venables and ripley's MASS. R

- package version 7.3-51.4. Retrieved from <https://cran.r-project.org/web/packages/MASS>.
- Roberts, M. R., & Gierl, M. J. (2010). Developing score reports for cognitive diagnostic assessment. *Educational Measurement: Issues and Practice*, 29(3), 25-38.
- Robitzsch A., Kiefer T., George A. C., & Uenlue A. (2019). CDM: cognitive diagnosis modeling. R package version 7.4-19. Retrieved from <http://CRAN.R-project.org/package=CDM>
- Rosseel, Y., Oberski, Jorgensen., T. D., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., & Scharf, F. (2019). Llavaan: Latent Variable Analysis. R package version 0.6-5. Retrieved from <http://CRAN.R-project.org/package=lavaan>
- Roussos, L., Templin, J., & Henson, R. (2007). Skills Diagnosis using IRT-based latent class models. *Journal of Educational Measurement*, 44, 293-311.
- Rupp A. A. (2007). The answer is in the question: a guide for describing and investigating the conceptual foundations and statistical properties of cognitive psychometric models. *International Journal of Testing*, 7, 95-125.
- Rupp, A. A., & Mislevy, R. J. (2007). Cognitive foundations of structured item response models. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 205- 241). . Retrieved from <http://dx.doi.org/10.1017/CBO9780511611186.008>
- Rupp, A. A., & Templin, J. L. (2008b). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, 6, 219-262.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: theory, methods, and applications*. New York, NY: Guilford.
- Rupp, A., & Templin, J. (2008a). The effects of q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, 68, 78-96.

- Sawaki, Y., & Sinharay, S. (2017). Do the TOEFL iBT section scores provide value-added information to stakeholders?. *Language Testing*, 1-28. Retrieved from <https://doi.org/10.1177/0265532217716731>
- Schwarzer, G. (1976). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Sen, S., & Bradshaw, L. (2017). Comparison of relative fit indices for diagnostic model selection. *Applied Psychological Measurement*, 1-17.
- Sessoms, J., & Henson, R. A. (2018) Applications of diagnostic classification models: a literature review and critical commentary. *Measurement: Interdisciplinary Research and Perspectives*, 16(1), 1-17.
- Sheng Y., & Wikle C. K. (2007). Comparing multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement*, 68(3), 413-430.
- Shin, C. D., Ansley, T., Tsai, T., & Mao X. (2005). *A comparison of methods of estimating objective scores*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Quebec, Canada.
- Shin, D. (2007). *A comparison of methods of estimating subscale scores for mixed-format tests*. Pearson Educational Measurement.
- Sinharay, S. (2010). How often do subscores have added value? Results from operational and simulated data. *Journal of Educational Measurement*, 47, 150-174.
- Sinharay, S., & Haberman, S. J. (2014) An empirical investigation of population invariance in the value of subscores. *International Journal of Testing*, 14(1), 22-48.
- Sinharay, S., Haberman, S. J., & Puhan, G. (2007). Subscores based on classical test theory: to report or not to report. *Educational Measurement: Issues and Practice*, 26, 21–28.
- Sinharay, S., Puhan, G., & Haberman, S. J. (2010). Reporting diagnostic scores in educational testing: temptations, pitfalls, and some solutions. *Multivariate Behavioral Research*, 45(3), 553-573.

- Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice*, 30(3), 29-40.
- Snow, R.E., & D.F. Lohman (1989). *Implications of cognitive psychology for educational measurement*. In Educational measurement, ed. R.L. Linn, 3rd ed., 263–331. New York: American Council on Education and Macmillan.
- Sorrel, M. A., de la Torre, J., Abad, F. J., & Olea, J. (2017). Two-step likelihood ratio test for item-level model comparison in cognitive diagnosis models. *Methodology*, 13, 39-47.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.
- Stone, M. (1979). Comments on model selection criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society*, 41(2), 276–278. Retrieved from <https://doi.org/10.1111/j.2517-6161.1979.tb01084.x>
- Tatsuoka, K. K. (1983). Rule Space: an approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345-354.
- Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational and Behavioral Statistics*, 10, 55-73.
- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11, 287-305.
- Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, 30, 251-275.
- Templin, J., & Hoffman, L. (2013). Obtaining diagnostic classification model estimates using Mplus. *Educational Measurement: Issues and Practice*, 32(2), 37-50.
- Terzi, R., & de la Torre, J. (2018). An iterative method for empirically-based q-matrix validation. *International Journal of Assessment Tools in Education*, 5(2), 248-262.

- Tu, D., Wang, S., Cai, Y., Douglas J., & Chang, H. H. (2018). Cognitive diagnostic models with attribute hierarchies: model estimation with a restricted q-matrix design. *Applied Psychological Measurement*, 1–17.
- U.S. Department of Education. (2001). Public Law print of PL 107-110, the No Child Left Behind Act of 2001. Retrieved from <http://www2.ed.gov/policy/elsec/leg/esea02/index.html>
- van der Ark, L. A., Bolt, D. M., Wang, W. C., Douglas, J. A., & Chow, S. M. (Eds.) (2014). *Quantitative Psychology Research: The 79th Annual Meeting of the Psychometric Society*, Madison, Wisconsin, 2014. New York: Springer. ISBN 978-3-319-19976-4
- von Davier, M. (2005). *A general diagnostic model applied to language testing data* (ETS Research Report RR-05-16). Princeton, NJ: Educational Testing Service.
- Wang, W., Song, L., Chen, P., & Ding, S. (2019). An item-level expected classification accuracy and its applications in cognitive diagnostic assessment. *Journal of Educational Measurement*, 56(1), 51–75.
- Wang, Y. C. (2009). *Factor analytic models and cognitive diagnostic models: how comparable are they? a comparison of R-RUM and compensatory MIRT model with respect to cognitive feedback* (Doctoral dissertation). University of North Carolina, Greensboro.
- Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika*, 81, 625-649.
- Yakar, L. (2017). *Bilişsel tanı ve çok boyutlu madde tepki kuramı modellerinin karşılıklı uyumlarının incelenmesi* (Doktora tezi) Hacettepe Üniversitesi, Ankara.
- Yao, L., & Boughton, K. A. (2007). A multidimensional item response modeling approach for improving subscale proficiency estimation and classification. *Applied Psychological Measurement*, 31, 83-105.
- Yi, Y. S. (2017) In search of optimal cognitive diagnostic model(s) for ESL grammar test data. *Applied Measurement in Education*, 30(2), 82-101.

**EK-A: TIMSS Matematik Testinin 1, 2 ve 4 Boyutlu Formlarına İlişkin Faktör  
Yük Değerleri**

Testin Farklı Boyuttaki Formlarına İlişkin Faktör Yük Değerleri

Tek Boyutlu Yapı		İki Boyutlu Yapı			Dört Boyutlu Yapı				
Madde	1. boyut	Madde	1. boyut	2. boyut	Madde	1. boyut	2. boyut	3. boyut	4. boyut
M1	0.51	M1	0.38	0.35	M1	0.29	0.16	0.48	-0.01
M2	0.74	M2	0.64	0.37	M2	0.62	0.31	0.28	-0.08
M3	0.76	M3	0.62	0.43	M3	0.54	0.24	0.54	0.03
M4	0.75	M4	0.64	0.39	M4	0.60	0.31	0.33	0.01
M5	0.4	M5	0.12	0.52	M5	0.10	0.49	0.20	0.06
M6	0.38	M6	0.67	0.38	M6	0.63	0.29	0.32	0.15
M7	0.76	M7	0.56	0.30	M7	0.52	0.22	0.28	0.04
M8	0.63	M8	0.30	0.43	M8	0.31	0.46	0.12	-0.05
M9	0.5	M9	0.61	0.41	M9	0.57	0.34	0.30	0.18
M10	0.74	M10	0.22	0.55	M10	0.22	0.63	0.09	0.00
M11	0.5	M11	0.66	0.49	M11	0.62	0.40	0.36	0.06
M12	0.83	M12	0.48	0.18	M12	0.47	0.15	0.15	0.12
M13	0.49	M13	0.80	0.29	M13	0.77	0.24	0.26	-0.03
M14	0.81	M14	0.49	0.27	M14	0.49	0.27	0.15	-0.09
M15	0.56	M15	0.37	0.44	M15	0.25	0.21	0.61	0.01
M16	0.56	M16	0.69	0.28	M16	0.66	0.22	0.28	-0.06
M17	0.72	M17	0.46	0.33	M17	0.46	0.33	0.16	-0.13
M18	0.57	M18	0.46	0.33	M18	0.45	0.32	0.17	-0.10
M19	0.56	M19	0.74	0.32	M19	0.72	0.27	0.25	-0.09
M20	0.78	M20	0.67	0.38	M20	0.67	0.37	0.20	-0.09
M21	0.77	M21	0.23	0.64	M21	0.18	0.53	0.37	-0.02
M22	0.55	M22	0.36	0.61	M22	0.28	0.44	0.52	0.04
M23	0.65	M23	0.59	0.62	M23	0.51	0.45	0.52	0.13
M24	0.84	M24	0.49	0.53	M24	0.47	0.50	0.24	0.14
M25	0.71	M25	0.39	0.37	M25	0.37	0.34	0.17	0.13
M26	0.53	M26	0.50	0.35	M26	0.49	0.33	0.18	0.14
M27	0.61	M27	0.76	0.33	M27	0.71	0.21	0.37	0.14
M28	0.81	M28	0.54	0.16	M28	0.51	0.10	0.21	0.06
M29	0.53	M29	0.60	0.24	M29	0.58	0.19	0.20	0.23
M30	0.62								



**EK-B: Testin Farklı Sayıda Boyutlardan Oluşan Formlarına İlişkin EM,  
QMCEM ve MHRM Yöntemlerine Göre Kestirilen Mutlak ve Göreceli Model  
Uyumu İndeksleri**

**Testin Farklı Sayıda Boyutlardan Oluşan Formlarına İlişkin Mutlak Model Uyumu İndeksleri**

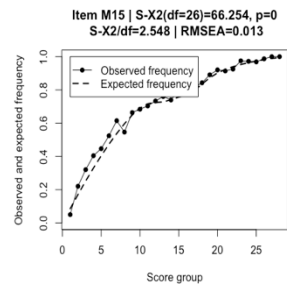
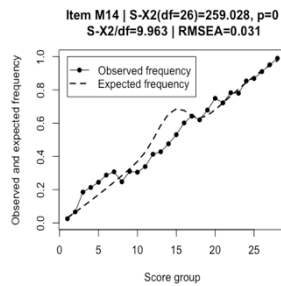
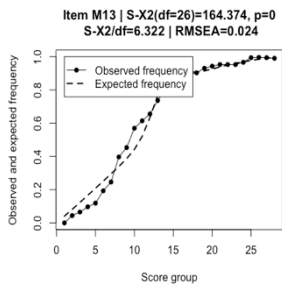
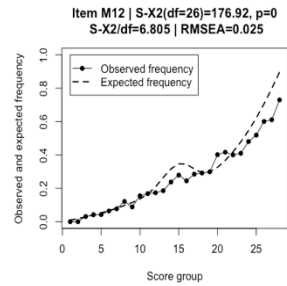
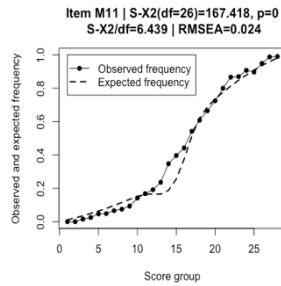
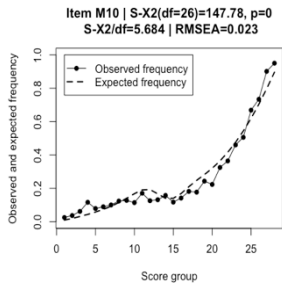
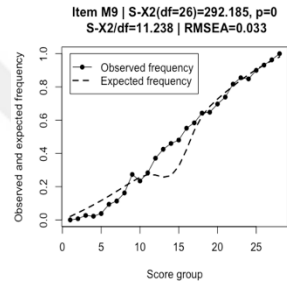
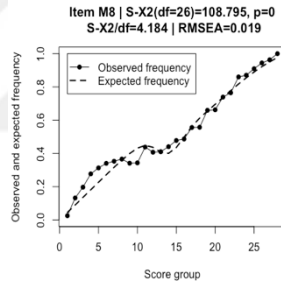
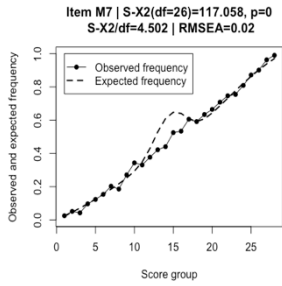
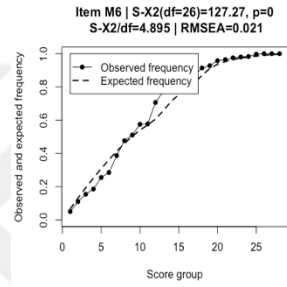
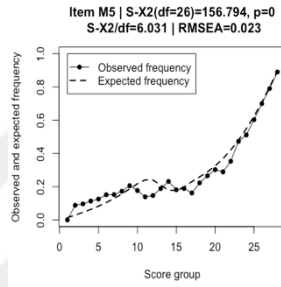
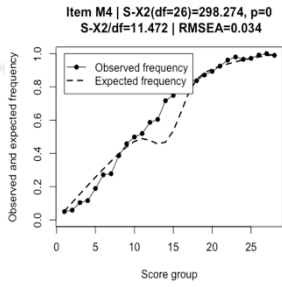
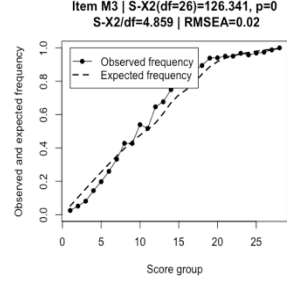
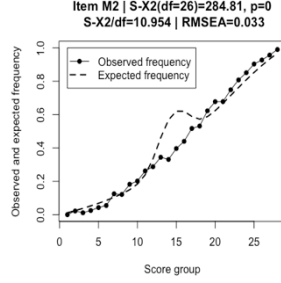
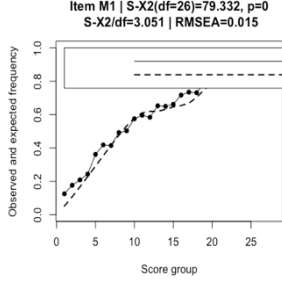
Boyut Sayısı	Kestirim Yöntemi	M2	df	RMSEA	SRMSR	TLI	CFI
1	EM	3365.200	377	0.0293	0.0260	0.9880	0.9889
	QMCEM	3365.285	377	0.0293	0.0260	0.9880	0.9889
	MHRM	3366.680	377	0.0293	0.0259	0.9880	0.9889
2	EM	1656.871	349	0.0202	0.0204	0.9943	0.9951
	QMCEM	1656.832	349	0.0202	0.0204	0.9943	0.9951
	MHRM	1659.312	349	0.0202	0.0204	0.9943	0.9951
3	EM	990.7818	322	0.0150	0.0187	0.9969	0.9975
	QMCEM	985.4691	322	0.0150	0.0178	0.9969	0.9975
	MHRM	986.6266	322	0.0150	0.0181	0.9969	0.9975
4	EM	822.7944	296	0.0139	0.0229	0.9973	0.9980
	QMCEM	858.1029	296	0.0144	0.0298	0.9971	0.9979
	MHRM	959.2871	296	0.0156	0.0326	0.9966	0.9975

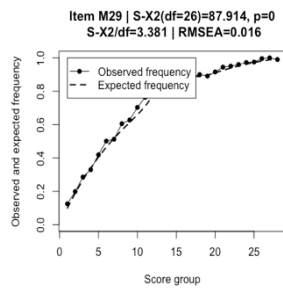
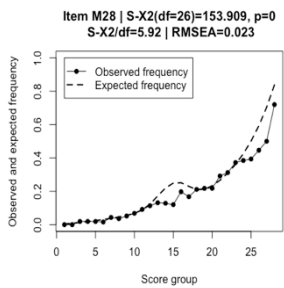
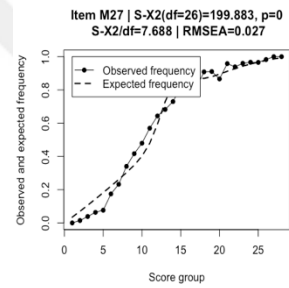
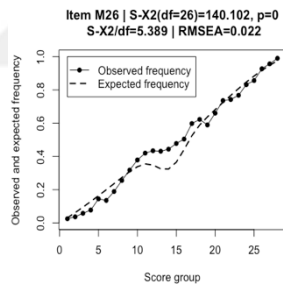
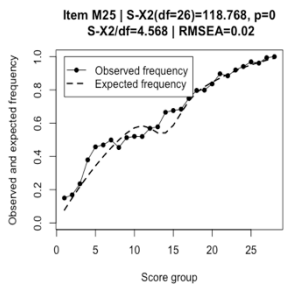
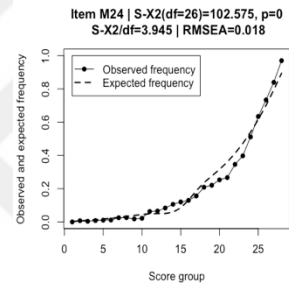
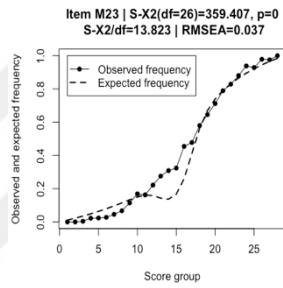
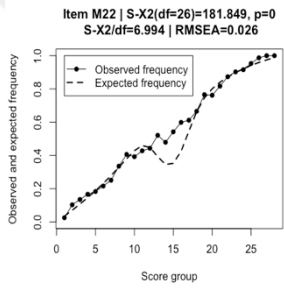
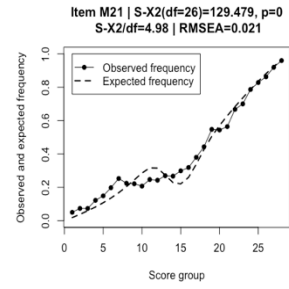
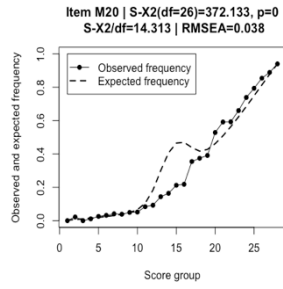
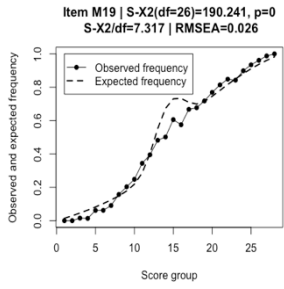
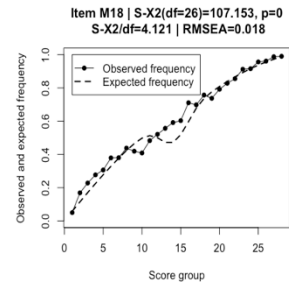
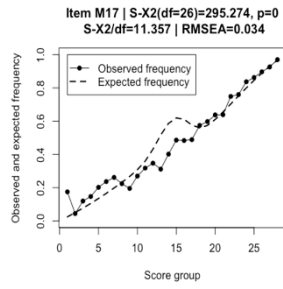
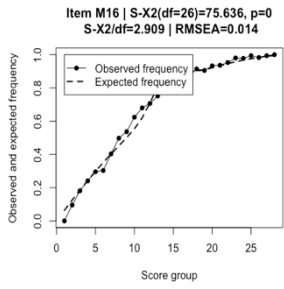
**Testin Farklı Sayıda Boyutlardan Oluşan Formlarına İlişkin Göreceli Model Uyumu İndeksleri**

Boyut Sayısı	Kestirim Yöntemi	-2LL	AIC	BIC	Estimated Parameters
1	EM	139332.5	278780.9	279194.3	58
	QMCEM	139332.5	278781.0	279194.5	58
	MHRM	139336.8	278789.6	279203.0	58
2	EM	138698.4	277568.9	278181.9	86
	QMCEM	138697.7	277567.3	278180.3	86
	MHRM	138710.1	277592.3	278205.3	86
3	EM	138500.4	277226.8	278032.3	113
	QMCEM	138482.3	277190.6	277996.1	113
	MHRM	138500.7	277227.4	278032.8	113
4	EM	138473.0	277224.0	278214.7	139
	QMCEM	138380.1	277038.3	278029.0	139
	MHRM	138409.0	277096.0	278086.8	139

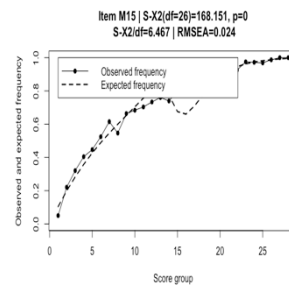
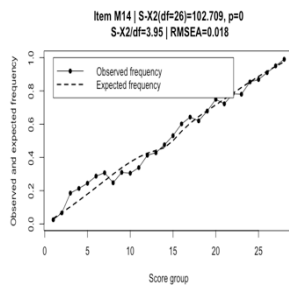
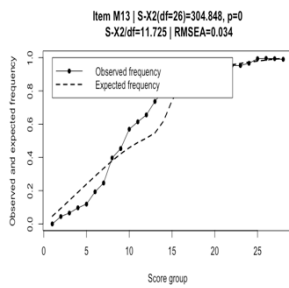
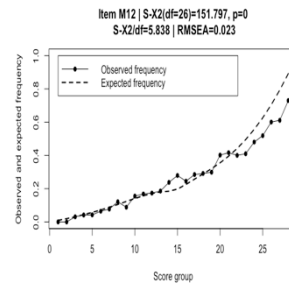
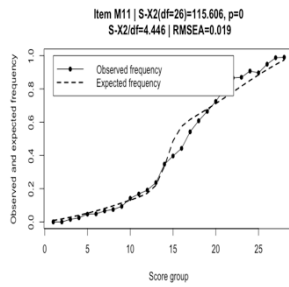
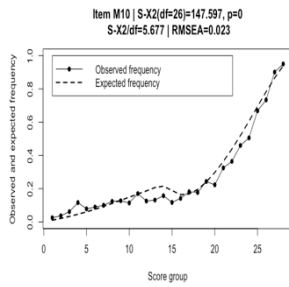
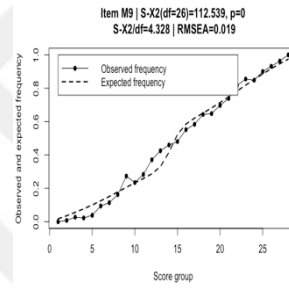
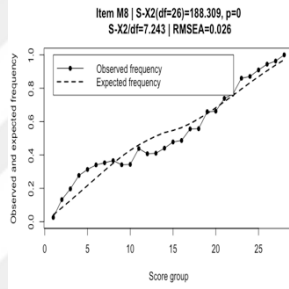
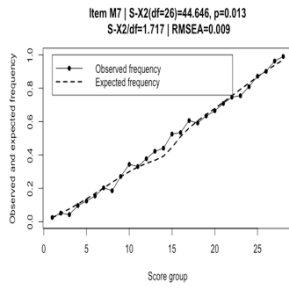
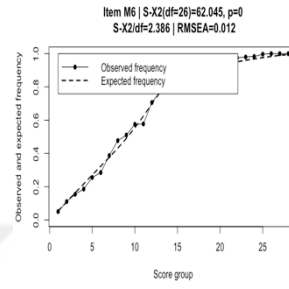
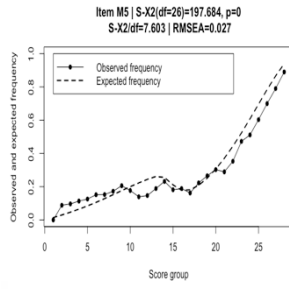
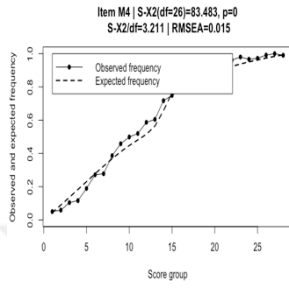
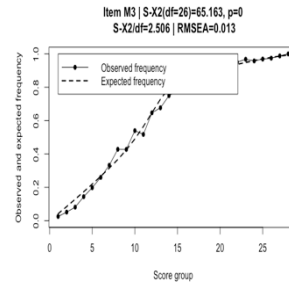
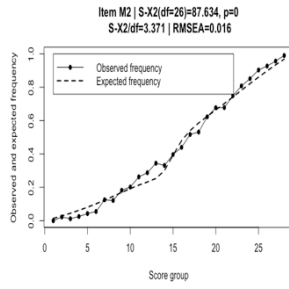
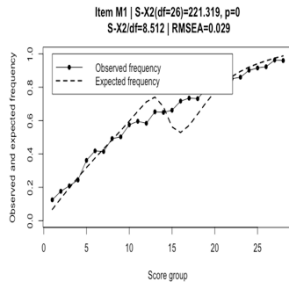
# EK-C: BTM'lere göre Kestirilen Madde Uyum İstatistikleri

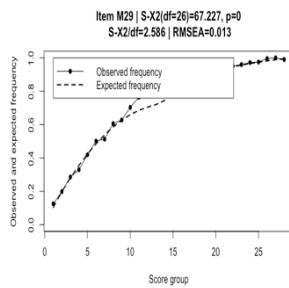
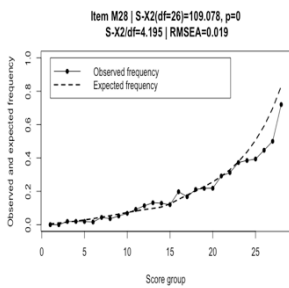
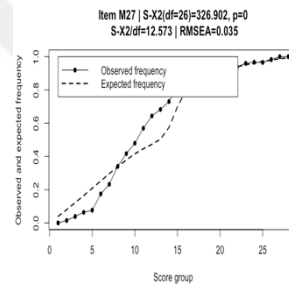
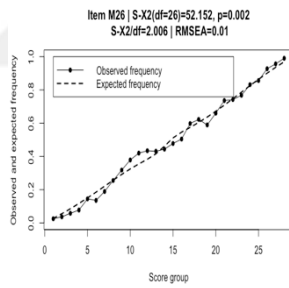
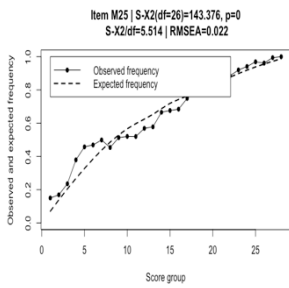
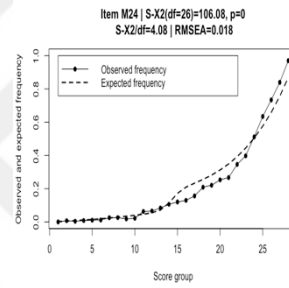
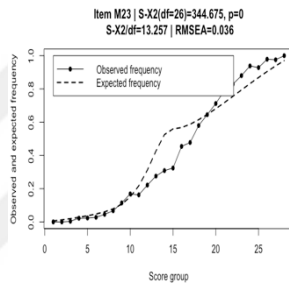
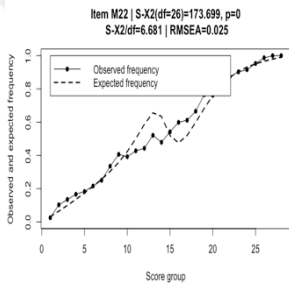
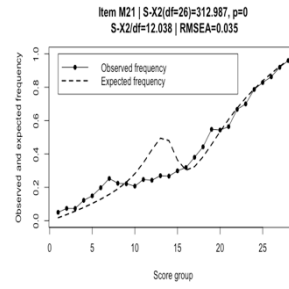
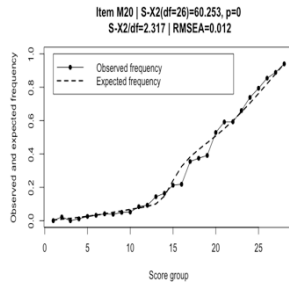
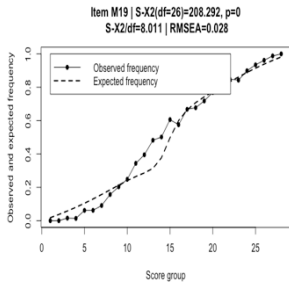
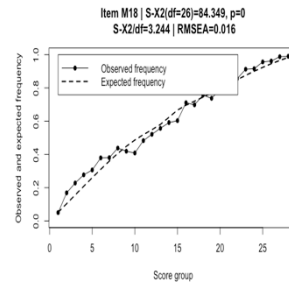
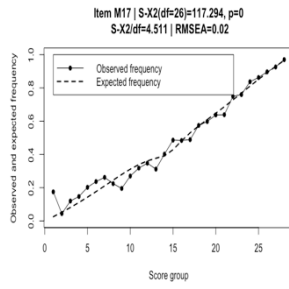
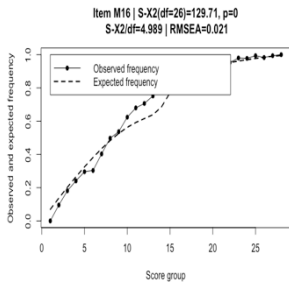
## DINA MODELİ



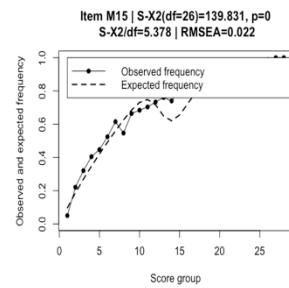
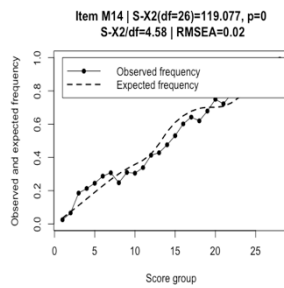
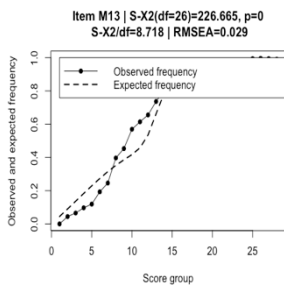
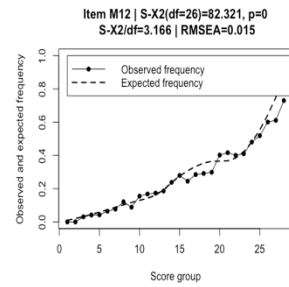
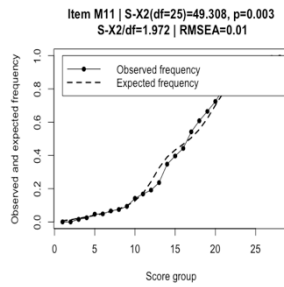
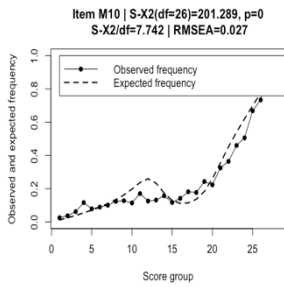
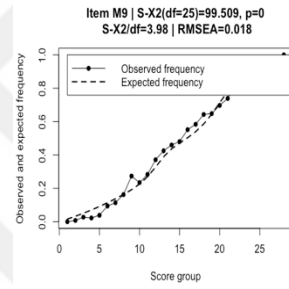
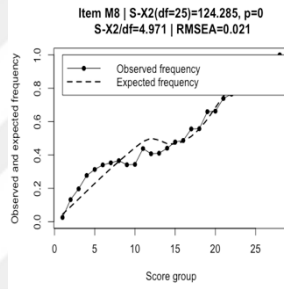
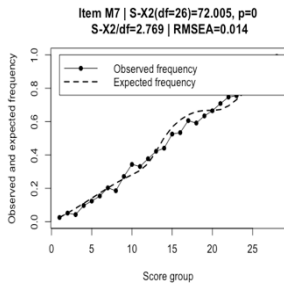
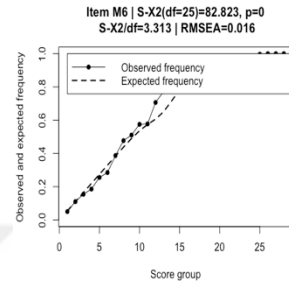
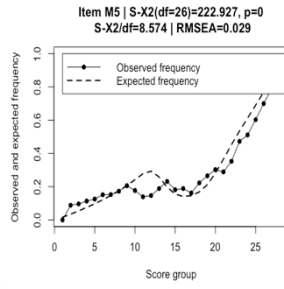
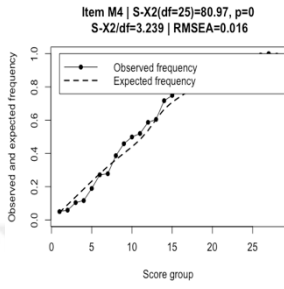
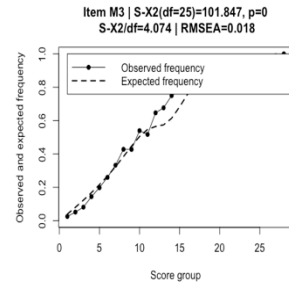
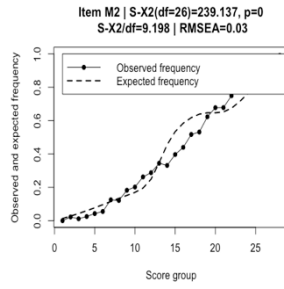
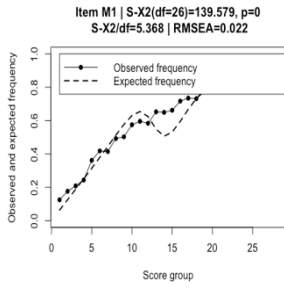


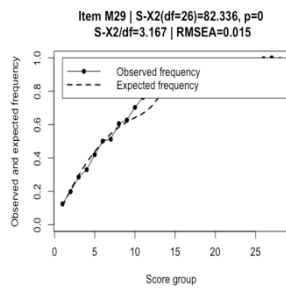
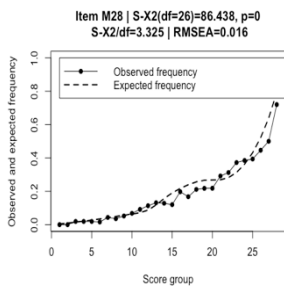
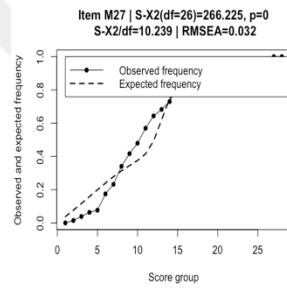
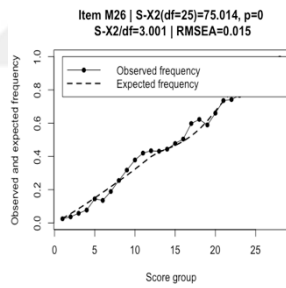
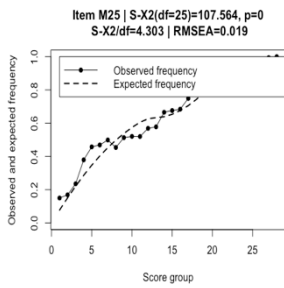
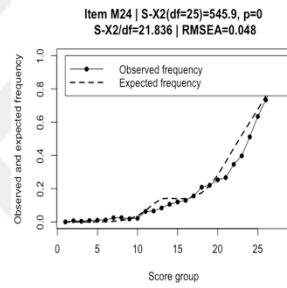
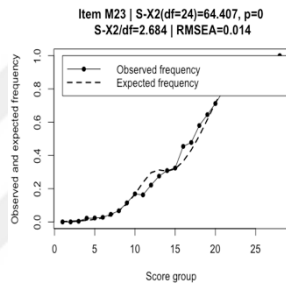
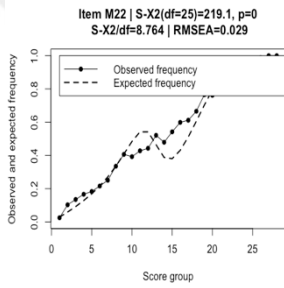
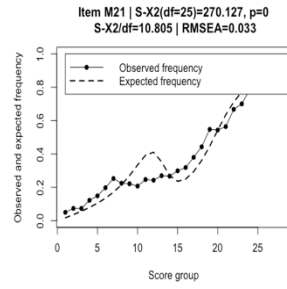
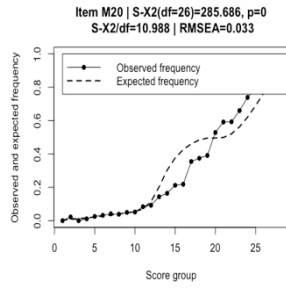
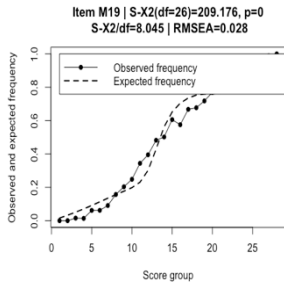
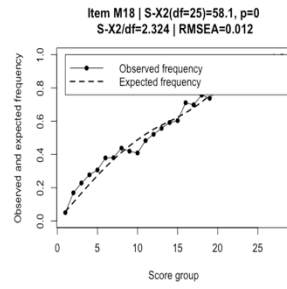
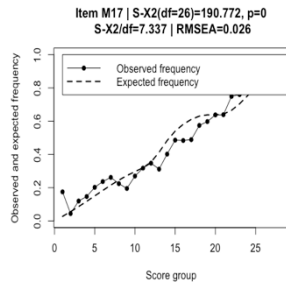
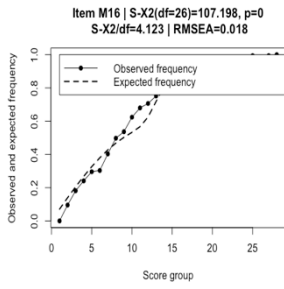
# DINO MODELİ





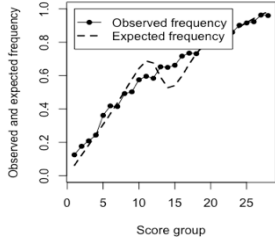
# A-CDM MODELİ



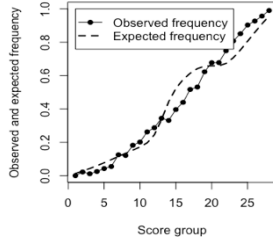


# LLM MODELİ

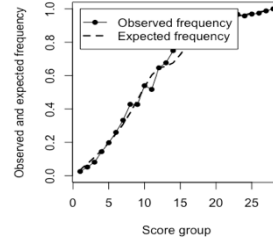
**Item M1** | S-X2(df=26)=142.254, p=0  
S-X2/df=5.471 | RMSEA=0.022



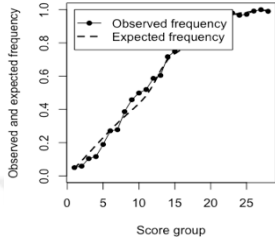
**Item M2** | S-X2(df=26)=235.527, p=0  
S-X2/df=9.059 | RMSEA=0.03



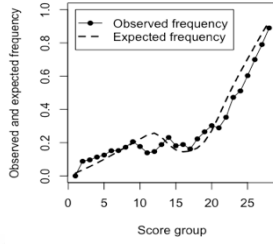
**Item M3** | S-X2(df=25)=46.568, p=0.006  
S-X2/df=1.863 | RMSEA=0.01



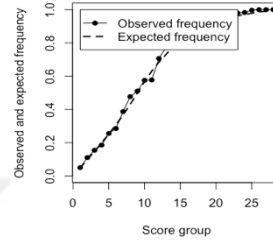
**Item M4** | S-X2(df=25)=63.319, p=0  
S-X2/df=2.533 | RMSEA=0.013



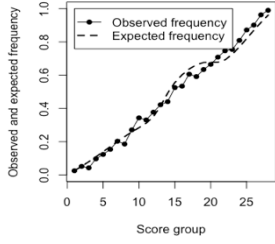
**Item M5** | S-X2(df=26)=201.93, p=0  
S-X2/df=7.767 | RMSEA=0.027



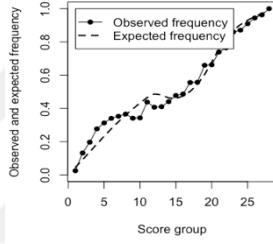
**Item M6** | S-X2(df=25)=48.963, p=0.003  
S-X2/df=1.959 | RMSEA=0.01



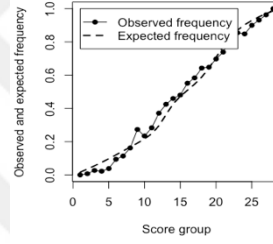
**Item M7** | S-X2(df=26)=71.367, p=0  
S-X2/df=2.745 | RMSEA=0.014



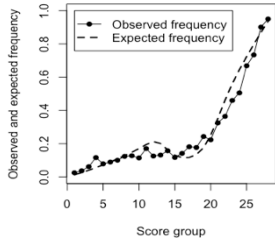
**Item M8** | S-X2(df=25)=121.602, p=0  
S-X2/df=4.864 | RMSEA=0.02



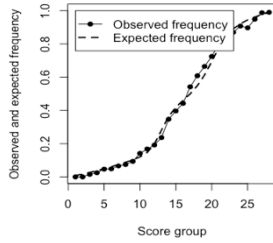
**Item M9** | S-X2(df=25)=116.541, p=0  
S-X2/df=4.662 | RMSEA=0.02



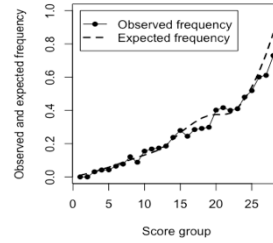
**Item M10** | S-X2(df=26)=178.699, p=0  
S-X2/df=6.873 | RMSEA=0.025



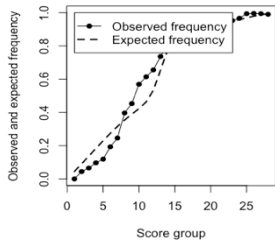
**Item M11** | S-X2(df=25)=50.891, p=0.001  
S-X2/df=2.036 | RMSEA=0.011



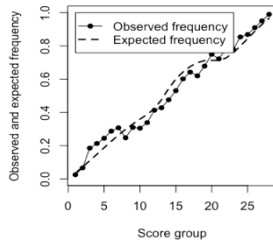
**Item M12** | S-X2(df=26)=71.823, p=0  
S-X2/df=2.762 | RMSEA=0.014



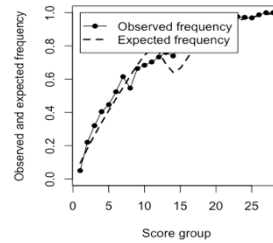
**Item M13** | S-X2(df=26)=222.521, p=0  
S-X2/df=8.559 | RMSEA=0.029



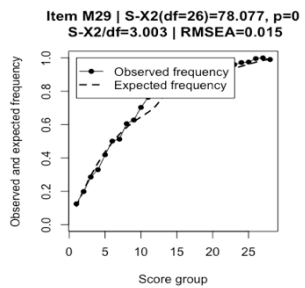
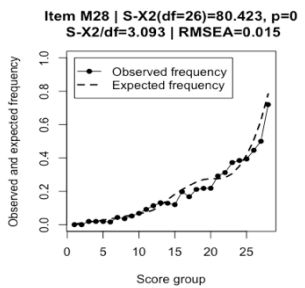
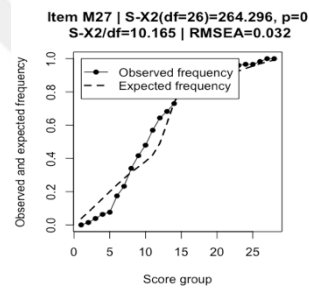
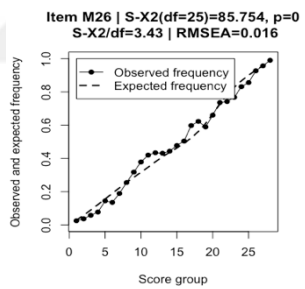
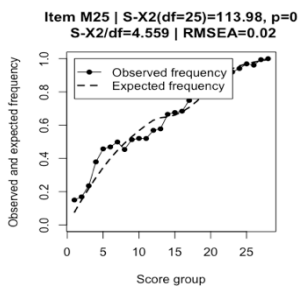
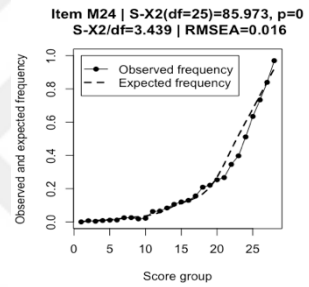
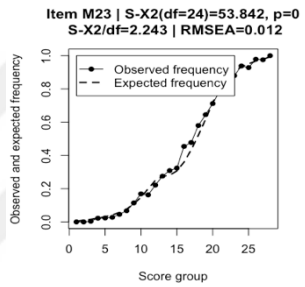
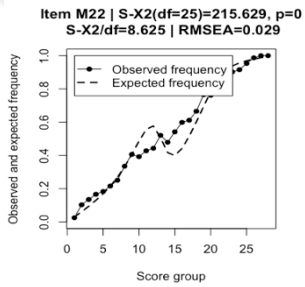
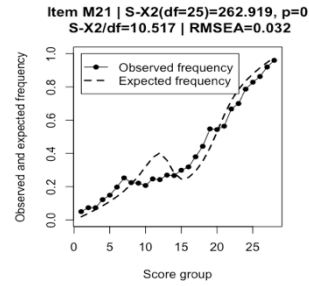
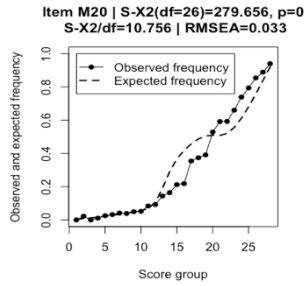
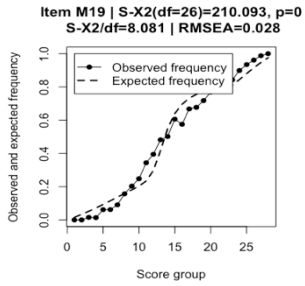
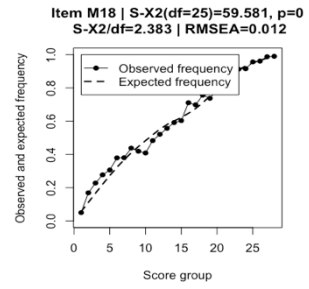
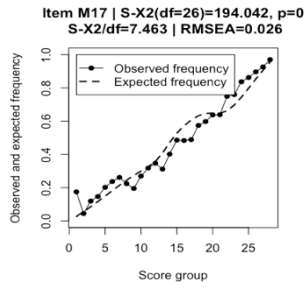
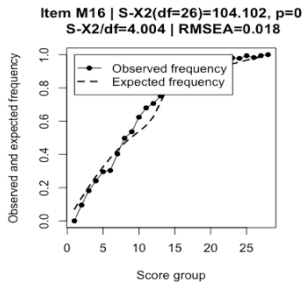
**Item M14** | S-X2(df=26)=121.028, p=0  
S-X2/df=4.655 | RMSEA=0.02



**Item M15** | S-X2(df=26)=137.73, p=0  
S-X2/df=5.297 | RMSEA=0.022

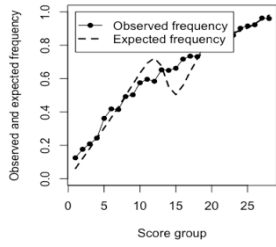




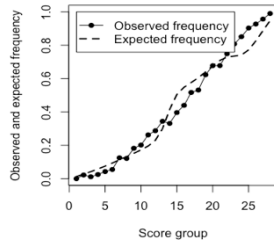


# R-RUM MODELİ

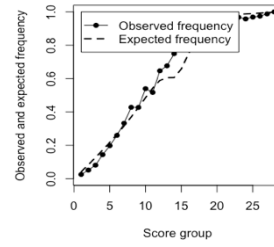
Item M1 | S-X2(df=26)=193.368, p=0  
S-X2/df=7.437 | RMSEA=0.026



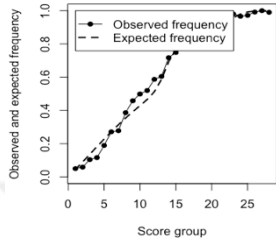
Item M2 | S-X2(df=26)=205.747, p=0  
S-X2/df=7.913 | RMSEA=0.027



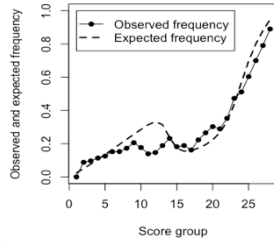
Item M3 | S-X2(df=25)=196.203, p=0  
S-X2/df=7.848 | RMSEA=0.027



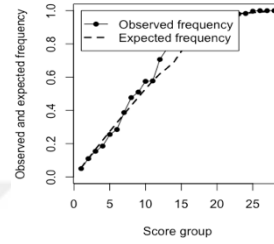
Item M4 | S-X2(df=25)=187.03, p=0  
S-X2/df=7.481 | RMSEA=0.027



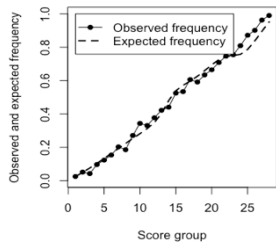
Item M5 | S-X2(df=26)=266.154, p=0  
S-X2/df=10.237 | RMSEA=0.032



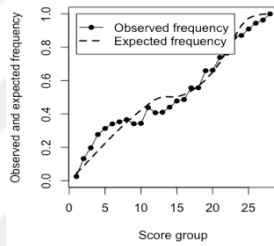
Item M6 | S-X2(df=25)=96.378, p=0  
S-X2/df=3.855 | RMSEA=0.018



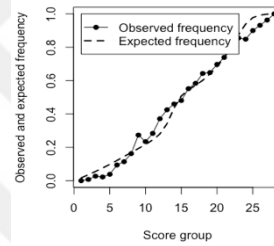
Item M7 | S-X2(df=26)=64.758, p=0  
S-X2/df=2.491 | RMSEA=0.013



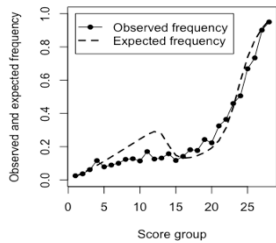
Item M8 | S-X2(df=25)=358.703, p=0  
S-X2/df=14.348 | RMSEA=0.038



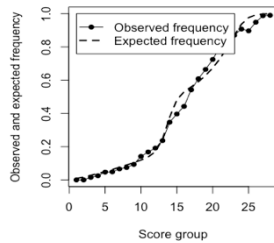
Item M9 | S-X2(df=25)=476.7, p=0  
S-X2/df=19.068 | RMSEA=0.044



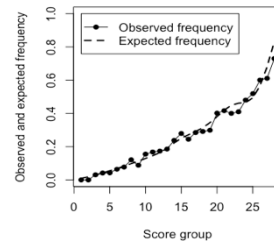
Item M10 | S-X2(df=26)=292.571, p=0  
S-X2/df=11.253 | RMSEA=0.033



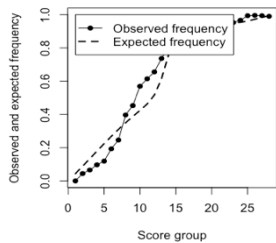
Item M11 | S-X2(df=25)=259.082, p=0  
S-X2/df=10.363 | RMSEA=0.032



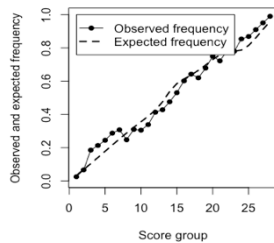
Item M12 | S-X2(df=26)=43.356, p=0.011  
S-X2/df=1.668 | RMSEA=0.009



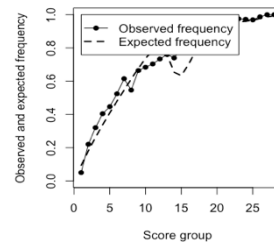
Item M13 | S-X2(df=26)=231.837, p=0  
S-X2/df=8.917 | RMSEA=0.029



Item M14 | S-X2(df=26)=126.14, p=0  
S-X2/df=4.852 | RMSEA=0.02



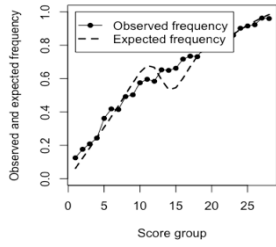
Item M15 | S-X2(df=26)=170.239, p=0  
S-X2/df=6.548 | RMSEA=0.025



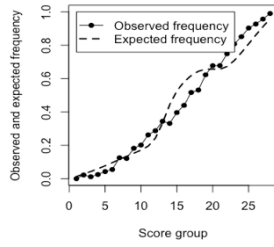


# G-DINA MODELİ

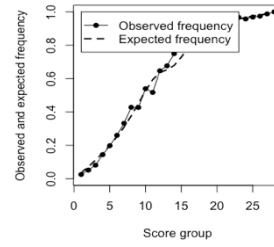
**Item M1** | S-X2(df=26)=129.289, p=0  
S-X2/df=4.973 | RMSEA=0.021



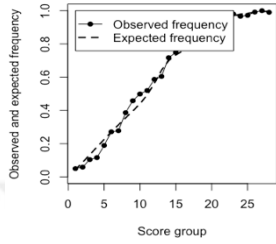
**Item M2** | S-X2(df=26)=240.254, p=0  
S-X2/df=9.241 | RMSEA=0.03



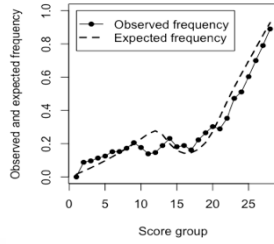
**Item M3** | S-X2(df=24)=43.254, p=0.009  
S-X2/df=1.802 | RMSEA=0.009



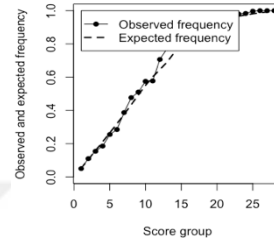
**Item M4** | S-X2(df=24)=57.505, p=0  
S-X2/df=2.396 | RMSEA=0.012



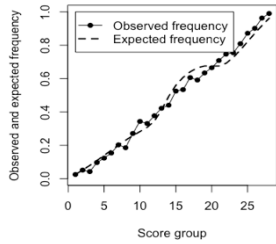
**Item M5** | S-X2(df=26)=214.587, p=0  
S-X2/df=8.253 | RMSEA=0.028



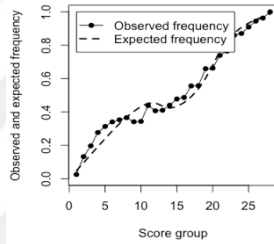
**Item M6** | S-X2(df=24)=56.712, p=0  
S-X2/df=2.363 | RMSEA=0.012



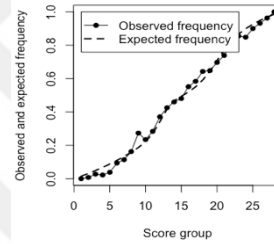
**Item M7** | S-X2(df=26)=73.892, p=0  
S-X2/df=2.842 | RMSEA=0.014



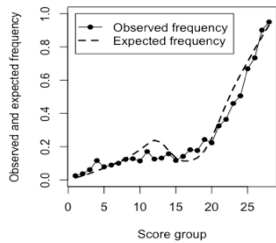
**Item M8** | S-X2(df=24)=109.798, p=0  
S-X2/df=4.575 | RMSEA=0.02



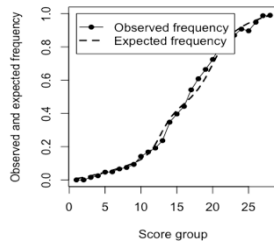
**Item M9** | S-X2(df=24)=82.384, p=0  
S-X2/df=3.433 | RMSEA=0.016



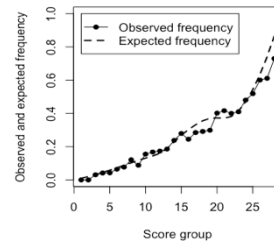
**Item M10** | S-X2(df=26)=191.945, p=0  
S-X2/df=7.382 | RMSEA=0.026



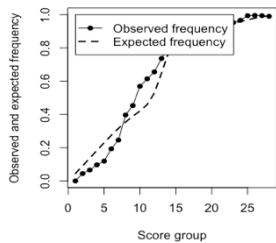
**Item M11** | S-X2(df=24)=48.879, p=0.00;  
S-X2/df=2.037 | RMSEA=0.011



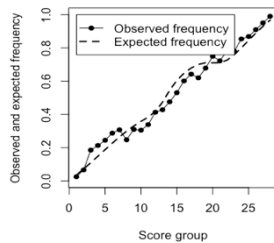
**Item M12** | S-X2(df=26)=75.892, p=0  
S-X2/df=2.919 | RMSEA=0.014



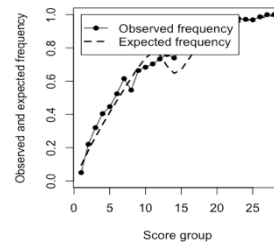
**Item M13** | S-X2(df=26)=230.079, p=0  
S-X2/df=8.849 | RMSEA=0.029



**Item M14** | S-X2(df=26)=121.445, p=0  
S-X2/df=4.671 | RMSEA=0.02



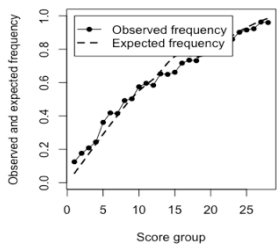
**Item M15** | S-X2(df=26)=127.149, p=0  
S-X2/df=4.89 | RMSEA=0.021



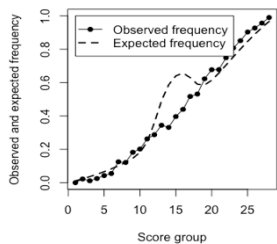


# HO-DINA MODELİ

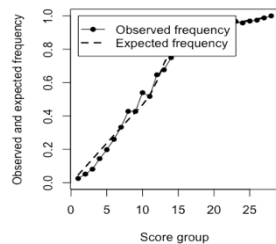
**Item M1** | S-X2(df=26)=87.59, p=0  
S-X2/df=3.369 | RMSEA=0.016



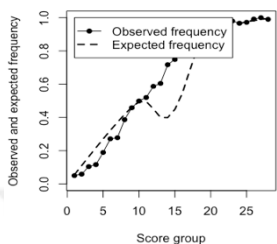
**Item M2** | S-X2(df=26)=374.996, p=0  
S-X2/df=14.423 | RMSEA=0.038



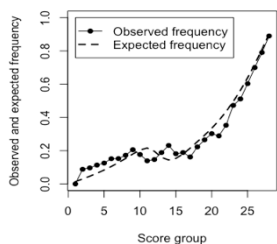
**Item M3** | S-X2(df=26)=81.127, p=0  
S-X2/df=3.12 | RMSEA=0.015



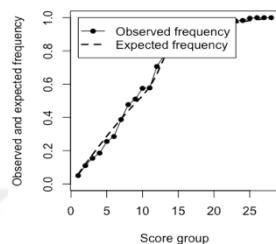
**Item M4** | S-X2(df=26)=560.737, p=0  
S-X2/df=21.567 | RMSEA=0.047



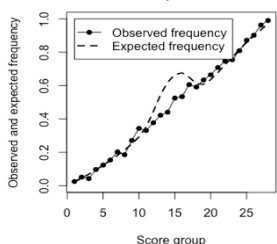
**Item M5** | S-X2(df=26)=160.047, p=0  
S-X2/df=6.156 | RMSEA=0.024



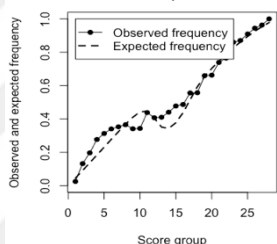
**Item M6** | S-X2(df=26)=64.396, p=0  
S-X2/df=2.477 | RMSEA=0.013



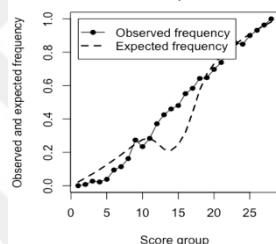
**Item M7** | S-X2(df=26)=158.09, p=0  
S-X2/df=6.08 | RMSEA=0.023



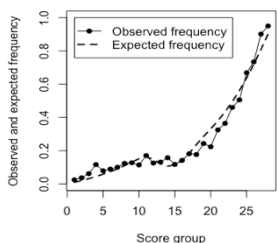
**Item M8** | S-X2(df=26)=135.561, p=0  
S-X2/df=5.214 | RMSEA=0.021



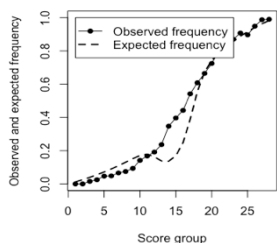
**Item M9** | S-X2(df=26)=566.691, p=0  
S-X2/df=21.796 | RMSEA=0.048



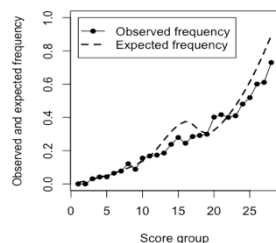
**Item M10** | S-X2(df=26)=139.244, p=0  
S-X2/df=5.356 | RMSEA=0.022



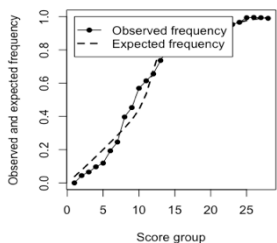
**Item M11** | S-X2(df=26)=457.981, p=0  
S-X2/df=17.615 | RMSEA=0.042



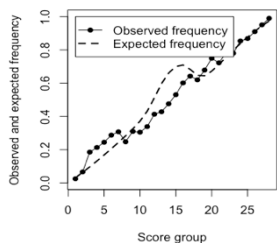
**Item M12** | S-X2(df=26)=185.562, p=0  
S-X2/df=7.137 | RMSEA=0.026



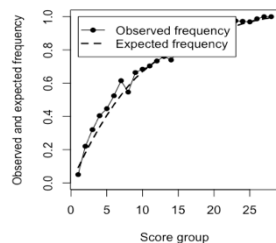
**Item M13** | S-X2(df=26)=206.804, p=0  
S-X2/df=7.954 | RMSEA=0.027



**Item M14** | S-X2(df=26)=303.891, p=0  
S-X2/df=11.688 | RMSEA=0.034



**Item M15** | S-X2(df=26)=101.206, p=0  
S-X2/df=3.893 | RMSEA=0.018





## EK-Ç: Çalışmada Kullanılan Q-Matrisleri

Test Uzunluğunun 30 Madde Olduğu Test Koşulları için Kullanılan Q-Matrisleri

Gerçek Verinin Faktör Yapısına göre Oluşturulan Q-Matrisi				Nitelik ve Nitelik Profillerinin Eşit Sayıda Ölçüldüğü Q-Matrisi			
Nitelik				Nitelik			
Madde	$\alpha_1$	$\alpha_2$	$\alpha_3$	Madde	$\alpha_1$	$\alpha_2$	$\alpha_3$
1	0	1	1	1	1	0	0
2	0	1	1	2	1	0	0
3	0	1	1	3	1	0	0
4	0	0	1	4	1	0	0
5	0	0	1	5	1	0	0
6	1	0	1	6	1	0	1
7	1	0	1	7	1	0	1
8	1	0	1	8	1	0	1
9	1	0	0	9	1	0	1
10	1	0	0	10	1	0	1
11	1	0	0	11	1	1	0
12	1	0	0	12	1	1	0
13	1	0	0	13	1	1	0
14	1	0	0	14	1	1	0
15	1	0	0	15	1	1	0
16	1	0	0	16	0	1	0
17	1	0	0	17	0	1	0
18	1	0	0	18	0	1	0
19	1	0	0	19	0	1	0
20	1	1	0	20	0	1	0
21	1	1	0	21	0	1	1
22	1	1	0	22	0	1	1
23	1	1	0	23	0	1	1
24	1	1	0	24	0	1	1
25	1	1	0	25	0	1	1
26	1	1	0	26	0	0	1
27	0	1	0	27	0	0	1
28	0	1	0	28	0	0	1
29	0	1	0	29	0	0	1
30	0	1	0	30	0	0	1



Test Uzunluğunun 15 Madde Olduğu Test Koşulları için Kullanılan Q-Matrisleri

Gerçek Verinin Faktör Yapısına göre Oluşturulan Q-Matrisi				Nitelik ve Nitelik Profillerinin Eşit Sayıda Ölçüldüğü Q-Matrisi			
Nitelik				Nitelik			
Madde	$\alpha_1$	$\alpha_2$	$\alpha_3$	Madde	$\alpha_1$	$\alpha_2$	$\alpha_3$
1	0	1	1	1	1	0	1
2	0	1	1	2	1	0	1
3	0	0	1	3	1	0	1
4	1	0	1	4	1	0	0
5	1	0	0	5	1	1	0
6	1	0	0	6	1	1	0
7	1	0	0	7	1	1	0
8	1	0	0	8	1	1	0
9	1	0	0	9	0	1	0
10	1	0	0	10	0	1	0
11	1	1	0	11	0	1	0
12	1	1	0	12	0	0	1
13	1	1	0	13	0	0	1
14	1	1	0	14	0	0	1
15	0	1	0	15	0	0	1

## EK-D: BTM'lere göre Kestirilen Model Uyum İstatistikleri

Bilişsel Tanı Modelleri için Kestirilen Göreceli ve Mutlak Model Uyum İndeksleri

		Modeller						
		DINA	DINO	A-CDM	LLM	R-RUM	G-DINA	HO-DINA
Göreceli	2LL	285134	285544	281828	281771	281946	281666	285148
	AIC	285264	285674	281986	281929	282104	281856	285276
	BIC	285727	286137	282549	282492	282667	282533	285732
	CAIC	285792	286202	282628	282571	282746	282628	285796
	SABIC	285521	285930	282298	282241	282416	282231	285528
	PS	65	65	79	79	79	95	64
Mutlak	M2	2451.34	2531.69	1955.96	1917.41	1896.32	1865.51	2481.55
		4	4	3	1	9	2	7
	df	370	370	356	356	356	340	371
	RMSEA	0.0247	0.0252	0.0221	0.0218	0.0217	0.0221	0.0249
	SRMSR	0.0558	0.0531	0.0449	0.0444	0.0450	0.0448	0.0567

PS: Parametre sayısı

## EK-E: Etik Kurul İzin Muafiyeti Formu

Form: 40

### Tez Çalışması Etik Kurul İzin Muafiyeti Formu

09 / 03 / 2016

Hacettepe Üniversitesi  
Eğitim Bilimleri Enstitüsü  
Eğitim Bilimleri Anabilim Dalı Başkanlığı'na

**Tez Başlığı / Konusu:** HİYERARŞİK NİTELİK YAPILARI ALTINDA BİLİŞSEL TANI VE ÇOK BOYUTLU MADDE TEPKİ MODELLERİNİN KARŞILAŞTIRILMASI

Yukarıda başlığı/konusu gösterilen tez çalışmam:

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır,
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne müdahale içermemektedir.
4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması, sistem-model geliştirme çalışmaları) niteliğinde değildir.

Hacettepe Üniversitesi Etik Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabül ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

  
Elif Özlem ARDIÇ  
(Öğrencinin Adı Soyadı, İmzası)

#### Öğrenci Bilgileri

Adı Soyadı	Elif Özlem Ardiç
Öğrenci No	N12240666
Anabilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr.

#### Danışman Görüşü ve Onayı

Prof.Dr. Hülya KELECİOĞLU  
(İmza)  
(Danışmanın Ünvanı, Adı ve Soyadı)



## 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 bütününe kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

29/ 01/ 2020



Elif Özlem ARDIÇ

## EK-G: Doktora Tez Çalışması Orijinallik Raporu

29/ 01 /2020

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

Tez Başlığı: Bilişsel Tanı ve Çok Boyutlu Madde Tepki Modellerinin Sınıflama Doğruluğu ve Parametrelerinin Karşılaştırılması

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

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
27/ 01/ 2020	128	215,769	16/ 01/ 2020	%4	1247246332

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:** Elif Özlem ARDIÇ

**Öğrenci No.:** N12240666

**Ana Bilim Dalı:** Eğitim Bilimleri

**Programı:** Eğitimde Ölçme ve Değerlendirme

**Statüsü:**  Y.Lisans  Doktora  Bütünleşik Dr.

### DANIŞMAN ONAYI

UYGUNDUR.  
Prof. Dr. Hülya KELECİOĞLU

## EK-Ğ: Dissertation Originality Report

29/ 01/ 2020

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

Thesis Title: Comparison of Classification Accuracy and Parameters of Cognitive Diagnostic and Multidimensional Item Response Models

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
27/ 01/ 2020	128	215,769	16/ 01/ 2020	%4	1247246332

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:** Elif Özlem ARDIÇ

**Student No.:** N12240666

**Department:** Educational Science

**Program:** Educational Measurement and Evaluation

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



### ADVISOR APPROVAL

APPROVED  
Prof. Dr. Hülya KELECİOĞLU



## 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.

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

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

- o Enstitü/ Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. <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>

29 /01/ 2020



Elif Özlem ARDIÇ

"*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 internette paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.*

(3) *Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir\*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.*

*Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir*

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

